Computer Modeling in Engineering & Sciences

Tech Science Press

https://doi.org/10.32604/cmes.2026.084193

ARTICLE

Learning Scenario-Dependent Construction Strategy Selection Policies Using a Hybrid T-Spherical Fuzzy CRITIC–CoCoSo RankNet Framework

Yih-Tzoo Chen[1] **and Thi-Hien Dao**[2,*]

[1]Department of Construction Engineering, College of Engineering, National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan

[2]Ph.D. Program in Engineering Science and Technology, College of Engineering, National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan

*Corresponding Author: Thi-Hien Dao. Email: I112109110@nkust.edu.tw

Received: 17 April 2026; Accepted: 23 July 2026; Published: 28 August 2026

ABSTRACT: Construction strategy selection is a context-dependent decision problem in which multiple conflicting criteria must be considered under changing project conditions. Conventional multi-criteria decision-making (MCDM) methods generally produce rankings for predefined decision matrices but do not learn transferable preference structures across project scenarios. This study proposes a hybrid framework integrating T-spherical fuzzy sets, the Criteria Importance Through Intercriteria Correlation (CRITIC) method, the Combined Compromise Solution (CoCoSo) method, and RankNet-based pairwise learning. Fifty construction scenarios were designed using six contextual variables, and eight construction strategies were evaluated against eight criteria. Linguistic evaluations were transformed using a seven-level T-spherical fuzzy scale and the neutrality-aware score function proposed by Munir et al. Scenario-specific CRITIC–CoCoSo rankings were subsequently converted into 1400 unordered pairwise observations. To prevent information leakage and evaluate transferability, all pairs belonging to the same scenario were assigned to the same subset, with 35 scenarios used for training, seven for validation, and eight unseen scenarios for testing in each repeated split. Across ten random seeds, RankNet achieved a mean test accuracy of 0.9018 ± 0.0234, a mean Spearman rank correlation of 0.8997 ± 0.0308, and a local perturbation robustness correlation of 0.9966 ± 0.0020. Comparative analysis showed that RankNet performed comparably to logistic regression and the multilayer perceptron classifier while outperforming the examined tree-based ensemble models in ranking recovery. The ablation results further showed that interaction features improved the recovery of scenario-dependent ranking structures more clearly than direct scenario variables alone. SHAP analysis identified flexibility, risk, and their interactions with project urgency, supply instability, budget pressure, and weather uncertainty as major drivers of the learned preference policy. The proposed framework provides an interpretable and computationally efficient approach for learning construction strategy preferences under dynamic and uncertain project conditions.

KEYWORDS: T-spherical fuzzy sets; multi-criteria decision-making; CRITIC method; CoCoSo method; RankNet; pairwise learning-to-rank; construction strategy selection; scenario-based decision analysis; interaction features; SHAP explainability

1 Introduction

Construction projects are inherently complex and dynamic systems characterized by uncertainty, resource constraints, and the need to simultaneously satisfy multiple conflicting objectives such as cost, time, quality, safety, and sustainability [1–3]. The selection of an appropriate construction strategy is therefore a

critical decision that significantly influences project outcomes. This decision is not only multi-dimensional but also highly sensitive to contextual conditions, including labor availability, weather variability, supply chain stability, environmental requirements, budget pressure, and project urgency.

In recent years, multi-criteria decision-making (MCDM) methods have been widely applied to support decision-making in construction management [4–6]. These approaches provide structured mechanisms for evaluating alternatives under multiple criteria and have been used in contractor selection, construction technology assessment, risk prioritization, and sustainability-oriented decision problems [4,5]. Classical methods such as TOPSIS, VIKOR, and PROMETHEE remain widely adopted due to their ability to model trade-offs among competing objectives [7–9]. Meanwhile, objective weighting techniques such as the CRITIC method allow criteria importance to be derived from data variability and inter-criteria conflict [10], while the CoCoSo method has emerged as an effective ranking approach for balancing additive and multiplicative compromise solutions [11].

Beyond individual MCDM techniques, hybrid MCDM frameworks have increasingly been developed to address complex engineering decision problems involving heterogeneous information, uncertain preferences, and multiple stakeholder perspectives. Recent studies have combined fuzzy, linguistic, hesitant, and spherical fuzzy representations with weighting and ranking procedures to improve the flexibility of decision modeling in areas such as supplier selection, infrastructure assessment [12], and other decision-making [13,14]. These studies indicate that hybrid MCDM methods are useful for structuring complex judgments and improving decision transparency. However, most of them remain evaluation-oriented: they produce rankings for a given decision matrix but do not convert accumulated scenario-specific decisions into a reusable decision policy. Recent studies on complex heterogeneous multi-attribute group decision-making have emphasized the importance of integrating different information formats, preference structures, and aggregation mechanisms when decision environments involve multiple stakeholders and uncertain judgments [15–17]. Such frameworks are useful because real decision problems often contain heterogeneous evaluation information rather than a single homogeneous data type. Similarly, linguistic hesitant fuzzy multi-criteria group decision-making has been developed to represent situations in which experts hesitate among several linguistic terms when expressing their preferences [18]. These developments indicate a broader movement toward richer uncertainty representation and more flexible group-decision mechanisms. Nevertheless, such methods generally focus on obtaining a ranking within a specified decision setting. Relatively little attention has been given to learning a transferable decision rule from multiple scenario-specific fuzzy MCDM outcomes.

Despite these advances, a key limitation of existing MCDM approaches is their reliance on static evaluation structures, where a single ranking is produced under fixed assumptions [19,20]. In practice, construction environments are highly dynamic, and the performance of strategies may vary significantly across project scenarios [21,22]. For example, a cost-minimization strategy may be effective under stable conditions but become less suitable under severe labor shortages or supply disruptions, whereas sustainability-oriented strategies may become more favorable when environmental or regulatory requirements become stricter [23,24]. Accordingly, a single global ranking may obscure the conditional suitability of strategies and provide limited guidance when project conditions change.

To address uncertainty in expert judgment, fuzzy set theory has been extensively integrated into MCDM models [12,14]. Advanced fuzzy extensions, including Pythagorean fuzzy sets, q-rung orthopair fuzzy sets, and T-spherical fuzzy sets, have been proposed to better represent uncertainty and hesitation [25–27]. Among these, T-spherical fuzzy sets provide a flexible and generalized representation by incorporating adjustable parameters that allow simultaneous modeling of membership, non-membership, and hesitation degrees [13]. Nevertheless, the quality of a T-spherical fuzzy ranking depends on the score function used to transform fuzzy evaluations into comparable numerical values. A score function that omits the neutrality

degree may discard part of the original assessment information. To address this issue, the present study adopts the established neutrality-aware score function proposed by Munir et al. [28], rather than introducing an additional study-specific scoring parameter.

Although these approaches improve uncertainty modeling, they still primarily operate within scenario-specific evaluation frameworks, generating isolated rankings without capturing generalizable decision patterns across multiple contexts [5]. Artificial intelligence (AI), particularly deep learning, has demonstrated strong capabilities in modeling nonlinear relationships and learning from high-dimensional data [29,30]. In the construction domain, AI techniques have been increasingly applied to tasks such as cost estimation, schedule prediction, productivity analysis, and risk assessment [31,32]. However, the application of AI to multi-criteria decision-making problems, especially those involving dynamic and scenario-dependent environments, remains limited [33]. Most existing AI applications in construction focus on prediction tasks, whereas fewer studies examine how AI can learn decision preferences or decision policies from structured MCDM outputs.

Pairwise learning models provide a promising way to address this limitation because they learn preference relationships between alternatives rather than only predicting numerical outcomes [34]. RankNet, originally introduced by Burges et al. [35], models the preference probability between two alternatives from the difference between their predicted scores and optimizes a pairwise cross-entropy objective. This characteristic makes it suitable for construction strategy selection, where the preferred alternative depends not only on the strategy itself but also on the project scenario in which it is implemented. However, applying such a model requires structured scenario-dependent decision data, which are rarely available in real construction datasets. A scenario-driven design is therefore required to generate sufficiently diverse preference observations while retaining transparent links between contextual conditions, fuzzy evaluations, and MCDM-derived reference rankings.

Based on the above discussion, four research gaps can be identified. First, most MCDM studies in construction rely on static decision matrices and do not explicitly incorporate dynamic scenario variations [6]. Second, although AI methods are increasingly used in construction, their integration with fuzzy MCDM frameworks for learning and generalizing decision behavior remains underexplored [36,37]. Third, existing hybrid MCDM–AI studies have not sufficiently examined whether scenario variables and interaction features contribute to learning decision policies beyond simply reproducing MCDM-generated rankings. This issue is important because a learning model should not only imitate a ranking method but also capture transferable preference patterns across changing project conditions. Fourth, model evaluation may overstate generalization when pairwise observations from the same scenario are randomly distributed across training and testing subsets. A more rigorous assessment requires all observations belonging to a scenario to remain within the same data partition so that the model is evaluated on genuinely unseen scenarios.

To address these gaps, this study proposes a hybrid framework that integrates T-spherical fuzzy evaluation, CRITIC weighting, CoCoSo ranking, and RankNet-based pairwise learning for scenario-dependent construction strategy selection. Fifty project scenarios are constructed using six contextual variables: budget pressure, labor shortage, weather uncertainty, supply instability, sustainability requirement, and project urgency. Eight construction strategies are evaluated against eight economic, temporal, environmental, technical, safety, risk, labor, and flexibility criteria. The scenario definitions and baseline strategy profiles are calibrated to preserve managerial meaning; for example, fast-track, labor-saving, sustainability-oriented, and risk-averse strategies are designed to exhibit their expected comparative advantages under the corresponding contextual conditions.

Linguistic evaluations are represented using a seven-level T-spherical fuzzy scale and converted into crisp values using the neutrality-aware score function of Munir et al. [28]. CRITIC is then used to obtain

scenario-specific objective criteria weights, and CoCoSo is applied to generate structured scenario-specific reference rankings. For each scenario, the ranking of eight strategies is transformed into 28 unordered pairwise observations, producing 1400 preference observations across the 50 scenarios. Reverse-pair duplication is excluded so that each alternative pair contributes only one observation per scenario.

The pairwise observations are used to train a RankNet scoring model based on alternative-level attributes, scenario-level variables, and their interaction features. To prevent scenario leakage, data partitioning is performed at the scenario level rather than at the individual-pair level. In each repeated experiment, 35 scenarios are assigned to training, seven to validation, and eight previously unseen scenarios to testing. All pairwise observations belonging to the same scenario remain in the same subset. This scenario-group evaluation provides a more rigorous test of whether the learned preference policy can be transferred to project conditions not observed during training.

The role of the AI component is not to replace fuzzy MCDM or to establish an independent empirical ground truth. Instead, RankNet learns a transferable scoring policy from structured MCDM-derived reference rankings. Its added value is evaluated through comparisons with logistic regression, random forest, gradient boosting, and multilayer perceptron models. Feature ablation is also used to distinguish the contribution of alternative-level, direct scenario-level, and interaction features. Because direct scenario variables are identical for all strategies within the same scenario, particular attention is given to interaction features that convert contextual conditions into alternative-specific preference information.

Finally, SHapley Additive exPlanations (SHAP), introduced by Lundberg and Lee [38], are used to interpret the learned RankNet scoring policy. SHAP values quantify how alternative attributes and alternative–scenario interaction terms contribute to predicted preference scores. The explanations are computed on observations from unseen test scenarios, thereby linking model interpretation to the same scenario-level generalization setting used for performance evaluation. SHAP values are treated as model-attribution measures rather than causal effects.

Accordingly, the main objectives of this study are as follows:

1. To develop a scenario-driven framework for construction strategy selection under dynamic and uncertain project conditions.
2. To integrate a seven-level T-spherical fuzzy representation and the neutrality-aware score function of Munir et al. with CRITIC weighting and CoCoSo ranking.
3. To construct non-duplicated pairwise preference data from scenario-specific MCDM-derived rankings and train a RankNet-based neural ranking policy.
4. To evaluate preference-policy transferability using repeated scenario-group holdout experiments on unseen project scenarios.
5. To assess the contribution of alternative, scenario, and interaction features through baseline comparison and feature-ablation analysis.
6. To evaluate the predictive performance, local perturbation stability, computational efficiency, and interpretability of the framework using multi-seed experiments and SHAP analysis.

The remainder of this paper is organized as follows. Section 2 presents the T-spherical fuzzy preliminaries and the proposed scenario-generation, CRITIC–CoCoSo, pairwise-learning, and SHAP procedures. Section 3 reports the scenario design, fuzzy MCDM results, scenario-group learning performance, baseline and ablation comparisons, and explainability analysis. Section 4 discusses the theoretical and managerial implications, together with the advantages and limitations of the proposed framework. Section 5 concludes the study and identifies directions for empirical validation and methodological extension.

2 Materials and Methods

2.1 Preliminaries of T-Spherical Fuzzy Sets

T-spherical fuzzy sets (T-SFSs) are employed to represent uncertainty, neutrality, and non-membership in linguistic evaluations of construction strategies. Compared with conventional fuzzy and intuitionistic fuzzy sets, T-SFSs provide a more flexible assessment domain because the membership, neutrality, and non-membership degrees are constrained through an adjustable exponent t.

Definition 1 [39]: *Let X be a universe of discourse. A T-spherical fuzzy set (T-SFS) $\tilde{A}$ on X is defined as*

$$\tilde{A} = \{(x, \alpha_{\tilde{A}}(x), \gamma_{\tilde{A}}(x), \beta_{\tilde{A}}(x)) | x \in X\} \tag{1}$$

$$where\ \alpha, \gamma, \beta: X \to [0,1]\ \ and\ 0 \le \alpha_{\tilde{A}}^{t}(x) + \gamma_{\tilde{A}}^{t}(x) + \beta_{\tilde{A}}^{t}(x) \le 1 \quad \forall t \in \mathbb{N}, x \in X \tag{2}$$

for a prescribed positive integer t. Here, $\alpha_{\tilde{A}}(x)$, $\gamma_{\tilde{A}}(x)$, and $\beta_{\tilde{A}}(x)$ denote the membership, neutrality, and non-membership degrees of x, respectively. The corresponding refusal degree is given by

$$r(x) = (1 - (\alpha^{t}(x) + \gamma^{t}(x) + \beta^{t}(x)))^{\frac{1}{t}} \tag{3}$$

Accordingly, the quadruple $(\alpha_{\tilde{A}}(x), \gamma_{\tilde{A}}(x), \beta_{\tilde{A}}(x))$ characterizes the membership, neutrality, non-membership, and refusal information associated with x. For convenience, a T-spherical fuzzy number (T-SFN) is denoted by $\tilde{A} = (\alpha, \gamma, \beta)$.

Definition 2 [28]: *For a T-SFN $\tilde{A} = (\alpha, \gamma, \beta)$, its score value is defined as*

$$SV(\tilde{A}) = \alpha^{t} - \gamma^{t} - \beta^{t} \tag{4}$$

where

$$-1 \le SV(\tilde{A}) \le 1 \tag{5}$$

Unlike a score function based only on membership and non-membership, Eq. (4) explicitly incorporates the neutrality or abstinence degree γ. Therefore, all three principal components of a T-SFN are retained when the linguistic evaluation is transformed into a numerical score. A larger value of $SV(\tilde{A})$ indicates a more favorable assessment. The accuracy function of a T-SFN is defined as:

$$AV(\tilde{A}) = \alpha^{t} + \gamma^{t} + \beta^{t} \tag{6}$$

For any valid T-SFN, the accuracy value lies in the interval $[0,1]$ because it is equal to the feasibility sum $\alpha^{t} + \gamma^{t} + \beta^{t}$. A larger accuracy value indicates a more determinate assessment. The accuracy function is used only as a secondary comparison rule when two T-SFNs have identical score values.

Remark 1: *Definitions 1 and 2 can be reduced to the following special cases:*

(1) *PFS if $t = 1$;*
(2) *SFS if $t = 2$;*
(3) *q-ROFS if $\gamma = 0$;*
(4) *PyFS if $t = 2$ and $\gamma = 0$;*
(5) *IFS if $t = 1$ and $\gamma = 0$;*
(6) *FS if $t = 1$, $\gamma = 0$, and $\beta = 0$.*

Definition 3 [39]: *Consider two T-SFNs $\tilde{A}_1 = (\alpha_1, \gamma_1, \beta_1)$ and $\tilde{A}_2 = (\alpha_2, \gamma_2, \beta_2)$, the comparison is defined as*

$\tilde{A}_1 < \tilde{A}_2$ *if and only if*

$$SV\left(\tilde{A}_1\right) < SV\left(\tilde{A}_2\right) or$$
$$SV\left(\tilde{A}_1\right) = SV\left(\tilde{A}_2\right) \ and \ AV\left(\tilde{A}_1\right) < AV\left(\tilde{A}_2\right) \quad (7)$$

A seven-level linguistic scale is used to transform qualitative strategy evaluations into T-SFNs as shown in Table 1. The scale consists of extremely low, very low, low, medium, high, very high, and extremely high assessments. Each triplet is designed to satisfy the T-spherical feasibility condition for $t = 2$, while the associated score values preserve the intended linguistic ordering. The medium level is calibrated to produce a score close to zero and therefore serves as the central reference level of the scale.

Table 1: T-spherical fuzzy linguistic scale.

Linguistic Judgments	Notation	T-SFN (α, γ, β)	Feasibility Sum $(t = 2)$	Score Value $(t = 2)$
Extreme low	EL	(0.050, 0.050, 0.950)	0.9075	−0.9025
Very low	VL	(0.200, 0.100, 0.850)	0.7725	−0.6925
Low	L	(0.350, 0.200, 0.750)	0.7250	−0.4800
Medium	M	(0.583, 0.200, 0.548)	0.6802	−0.0004
High	H	(0.750, 0.200, 0.350)	0.7250	0.4000
Very high	VH	(0.850, 0.100, 0.200)	0.7725	0.6725
Extreme high	EH	(0.950, 0.050, 0.050)	0.9075	0.8975

Note: The feasibility sum is calculated as $\alpha^t + \gamma^t + \beta^t$, while the score value is calculated as $\alpha^t - \gamma^t - \beta^t$, with $t = 2$. The scale was calibrated by the authors in accordance with the feasibility condition and the score function of Munir et al. [28].

As shown in Table 1, all feasibility sums are below one, and the score values increase strictly from EL to EH. The scale is applied consistently to all 50 scenario-specific linguistic decision matrices. Because the numerical triplets were calibrated for the scenario-based framework developed in this study, Table 1 is presented as a study-defined linguistic scale constructed in accordance with the T-spherical feasibility condition and the neutrality-aware scoring principle proposed by Munir et al. [28].

2.2 The Proposed Hybrid T-Spherical Fuzzy CRITIC–CoCoSo RankNet Framework

This study proposes a hybrid decision-support framework that integrates T-spherical fuzzy evaluation, objective criteria weighting, compromise-based ranking, pairwise learning-to-rank, and model interpretation for scenario-dependent construction strategy selection. The framework addresses the limitation of static decision models by connecting scenario-specific evaluation with a reusable preference-scoring policy. As illustrated in Fig. 1, the framework comprises four sequential stages. Stage I constructs scenario-specific linguistic decision matrices. Stage II transforms the linguistic evaluations into T-spherical fuzzy numbers and applies the neutrality-aware score function. Stage III derives scenario-specific CRITIC weights and CoCoSo rankings. Stage IV converts the rankings into pairwise preferences, trains RankNet using scenario-group partitions, and interprets the learned scoring policy using SHAP. The detailed procedures are presented in Sections 2.2.1–2.2.3. In summary, the framework does not use RankNet to replace the fuzzy MCDM procedure. Instead, the MCDM component generates structured scenario-specific reference rankings, while RankNet learns an approximate scoring policy that can be evaluated under unseen simulated project conditions.

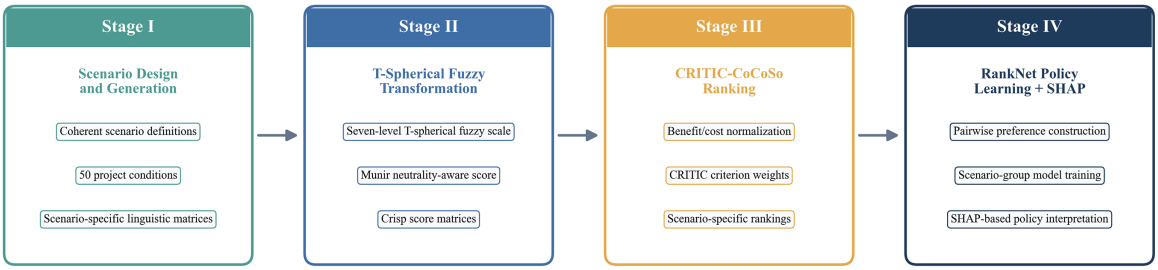

Figure 1: The proposed hybrid T-spherical fuzzy CRITIC–CoCoSo RankNet framework. Stage I generates scenario-specific linguistic decision matrices; Stage II performs T-spherical fuzzy transformation and neutrality-aware scoring; Stage III derives CRITIC weights and CoCoSo rankings; and Stage IV constructs pairwise preferences, trains RankNet using scenario-group partitions, and interprets the learned policy using SHAP.

2.2.1 Stage I: Scenario Design and Generation

Stage I aims to construct a scenario-dependent decision environment for construction strategy selection. Instead of assuming a single fixed decision matrix, this stage generates a set of scenario-specific decision matrices that reflect changes in project conditions. These matrices are later transformed into T-spherical fuzzy evaluations in Stage II. Therefore, Stage I provides the contextual foundation for the subsequent fuzzy MCDM and RankNet-based preference-learning processes.

Step 1. Scenario Design

Let $V = \{V_1, V_2, \ldots, V_k\}$ denote the set of scenario variables, where each variable represents a contextual condition such as labor shortage, weather uncertainty, sustainability requirement, supply instability, budget pressure, and project urgency.

Each scenario s is represented by a scenario vector:

$$\mathrm{v}^{(s)} = \left(v_1^{(s)}, v_2^{(s)}, \ldots, v_k^{(s)}\right), v_l^{(s)} \in [0,1], s = 1, 2, \ldots, N_s \tag{8}$$

where $v_l^{(s)}$ indicates the intensity of the l-th scenario variable in scenario s. A higher value represents a stronger contextual effect. For example, a higher value of labor shortage indicates more severe workforce constraints, while a higher value of project urgency indicates tighter schedule pressure.

In the present study, $N_s = 50$ manually specified and named scenarios are constructed to represent a broad range of project conditions. The scenario set comprises six dominant-constraint scenarios, 15 combined-constraint scenarios, three benchmark environments, and 26 applied construction contexts. The dominant-constraint scenarios isolate individual project pressures, whereas the combined-constraint scenarios represent the simultaneous occurrence of two or more adverse conditions. The benchmark environments represent stable, low-constraint, and balanced project settings, while the applied contexts describe practical situations such as rainy-season construction, supply-chain disruption, disaster response, remote-site operations, and green urban development.

The scenario vectors are specified deliberately rather than generated through unrestricted random sampling. This design ensures consistency among each scenario name, its managerial interpretation, and the corresponding contextual intensities. For example, the Rainy Season Project is assigned a high weather-uncertainty value, the Strict Sustainability scenario has a high sustainability-requirement value, the Critical Urgency scenario has a high project-urgency value, and the Stable Environment has low values across all six contextual variables. The simulation-based design is adopted because real-world construction datasets rarely provide repeated strategy evaluations across a sufficiently broad set of controlled project conditions.

Step 2. Scenario-Based Performance Generation

Let S_i denote the i-th construction strategy $(i = 1, 2, \ldots, m)$, and let C_j denote the j-th evaluation criterion $(j = 1, 2, \ldots, n)$. The scenario-dependent performance magnitude of strategy S_i under criterion C_j in scenario s is generated as:

$$x_{ij}^{(s)} = \text{clip}_{[0.03,0.97]}\left[b_{ij} + \sum_{l=1}^{k} \lambda_{ijl}\left(v_l^{(s)} - v_0\right) + \varepsilon_{ij}^{(s)}\right] \tag{9}$$

where b_{ij} is the baseline criterion magnitude of strategy S_i under criterion C_j; λ_{ijl} is the sensitivity of the specific strategy–criterion combination $\left(S_i, C_j\right)$ to scenario variable V_l; $v_l^{(s)}$ is the intensity of variable V_l in scenario s; $v_0 = 0.5$ is the moderate contextual reference level; and $\varepsilon_{ij}^{(s)}$ is a small pseudo-random Gaussian perturbation sampled from $\varepsilon_{ij}^{(s)} = N(0, 0.012^2)$. A fixed random seed of 2026 was used to ensure that the sampled perturbations and the resulting synthetic dataset were reproducible. The clipping operator restricts the generated magnitude to the interval $[0.03, 0.97]$. The first term in Eq. (9) represents the intrinsic criterion profile of each strategy under moderate project conditions. The second term models the context-dependent response of a particular strategy–criterion combination. Unlike a formulation based on separate criterion-level direct and interaction effects, the coefficient λ_{ijl} is indexed by strategy, criterion, and scenario variable. It therefore captures how the same contextual condition may affect different strategies differently. The Gaussian perturbation introduces a small amount of controlled stochastic variation, while the fixed seed ensures exact reproducibility of the generated values.

The centered expression $v_l^{(s)} - v_0$ represents the deviation of a scenario condition from the moderate reference level. A positive value indicates that the contextual pressure is stronger than the reference condition, whereas a negative value indicates a less constrained environment. The sign and magnitude of λ_{ijl} determine the direction and intensity of the resulting performance adjustment.

The baseline profiles and sensitivity coefficients are calibrated to preserve the intended managerial specialization of the strategies. For example, S1 has the lowest baseline cost magnitude, S2 has the lowest baseline duration magnitude, S3 has the lowest baseline carbon magnitude, S4 combines low risk with high safety, S6 has the lowest labor requirement, S7 has the highest quality, and S8 has a balanced profile. The sensitivity coefficients further allow S2 to become more competitive under urgency, S3 under sustainability pressure, S4 under weather and supply uncertainty, and S6 under labor shortage. The perturbation term follows a Gaussian distribution with standard deviation $\sigma = 0.012$. It is included only to reduce exact ties among otherwise similar evaluations. To ensure reproducibility, the perturbation is generated deterministically using a fixed generation seed of 2026 and a stable identifier constructed from the scenario, strategy, and criterion. Consequently, repeated execution of the generation procedure produces the same input dataset.

The generated numerical value $x_{ij}^{(s)}$ is subsequently mapped into a linguistic evaluation $l_{ij}^{(s)}$ using the predefined linguistic scale $\{EL, VL, L, M, H, VH, EH\}$. This step produces a scenario-specific linguistic decision matrix:

$$L^{(s)} = \left[l_{ij}^{(s)}\right]_{m\times n}, s = 1, 2, \ldots, N_s \tag{10}$$

The matrix $L^{(s)}$, rather than the intermediate numerical matrix $X^{(s)}$, is used as the direct input for the T-spherical fuzzy transformation in Stage II. This clarification ensures consistency between Stage I and the subsequent fuzzy evaluation process.

The scenario-based data generation procedure is summarized in Algorithm 1.

Algorithm 1: Scenario-based data generation

Input:

- Set of construction strategies $S = \{S_1, S_2, \ldots, S_m\}$;
- Set of criteria $C = \{C_1, C_2, \ldots, C_n\}$;
- Set of scenario variables $V = \{V_1, V_2, \ldots, V_k\}$;
- Number of scenarios N_s;
- Baseline performance matrix $B = [b_{ij}]$;
- Strategy–criterion–context sensitivity parameters λ_{jll};
- Moderate contextual reference level $v_0 = 0.5$;
- Noise level $\sigma = 0.012$;
- Fixed generation seed 2026;
- Clipping interval $[0.03, 0.97]$;
- Linguistic scale $\Omega = \{EL, VL, L, M, H, VH, EH\}$.

Output:

- Scenario-specific linguistic decision matrices $L^{(s)} = [l_{ij}^{(s)}], s = 1, 2, \ldots, N_s$.

for $s = 1$ to N_s **do**
  Read the manually specified scenario vector $\mathbf{v}^{(s)} = (v_1^{(s)}, v_2^{(s)}, \ldots, v_k^{(s)})$.
  for each strategy S_i **do**
    for each criterion C_j **do**
      Retrieve the baseline performance value b_{ij}
      Compute the centered context-dependent adjustment: $\Delta_{ij}^{(s)} = \sum_{l=1}^{k} \lambda_{ijl} \left(v_l^{(s)} - 0.5 \right)$
      Generate the deterministic perturbation: $\varepsilon_{ij}^{(s)} \sim N(0, 0.012^2)$
      Compute the unbounded numerical magnitude: $\tilde{x}_{ij}^{(s)} = b_{ij} + \Delta_{ij}^{(s)} + \varepsilon_{ij}^{(s)}$
      Apply the clipping operation: $x_{ij}^{(s)} = \text{clip}_{[0.03,0.97]} \left(\tilde{x}_{ij}^{(s)} \right)$.
      Map $x_{ij}^{(s)}$ into the linguistic term $l_{ij}^{(s)} \in \{EL, VL, L, M, H, VH, EH\}$.
    end for
  end for
  Construct the linguistic decision matrix $L^{(s)} = [l_{ij}^{(s)}]$
end for

The output of Stage I is therefore a collection of scenario-specific linguistic decision matrices:

$$\mathcal{L} = \left\{ L^{(1)}, L^{(2)}, \ldots, L^{(N_s)} \right\} \tag{11}$$

These matrices capture the variability of strategy performance under different contextual conditions and serve as the input for the T-spherical fuzzy transformation in Stage II.

2.2.2 Stages II and III: T-Spherical Fuzzy Evaluation, CRITIC Weighting, and CoCoSo Ranking

Stage II and Stage III transform the scenario-specific linguistic decision matrices generated in Stage I into scenario-specific strategy rankings. Stage II performs T-spherical fuzzy transformation and score-based defuzzification, while Stage III applies CRITIC weighting and CoCoSo ranking. The output of these two

stages is a set of MCDM-derived reference rankings that are later transformed into pairwise preference data for RankNet training.

Step 1. Linguistic Assessment and T-Spherical Fuzzy Transformation

Let $L^{(s)} = [l_{ij}^{(s)}]_{m\times n}$ denote the linguistic decision matrix under scenario s, where $l_{ij}^{(s)}$ represents the linguistic evaluation of strategy S_i with respect to criterion C_j. Each linguistic term is mapped into a T-spherical fuzzy number according to the predefined linguistic scale shown in Table 1.

The T-spherical fuzzy decision matrix under scenario s is denoted as:

$$\tilde{X}^{(s)} = \left[\tilde{x}_{ij}^{(s)}\right]_{m\times n} \tag{12}$$

where

$$\tilde{x}_{ij}^{(s)} = \left(\alpha_{ij}^{(s)}, \gamma_{ij}^{(s)}, \beta_{ij}^{(s)}\right) \tag{13}$$

here, $\alpha_{ij}^{(s)}$, $\gamma_{ij}^{(s)}$, and $\beta_{ij}^{(s)}$ represent the membership, neutrality or abstinence, and non-membership degrees of the linguistic evaluation, respectively.

Step 2. Neutrality-Aware T-Spherical Fuzzy Scoring

To enable numerical computation while retaining the three principal components of each T-SFN, the linguistic fuzzy evaluations are transformed into crisp values using the neutrality-aware score function:

$$d_{ij}^{(s)} = SF\left(\tilde{x}_{ij}^{(s)}\right) = \left(\alpha_{ij}^{(s)}\right)^t - \left(\gamma_{ij}^{(s)}\right)^t - \left(\beta_{ij}^{(s)}\right)^t \tag{14}$$

In this study, $t = 2$. A larger score indicates a more favorable linguistic evaluation. Eq. (14) is used consistently for all scenario-specific matrices and throughout the subsequent CRITIC–CoCoSo and feature-construction procedures:

$$D^{(s)} = \left[d_{ij}^{(s)}\right]_{m\times n} \tag{15}$$

The resulting matrix $D^{(s)}$ is used directly as the crisp decision matrix for normalization, objective weighting, and compromise ranking.

Step 3. Beneficial and Non-Beneficial Normalization

Before CRITIC weighting and CoCoSo ranking, the crisp matrix $D^{(s)}$ is normalized. Because the evaluation criteria include both beneficial and non-beneficial criteria, two normalization equations are used.

For a beneficial criterion C_j, the normalized value is calculated as:

$$z_{ij}^{(s)} = \frac{d_{ij}^{(s)} - \min_i d_{ij}^{(s)}}{\max_i d_{ij}^{(s)} - \min_i d_{ij}^{(s)}} \tag{16}$$

For a non-beneficial criterion C_j, the normalized value is calculated as:

$$z_{ij}^{(s)} = \frac{\max_i d_{ij}^{(s)} - d_{ij}^{(s)}}{\max_i d_{ij}^{(s)} - \min_i d_{ij}^{(s)}} \tag{17}$$

If all strategies have the same score under a criterion, the denominator in Eqs. (16) and (17) becomes zero. In this exceptional case, the normalized values for that criterion are assigned a neutral value of 0.5 for all strategies. The normalized decision matrix under scenario s is denoted as:

$$Z^{(s)} = \left[z_{ij}^{(s)}\right]_{m\times n} \tag{18}$$

In this study, quality, safety, and flexibility are treated as beneficial criteria, whereas cost, duration, carbon, risk, and labor are treated as non-beneficial criteria. This distinction ensures that all normalized values follow the same preference direction, where larger values indicate better performance.

Step 4. CRITIC-Based Objective Weighting

The CRITIC method is applied to determine the objective importance of each criterion under scenario s. First, the standard deviation of criterion C_j is calculated as:

$$\sigma_j^{(s)} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(z_{ij}^{(s)} - \bar{z}_j^{(s)}\right)^2} \tag{19}$$

where $\bar{z}_j^{(s)}$ is the mean normalized value of criterion C_j under scenario s.

The correlation coefficient between criteria C_j and C_k is calculated as:

$$r_{jk}^{(s)} = \frac{\sum_{i=1}^{m}\left(z_{ij}^{(s)} - \bar{z}_j^{(s)}\right)\left(z_{ik}^{(s)} - \bar{z}_k^{(s)}\right)}{\sqrt{\sum_{i=1}^{m}\left(z_{ij}^{(s)} - \bar{z}_j^{(s)}\right)^2 \sum_{i=1}^{m}\left(z_{ik}^{(s)} - \bar{z}_k^{(s)}\right)^2}} \tag{20}$$

The information content of criterion C_j is defined as:

$$I_j^{(s)} = \sigma_j^{(s)} \sum_{k=1}^{n}\left(1 - r_{jk}^{(s)}\right) \tag{21}$$

The notation $I_j^{(s)}$ is used for information content to avoid confusion with the criterion symbol C_j.

The CRITIC weight of criterion C_j is then obtained as:

$$w_j^{(s)} = \frac{I_j^{(s)}}{\sum_{j=1}^{n} I_j^{(s)}} \tag{22}$$

where

$$0 \le w_j^{(s)} \le 1, \sum_{j=1}^{n} w_j^{(s)} = 1 \tag{23}$$

The CRITIC weights are recalculated independently for each scenario. If the total information content is numerically zero, equal weights $w_j^{(s)} = 1/n$ are assigned as a computational safeguard.

Step 5. CoCoSo-Based Ranking

Using the CRITIC weights, the CoCoSo method is applied to calculate the compromise score of each construction strategy under scenario s. The additive measure is calculated as:

$$AS_i^{(s)} = \sum_{j=1}^{n} w_j^{(s)} z_{ij}^{(s)} \tag{24}$$

For numerical stability in the multiplicative aggregation, normalized values equal to zero are replaced by a small positive constant:

$$PS_i^{(s)} = \prod_{j=1}^{n} \left(\hat{z}_{ij}^{(s)}\right)^{w_j^{(s)}} \text{ where } \hat{z}_{ij}^{(s)} = \max\left(z_{ij}^{(s)}, 10^{-9}\right) \tag{25}$$

This adjustment prevents undefined logarithmic operations without materially changing the normalized criterion values.

The final compromise score is calculated as:

$$Q_i^{(s)} = AS_i^{(s)} + PS_i^{(s)} + \frac{AS_i^{(s)} PS_i^{(s)}}{AS_i^{(s)} + PS_i^{(s)}} \tag{26}$$

Strategies are ranked in descending order of $Q_i^{(s)}$. The resulting scenario-specific ranking is denoted as:

$$R^{(s)} = \text{rank}\left(S_1, S_2, \ldots, S_m \mid Q_i^{(s)}\right), \; s = 1, 2, \ldots, N_s \tag{27}$$

The ranking $R^{(s)}$ represents the MCDM-derived preference order of construction strategies under scenario s. These rankings are not treated as external empirical ground truth; rather, they serve as structured reference rankings from which pairwise preference labels are constructed for the subsequent RankNet-based learning-to-rank stage. The T-spherical fuzzy CRITIC-CoCoSo evaluation is summarized in Algorithm 2.

Algorithm 2: T-spherical fuzzy CRITIC–CoCoSo evaluation

Input:

- Scenario-specific linguistic decision matrices $\mathcal{L} = \left\{L^{(1)}, L^{(2)}, \ldots, L^{(N_s)}\right\}$;
- Linguistic scale $\Omega = \{EL, VL, L, M, H, VH, EH\}$;
- T-spherical parameter $t = 2$;
- Neutrality-aware score function $SV(\tilde{A}) = \alpha^t - \gamma^t - \beta^t$;
- Set of beneficial and non-beneficial criteria.

Output:

- Scenario-specific MCDM-derived rankings $\mathcal{R} = \{R^{(1)}, R^{(2)}, \ldots, R^{(N_s)}\}$.

for each scenario $s = 1, 2, \ldots, N_s$ **do**
  Read the linguistic decision matrix $L^{(s)} = [l_{ij}^{(s)}]$
  Map each linguistic element $l_{ij}^{(s)}$ to the T-SFN $\tilde{x}_{ij}^{(s)} = (\alpha_{ij}^{(s)}, \gamma_{ij}^{(s)}, \beta_{ij}^{(s)})$
  Construct the T-spherical fuzzy decision matrix $\tilde{X}^{(s)}$
  Compute the neutrality-aware crisp score matrix $D^{(s)} = [d_{ij}^{(s)}]$ using Eq. (14)

(Continued)

Algorithm 2 (continued)

Normalize $D^{(s)}$ into $Z^{(s)}$ using Eq. (16) for beneficial criteria and Eq. (17) for non-beneficial criteria; assign 0.5 when a criterion has zero range.
Compute the standard deviation $\sigma_j^{(s)}$ of each criterion.
Compute the intercriterion correlations $r_{jk}^{(s)}$ and information contents $I_j^{(s)}$.
Compute and normalize the CRITIC weights $w_j^{(s)}$.
for each strategy S_i **do**
Compute the additive measure $AS_i^{(s)}$
Compute the multiplicative measure $PS_i^{(s)}$ using the numerically adjusted normalized values $\hat{z}_{ij}^{(s)}$.
Compute the compromise score $Q_i^{(s)}$
end for
Rank strategies in descending order of $Q_i^{(s)}$ to obtain $R^{(s)}$
end for

The output of Stage II and Stage III is a collection of scenario-specific MCDM-derived rankings:

$$\mathcal{R} = \left\{R^{(1)}, R^{(2)}, \ldots, R^{(N_s)}\right\} \tag{28}$$

These rankings summarize the preferred construction strategies under different project conditions and are subsequently transformed into pairwise preference data for the RankNet-based neural ranking model in Stage IV.

2.2.3 Stage IV: Pairwise Preference Learning and Decision Policy Modeling

Stage IV transforms the scenario-specific rankings obtained from Stage III into a learnable decision policy using RankNet-based pairwise learning-to-rank. Unlike conventional MCDM approaches that generate rankings only for predefined decision matrices, this stage aims to learn how strategy preferences change across different scenario conditions. Therefore, the output of the CRITIC–CoCoSo stage is not treated as external empirical ground truth, but as a set of structured MCDM-derived reference rankings for training and evaluating the learning model.

Step 1. Feature Construction

For each construction strategy S_i under scenario s, a feature vector is constructed by combining three groups of features:

$$\mathrm{x}_i^{(s)} = \left[\mathrm{a}_i^{(s)}, \mathrm{v}^{(s)}, \mathrm{a}_i^{(s)} \otimes \mathrm{v}^{(s)}\right] \tag{29}$$

where $\mathrm{a}_i^{(s)}$ denotes the eight alternative-level strategy-performance features obtained from the scenario-specific fuzzy score matrix $D^{(s)}$ in Eq. (15), $\mathrm{v}^{(s)}$ is the six-dimensional scenario vector defined in Eq. (8), and $\mathrm{a}_i^{(s)} \otimes \mathrm{v}^{(s)}$ represents the outer-product interaction features between strategy attributes and scenario variables.

In this study, the final feature matrix contains 400 scenario–strategy instances, corresponding to 50 scenarios and 8 strategies. Each instance includes 8 alternative-level features, 6 scenario-level features, and 48 interaction features. Therefore, the complete feature vector contains 62 explanatory variables. The 48 interaction features are obtained by combining each of the eight strategy attributes with each of the six

scenario variables. This design allows the model to learn not only the direct effects of strategy attributes and scenario conditions, but also their conditional interactions.

The inclusion of interaction features is important because scenario variables alone may not distinguish alternatives within the same scenario. For example, all strategies in the same scenario share the same project urgency or budget pressure level. However, the interaction between strategy attributes and scenario variables allows the model to capture how a specific strategy responds to a given contextual condition.

Step 2. Pairwise Preference Construction

For each scenario s, the MCDM-derived ranking $R^{(s)}$ is transformed into pairwise preference labels. Each unordered pair of distinct strategies is included only once. Let

$$\mathcal{P}^{(s)} = \left\{ \{S_i, S_j\} \mid 1 \le i < j \le m \right\} \tag{30}$$

denote the set of unordered strategy pairs under scenario s.

For each unordered pair, one orientation $\left(S_i, S_j\right)$ is retained deterministically. The preference label is defined as:

$$y_{ij}^{(s)} = \begin{cases} 1, & \text{if } S_i \succ S_j \text{ according to } R^{(s)}, \\ 0, & \text{if } S_j \succ S_i \text{ according to } R^{(s)}. \end{cases} \tag{31}$$

where $S_i \succ S_j$ means that strategy S_i is ranked higher than strategy S_j under scenario s.

Because only one orientation is retained for each unordered pair, reverse duplicates such as (S_i, S_j) and (S_j, S_i) are not simultaneously included. With eight strategies, the number of pairwise observations generated under each scenario is: $\binom{8}{2} = 28$. Across 50 scenarios, the complete pairwise dataset therefore contains: $50 \times 28 = 1400$ pairwise observations.

To prevent information leakage, the pairwise observations are partitioned by scenario rather than by individual pair. For each random seed, 35 scenarios are assigned to training, seven scenarios to validation, and eight scenarios to testing. All 28 pairs generated from the same scenario remain in the same subset. Consequently, each split contains 980 training pairs, 196 validation pairs, and 224 test pairs. The test subset therefore represents project scenarios that are not observed during model training.

Step 3. RankNet Scoring Model

Following the pairwise probabilistic learning-to-rank formulation of Burges et al. [35], a neural scoring model is used to map the feature vector of each strategy under each scenario into a scalar preference score:

$$f_i^{(s)} = f(\mathrm{x}_i^{(s)}; \theta) \tag{32}$$

where θ denotes the trainable parameters of the neural network. The preference probability that S_i is preferred over S_j is modeled as:

$$P_{ij}^{(s)} = \frac{1}{1 + \exp\left[-\left(f_i^{(s)} - f_j^{(s)}\right)\right]} \tag{33}$$

A higher value of $f_i^{(s)}$ indicates a stronger preference for strategy S_i under scenario s. Therefore, a complete ranking for a scenario can be obtained by sorting the predicted scores in descending order.

In the computational implementation, RankNet is specified as a feed-forward neural scoring network with one hidden layer. The model receives the 62-dimensional feature vector as input, applies a fully

connected hidden layer containing 16 neurons, a rectified linear unit activation function, and a dropout rate of 0.10, and produces one scalar preference score.

The network is optimized using Adam with a learning rate of 0.003 and a weight-decay coefficient of 1×10^{-4}. The maximum number of training epochs is set to 10,000. Validation accuracy is evaluated every 20 epochs, and early stopping is applied with a patience of 300 epochs and a minimum improvement threshold of 1×10^{-4}. The model parameters associated with the highest validation accuracy are restored for final evaluation.

The experiment is repeated using ten random seeds: 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100. For each seed, the scenario groups are reshuffled before constructing the 35–7–8 training, validation, and test partition. Test observations are not used for parameter updating or checkpoint selection.

Step 4. Pairwise Loss Function

The RankNet model is trained by minimizing the binary cross-entropy loss over all pairwise comparisons:

$$\mathcal{L} = -\frac{1}{\sum_s \mid \mathcal{P}^{(s)} \mid} \sum_s \sum_{(i,j)\in\mathcal{P}^{(s)}} \left[y_{ij}^{(s)} \log P_{ij}^{(s)} + \left(1 - y_{ij}^{(s)}\right) \log\left(1 - P_{ij}^{(s)}\right) \right] \tag{34}$$

This binary cross-entropy objective encourages the model to assign a higher scalar score to the strategy preferred in the corresponding MCDM-derived reference ranking. The implemented loss is averaged over the retained unordered pairs.

After training, the learned decision policy can be expressed as:

$$\pi: \left[\mathbf{a}_i^{(s)}, \mathbf{v}^{(s)}, \mathbf{a}_i^{(s)} \otimes \mathbf{v}^{(s)} \right] \longmapsto f_i^{(s)} \tag{35}$$

The policy π maps the complete strategy–scenario feature vector to a scalar preference score. For a new scenario, the eight strategies are evaluated using the same feature-construction procedure and ranked by sorting their predicted scores in descending order. The pairwise learning and decision policy is summarized in Algorithm 3.

Algorithm 3: Pairwise learning and decision policy modeling

Input:

- Scenario-specific MCDM-derived rankings $\mathcal{R} = \{R^{(1)}, R^{(2)}, \ldots, R^{(N_s)}\}$;
- Feature vectors $\mathrm{x}_i^{(s)} = [\mathrm{a}_i^{(s)}, \mathrm{v}^{(s)}, \mathrm{a}_i^{(s)} \otimes \mathrm{v}^{(s)}]$;
- Scenario identifiers associated with all feature vectors;
- RankNet hyperparameters, including maximum epochs, learning rate, hidden dimension, dropout rate, weight decay, minimum improvement threshold, and early-stopping patience.

Output:

- Trained RankNet scoring model $f(\cdot;\theta)$;
- Predicted strategy rankings for unseen test scenarios.

Construct feature vectors $\mathbf{x}_i^{(s)}$ for all strategies and scenarios
for each scenario $s = 1, 2, \ldots, N_s$ **do**
    Generate the 28 unordered strategy pairs $\mathcal{P}^{(s)} = \left\{ \{S_i, S_j\} \mid i < j \right\}$.
    Retain one deterministic orientation $\left(S_i, S_j\right)$ for each unordered pair.
    Assign the pairwise label $y_{ij}^{(s)}$ using Eq. (31).
end for

(Continued)

Algorithm 3 (continued)

Randomly divide the 50 scenarios into 35 training, seven validation, and eight test scenario groups using the specified random seed.
Assign all pairwise observations from each scenario to the corresponding scenario-level subset.
Initialize the RankNet parameters $\boldsymbol{\theta}$.
for each training epoch **do**
  Compute the scores $f_i^{(s)}$ and $f_j^{(s)}$ for each pair
  Compute the preference logits: $f_i^{(s)} - f_j^{(s)}$.
  Compute the preference probabilities using Eq. (33).
  Compute the pairwise loss using Eq. (34).
  Update θ using backpropagation and Adam optimization.
  Evaluate validation pairwise accuracy every 20 epochs.
  Save the model parameters when validation accuracy improves by at least 10^{-4}.
  Terminate training when validation accuracy does not improve for 300 epochs.
end for
Restore the checkpoint with the highest validation accuracy.
Evaluate pairwise accuracy and ranking agreement on the eight unseen test scenarios.
Return the trained model $f(\cdot;\theta)$

To evaluate the learning performance of RankNet relative to conventional classifiers, logistic regression [40], random forest [41], gradient boosting [42], and a multilayer perceptron classifier [43] were trained using the same scenario-group partitions and pairwise labels. The four baseline models were implemented using scikit-learn [44], whereas RankNet followed the pairwise probabilistic formulation of Burges et al. [35]. The full 62-feature setting was used for the comparison of all five models. A separate feature-ablation analysis was conducted using logistic regression because its linear structure provides a transparent assessment of the information contributed by each feature group. Three settings are examined: (i) eight alternative-level features only, (ii) alternative-level plus six direct scenario features, and (iii) the complete feature set including 48 interaction features. This design evaluates whether scenario information improves preference learning directly or primarily through its interactions with strategy attributes.

Step 5. Explainability Using SHAP

To interpret the learned decision policy, SHAP values [38] are computed and analyzed for the trained RankNet model. For a prediction function f, the SHAP value of feature k is defined as:

$$\phi_k = \sum_{S \subseteq F \setminus \{k\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f(S \cup \{k\}) - f(S)] \tag{36}$$

where F is the set of all input features, and S denotes a subset of features excluding feature k.

SHAP values quantify the marginal contribution of each feature to the scalar preference score predicted by RankNet. In this study, SHAP was used to identify influential alternative-level attributes, direct scenario variables, and alternative–scenario interaction features. The SHAP analysis was conducted using the Seed 100 RankNet model. All 64 scenario–strategy instances belonging to its eight unseen test scenarios are explained. A representative background dataset containing 70 observations is constructed by selecting two strategy instances from each of the 35 training scenarios. SHAP DeepExplainer was used to calculate the feature attributions by combining SHAP principles with neural-network attribution mechanisms related to DeepLIFT [45]. Global importance was calculated as the mean absolute SHAP value across the 64 test

instances. A SHAP summary analysis was used to examine both attribution magnitude and direction, while a local waterfall explanation was used to decompose an individual strategy score into positive and negative feature contributions. It should be noted that SHAP values are interpreted as model-attribution measures rather than causal effects. They indicate how strongly each input contributes to the trained RankNet output but do not demonstrate that the feature causally determines construction-project performance.

The final output of Stage IV is a trained and interpretable decision policy model that can rank construction strategies under new or unseen project scenarios and explain the feature-level drivers behind the predicted ranking outcomes.

3 Results

3.1 Experimental Design and Data Description

The experimental dataset comprises 50 manually specified construction scenarios, eight strategies, eight evaluation criteria, and six contextual variables. The contextual variables are budget pressure, labor shortage, weather uncertainty, supply instability, sustainability requirement, and project urgency. Strategy performance was allowed to vary across scenarios according to the generation procedure described in Section 2.2.1. The resulting dataset contains 400 scenario–strategy instances. Each learning instance comprises eight alternative-level features, six scenario variables, and 48 alternative–scenario interactions, producing 62 explanatory features. The 50 MCDM-derived rankings were transformed into 1400 non-duplicated pairwise observations. In each repeated experiment, all observations from the same scenario remained within one training, validation, or test partition.

3.1.1 Scenario Variable Configuration

To represent realistic project environments, six contextual variables are defined: budget pressure, labor shortage, weather uncertainty, supply instability, sustainability requirement, and project urgency, as summarized in Table 2. These variables were selected because they represent common sources of uncertainty and constraint in construction project delivery. Budget pressure reflects financial limitation and cost sensitivity; labor shortage captures workforce availability constraints; weather uncertainty represents environmental disruption risk; supply instability reflects material and logistics uncertainty; sustainability requirement captures environmental and regulatory pressure; and project urgency represents schedule compression and deadline intensity.

Each variable was normalized to the range $[0,1]$, enabling consistent scaling across scenarios and facilitating integration into the scenario-based performance generation model. A higher value indicates a stronger contextual effect. For example, a high value of labor shortage indicates severe workforce constraints, while a high value of project urgency indicates tighter schedule pressure. Modeling these variables as continuous rather than categorical allows the framework to represent gradual changes in project conditions and avoids oversimplifying complex construction environments into discrete scenario classes.

Beyond their individual effects, the contextual variables are designed to influence each strategy–criterion combination differently. This setting reflects the practical observation that contextual impacts are conditional rather than uniform. For instance, labor shortage affects labor-intensive strategies more strongly than labor-saving strategies, whereas project urgency changes the comparative attractiveness of strategies with different duration and flexibility profiles. These conditional responses are represented in the scenario-generation model through strategy–criterion–context sensitivity coefficients. During feature construction, the resulting alternative attributes are further combined with the scenario variables to form 48 alternative–scenario interaction features.

Table 2: Scenario variables and descriptions.

Variable	Description	Operational Meaning (High Value)
Labor Shortage	Workforce availability constraint	Severe lack of skilled labor
Weather Uncertainty	Environmental variability	High probability of weather disruptions
Sustainability Requirement	Environmental/regulatory pressure	Strict sustainability targets/compliance
Supply Instability	Reliability of supply chain	Frequent material/logistics disruptions
Budget Pressure	Financial constraint intensity	Tight budget limits, cost sensitivity
Project Urgency	Schedule constraint intensity	Very tight deadlines

The 50 scenarios were manually specified to provide broad and managerially coherent coverage of the decision environment. They comprise six dominant-constraint scenarios, 15 combined-constraint scenarios, three benchmark environments, and 26 applied construction contexts. The scenario names and numerical intensities were aligned deliberately. For example, Tight Budget has very high budget pressure, Rainy Season Project has very high weather uncertainty, Green Regulation Strict has very high sustainability requirements, Critical Urgency has very high project urgency, and Stable Environment has low values across all six variables. Fig. 2 presents the six contextual variables across the 50 project scenarios. The heatmap shows single-constraint cases, combined pressures, relatively stable environments, and complex applied settings involving several simultaneous constraints. This structured diversity provides a broad basis for evaluating whether the learned preference policy can transfer across different construction conditions.

3.1.2 Strategy and Criteria Structure

The decision space comprises eight construction strategies designed to represent different managerial orientations in project delivery. These include S1 Cost Minimization, S2 Fast Track, S3 Sustainability, S4 Risk Averse, S5 Tech Intensive, S6 Labor Saving, S7 Quality Priority, and S8 Balanced. The inclusion of these strategies allows the decision model to represent different trade-offs among cost, time, quality, environmental performance, safety, risk exposure, labor requirement, and flexibility. The baseline profiles were calibrated to preserve the intended managerial meaning of each strategy. S1 was assigned the strongest baseline cost performance, S2 the strongest duration performance, S3 the lowest carbon burden, S4 the strongest risk and safety profile, S6 the lowest labor requirement, and S7 the highest quality. S8 was specified as a balanced strategy without a single extreme advantage, whereas S5 represented a broadly competitive technology-intensive configuration. Context-sensitivity coefficients were subsequently used to adjust these profiles under different scenarios.

Figure 2: Distribution of six contextual variables across 50 manually specified construction scenarios. The heatmap includes dominant constraints, combined constraints, benchmark environments, and applied project contexts ranging from relatively stable to highly constrained conditions.

Each strategy was evaluated using eight criteria, namely cost, duration, quality, carbon, safety, risk, labor, and flexibility, as shown in Table 3. These criteria reflect both conventional project performance dimensions and emerging requirements in modern construction management. Cost and duration represent economic and time-related performance; quality and safety capture technical and operational reliability; carbon reflects

environmental impact; risk represents exposure to uncertainty and disruption; labor captures workforce requirement; and flexibility reflects the ability of a strategy to adapt to changing conditions.

Table 3: Evaluation criteria, operational descriptions, and preference types.

Criterion	Description	Type
Cost	Total project cost or cost burden	Non-beneficial
Duration	Project completion time or schedule requirement	Non-beneficial
Quality	Expected output quality level	Beneficial
Carbon	Carbon emission or environmental burden	Non-beneficial
Safety	Safety performance and accident prevention capability	Beneficial
Risk	Exposure to uncertainty and disruption	Non-beneficial
Labor	Labor requirement or workforce intensity	Non-beneficial
Flexibility	Adaptability to changes in project conditions	Beneficial

The distinction between beneficial and non-beneficial criteria is essential for the normalization process in the CRITIC–CoCoSo procedure. In this study, quality, safety, and flexibility are treated as beneficial criteria because higher values are preferred. In contrast, cost, duration, carbon, risk, and labor are treated as non-beneficial criteria because lower values indicate more desirable performance. This classification ensures that all normalized values follow the same preference direction before objective weighting and compromise ranking are performed.

Strategy performance was treated as context-dependent rather than fixed. For example, duration-related performance becomes particularly relevant under urgent conditions, labor-related performance differentiates strategies under workforce shortages, and carbon-related performance becomes more influential under stringent sustainability requirements. These changes are first represented through the strategy–criterion–context sensitivities used in scenario generation. The learning dataset then includes explicit interaction features between the eight T-spherical fuzzy strategy scores and the six scenario variables, allowing RankNet to learn how project conditions modify alternative-specific preferences.

3.1.3 Scenario-Based Data Generation

For each scenario s, the performance of strategy S_i under criterion C_j was generated using the scenario-based performance generation mechanism described in Eq. (9). The numerical criterion magnitudes generated by Eq. (9) are mapped to the seven-level linguistic scale and subsequently transformed into T-spherical fuzzy scores. Across 50 scenarios and eight strategies, the resulting dataset contains 400 scenario–strategy instances.

For RankNet learning, each instance contains eight alternative-level features obtained from the scenario-specific T-spherical fuzzy score matrix, six scenario-level variables, and 48 outer-product interaction features. The final feature vector therefore contains 62 explanatory variables. The interaction features are important because direct scenario variables are shared by all strategies within the same scenario and do not, by themselves, distinguish one alternative from another. Their interactions with strategy attributes convert common contextual conditions into alternative-specific information.

For visualization, Fig. 3 compares the scenario-wise normalized CoCoSo compromise scores of the eight strategies under three representative environments: Tight Budget, Critical Urgency, and Stable Environment.

Unlike an unweighted aggregation of criterion values, the displayed measure incorporates the scenario-specific T-spherical fuzzy scores, CRITIC weights, and CoCoSo compromise calculation. The CoCoSo scores are normalized within each selected scenario to facilitate visual comparison. The results show clear and managerially coherent preference shifts. S1 Cost Minimization obtains the highest normalized CoCoSo score under Tight Budget, reflecting its cost-oriented baseline profile. S2 Fast Track becomes the preferred strategy under Critical Urgency because of its comparative duration advantage and urgency-related contextual response. Under Stable Environment, S8 Balanced receives the highest score, indicating that a non-specialized compromise strategy becomes attractive when no single project constraint dominates. These shifts provide a face-validity check for the corrected scenario and strategy definitions. They also demonstrate that no single strategy is uniformly preferred across all project conditions. This scenario dependency motivates the subsequent construction of scenario-specific CRITIC–CoCoSo rankings and the learning of a transferable preference policy through RankNet.

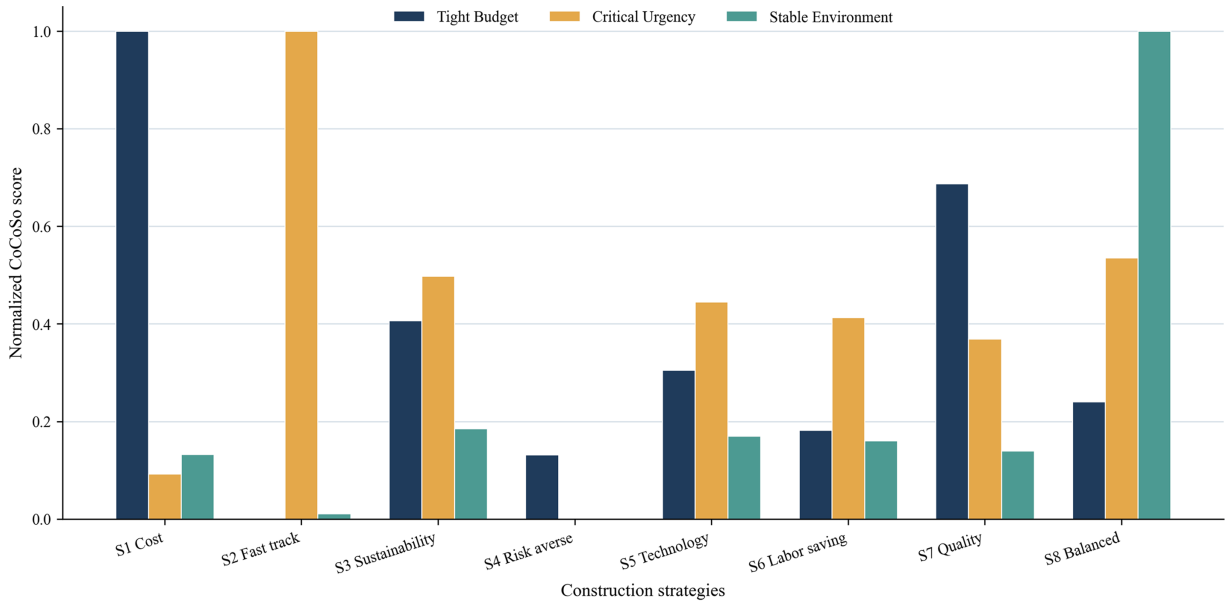

Figure 3: Scenario-dependent changes in normalized CoCoSo compromise scores under Tight Budget, Critical Urgency, and Stable Environment. S1 Cost Minimization, S2 Fast Track, and S8 Balanced become the preferred strategies under their corresponding managerial contexts.

3.2 T-Spherical Fuzzy CRITIC–CoCoSo Results

Building upon the scenario-based dataset described in Section 3.1, the generated scenario-specific linguistic matrices were processed using the T-spherical fuzzy CRITIC–CoCoSo procedure to derive MCDM-based evaluations of construction strategies. This stage transforms linguistic scenario-dependent assessments into structured decision outputs by incorporating T-spherical fuzzy representation, neutrality-aware numerical scoring, beneficial/non-beneficial normalization, objective criteria weighting, and compromise ranking. The resulting rankings provide the structured reference preferences from which the subsequent RankNet-based learning model learns scenario-dependent decision policies.

3.2.1 Linguistic and T-Spherical Fuzzy Evaluation

The scenario-specific criterion magnitudes generated in Stage I were first mapped to the seven linguistic levels defined in Table 1 and subsequently represented as T-spherical fuzzy numbers. This transformation

retains membership, neutrality or abstinence, and non-membership information before numerical weighting and ranking are performed. Table 4 presents an illustrative linguistic decision matrix for the Rainy Season Project scenario. The example shows that each strategy exhibits a distinct profile across cost, duration, quality, carbon, safety, risk, labor, and flexibility.

Table 4: Linguistic evaluation matrix for the Rainy Season Project scenario.

Strategy	Cost	Duration	Quality	Carbon	Safety	Risk	Labor	Flexibility
S1: Cost Minimization	VL	H	M	M	M	H	M	M
S2: Fast Track	M	VL	M	M	M	H	M	H
S3: Sustainability	H	H	H	VL	M	H	M	M
S4: Risk Averse	M	M	H	M	EH	EL	M	H
S5: Tech Intensive	VH	M	H	M	M	VH	L	VH
S6: Labor Saving	H	M	M	M	M	H	VL	H
S7: Quality Priority	H	H	EH	M	VH	M	H	M
S8: Balanced	M	H	H	M	M	H	M	H

Note: Source: Dataset generated by the authors.

Each linguistic term was converted into a T-spherical fuzzy number using the study-defined scale reported in Table 1. The fuzzy evaluations were then transformed into crisp values using the neutrality-aware score function in Eq. (4). The resulting scenario-specific score matrices were used directly as the inputs for beneficial/non-beneficial normalization, CRITIC weighting, and CoCoSo ranking.

Table 4 also illustrates the managerial coherence of the strategy profiles. S2 Fast Track receives a very low duration assessment, S3 Sustainability receives a very low carbon assessment, S4 Risk Averse combines extremely high safety with extremely low risk, S6 Labor Saving receives a very low labor requirement, and S7 Quality Priority receives an extremely high assessment on the quality criterion.

3.2.2 CRITIC Weighting Results

The CRITIC method was applied to each normalized scenario-specific decision matrix to determine objective criteria weights. Unlike subjective weighting methods, CRITIC derives weights from the contrast intensity of each criterion and the conflict among criteria. Therefore, a criterion receives a higher weight when it exhibits stronger variation across strategies and provides non-redundant information relative to other criteria.

Table 5 reports the mean CRITIC weights computed across the 50 scenario-specific decision matrices. Cost receives the highest average weight at 0.1326, followed by carbon at 0.1286, labor at 0.1256, quality at 0.1253, flexibility at 0.1237, risk at 0.1235, duration at 0.1225, and safety at 0.1182. The differences among the mean weights are relatively small, indicating that no single criterion dominates the complete scenario set. Instead, the information contribution of each criterion changes according to the contrast and correlation structure of the corresponding scenario-specific matrix. Risk exhibits the greatest cross-scenario variation, ranging from 0.0860 to 0.2181, which indicates that its objective importance increases substantially under some uncertainty-intensive project conditions.

It should be emphasized that Table 5 reports the average weighting structure across scenarios. Therefore, the mean values should not be interpreted as universal subjective priorities. They summarize the average contrast and non-redundant information contained in each criterion across the 50 scenarios. This scenario-specific weighting structure allows criteria such as risk, cost, carbon, or labor to become more informative under the project conditions in which they differentiate the strategies most clearly.

Table 5: Average CRITIC weights across 50 scenarios.

Criterion	Mean Weight	Criterion	Mean Weight
Cost	0.1326	Flexibility	0.1237
Carbon	0.1286	Risk	0.1235
Labor	0.1256	Duration	0.1225
Quality	0.1253	Safety	0.1182

3.2.3 CoCoSo Ranking and Scenario-Dependent Patterns

Using the scenario-specific CRITIC weights, the CoCoSo method was applied to compute compromise scores and rankings for the eight construction strategies under each scenario. Fig. 4 presents the scenario–strategy ranking heatmap across all 50 scenarios, while Fig. 5 summarizes the distribution of ranking positions for each strategy. Together, these figures show that construction strategy preference is strongly scenario-dependent rather than fixed across all project conditions.

The rankings reveal a distinction between scenario-specific dominance and cross-scenario consistency. S4 Risk Averse ranks first in 20 of the 50 scenarios, the highest first-place frequency among all strategies. It is particularly strong under extreme weather, supply disruption, disaster-response, climate-stress, logistics-constraint, and other uncertainty-intensive conditions. However, S4 also has the largest ranking variability, with a mean rank of 3.98 and a standard deviation of 3.01. It should therefore be interpreted as a specialized strategy that performs exceptionally well under severe uncertainty but is not uniformly competitive across all project environments. S5 Tech Intensive has the best mean rank at 3.08 and the lowest rank standard deviation at 0.92. It appears among the top three strategies in 36 scenarios, although it does not rank first in any scenario. This pattern indicates that S5 is the most consistently competitive strategy rather than the most frequently optimal strategy. S3 Sustainability ranks first in 11 scenarios and has a mean rank of 3.88. Its strongest results occur under stringent sustainability, environmental-regulation, green-city, and compliance-oriented conditions. S8 Balanced ranks first in seven scenarios, has a mean rank of 3.74, and performs strongly in stable or mixed project environments. S6 Labor Saving also ranks first in seven scenarios, particularly in cases characterized by severe labor shortages or low workforce availability. S2 Fast Track ranks first in four scenarios and has a mean rank of 5.60. Its first-place results occur primarily under critical urgency or combined urgency conditions, confirming its specialized schedule-oriented role. S1 Cost Minimization ranks first only in the pure Tight Budget scenario and has a mean rank of 7.04. This suggests that cost minimization becomes preferable when financial pressure is the dominant constraint but is less competitive when cost must be balanced against risk, sustainability, safety, labor, and flexibility. S7 Quality Priority does not rank first in any scenario but maintains a moderate mean rank of 4.62 and appears in the top three in nine scenarios. Its ranking distribution is less variable than those of the strongly specialized strategies, indicating moderate but not dominant performance across the scenario set. The CoCoSo rankings are summarized in Table 6.

The T-spherical fuzzy CRITIC–CoCoSo results demonstrate that criterion importance and strategy preference both vary across project scenarios. The findings also distinguish between strategies that are consistently competitive and those that are highly effective only under specific constraints. S5 provides the strongest cross-scenario consistency, whereas S4 exhibits the most pronounced scenario specialization. However, the MCDM procedure still produces rankings independently for each scenario and does not itself provide a transferable scoring policy. This limitation motivates the subsequent RankNet stage, which learns preference patterns from the complete set of scenario-specific reference rankings and is evaluated on unseen project scenarios.

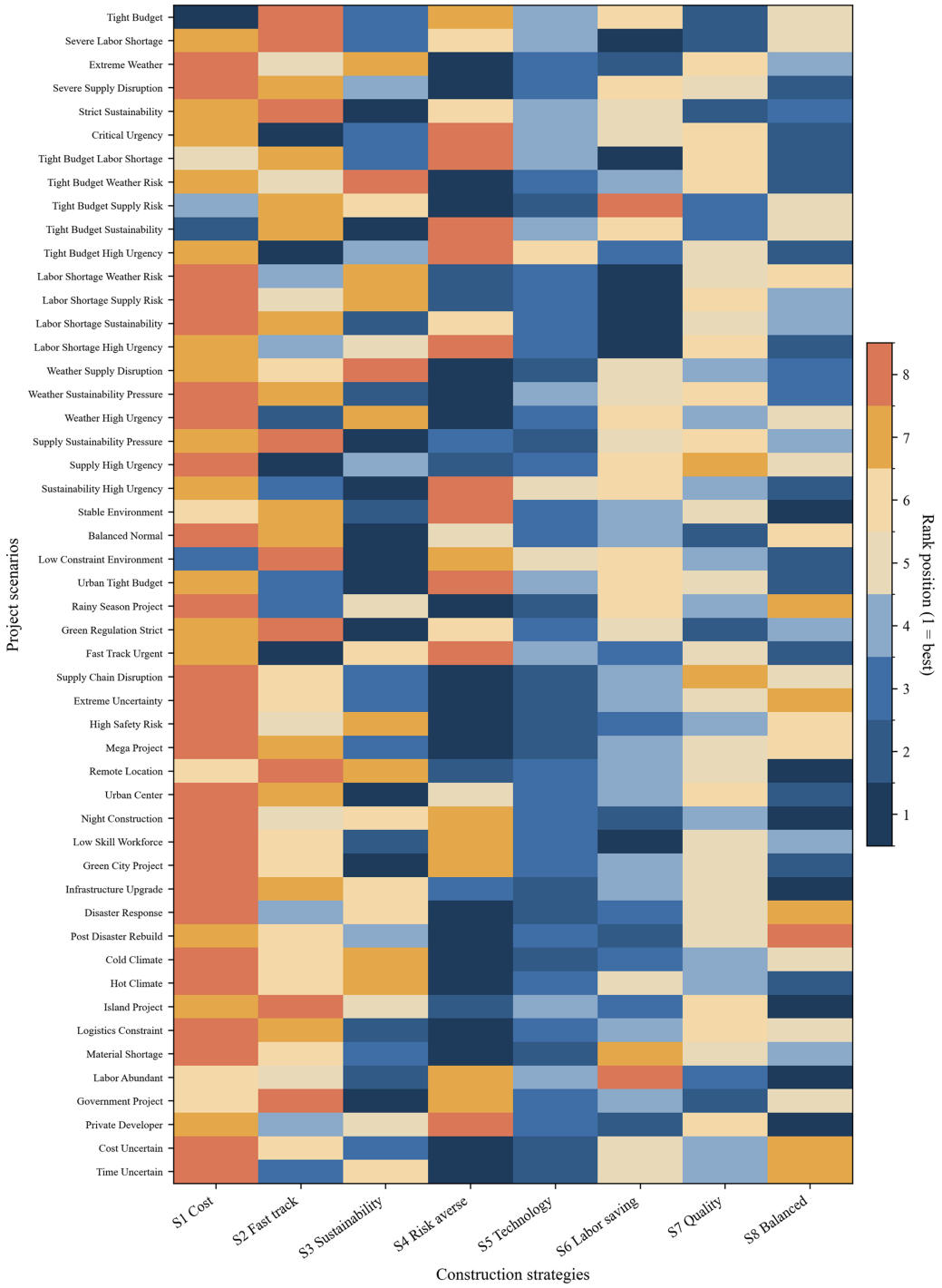

Figure 4: Scenario–strategy ranking heatmap across 50 construction scenarios. Lower numerical ranks indicate more favorable strategy positions, and the variation across rows demonstrates the scenario-dependent nature of the CoCoSo rankings.

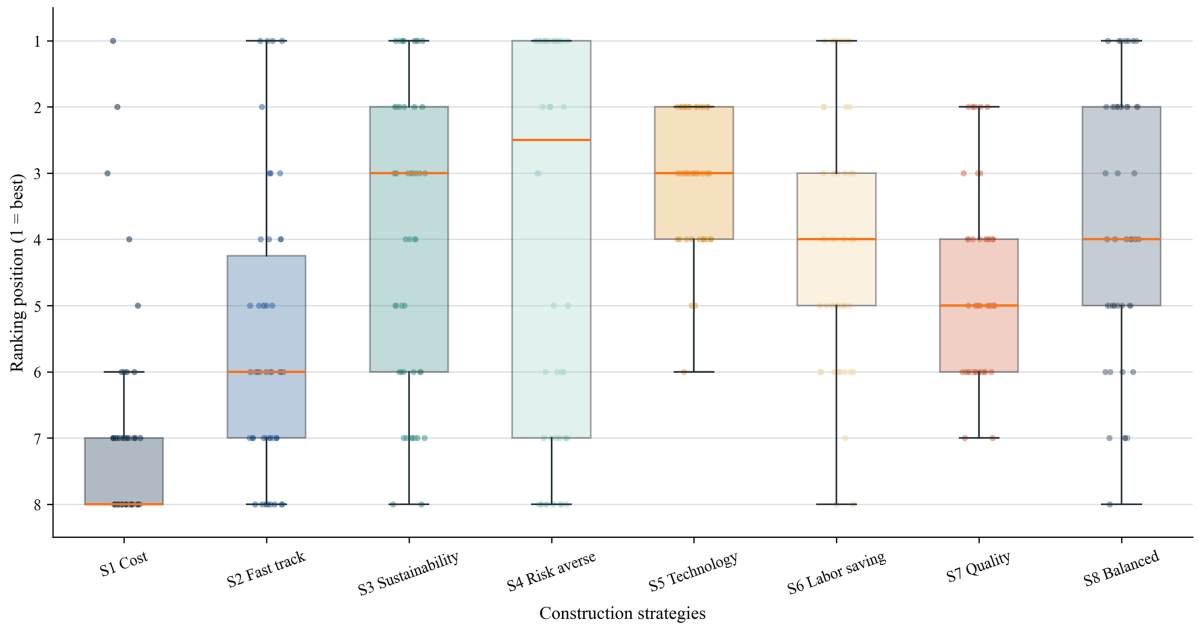

Figure 5: Distribution of CoCoSo ranking positions for the eight construction strategies. Narrow distributions indicate consistent cross-scenario performance, whereas wide distributions indicate stronger specialization and context sensitivity.

Table 6: Summary of CoCoSo ranking patterns across 50 scenarios.

Strategy	Mean Rank	Rank SD	First-Place Frequency	Top-Three Frequency
S1 Cost Minimization	7.04	1.56	1	3
S2 Fast Track	5.6	2.09	4	9
S3 Sustainability	3.88	2.32	11	26
S4 Risk Averse	3.98	3.01	20	27
S5 Tech Intensive	3.08	0.92	0	36
S6 Labor Saving	4.06	1.87	7	17
S7 Quality Priority	4.62	1.35	0	9
S8 Balanced	3.74	2.03	7	23

3.3 AI-Based Decision Policy Learning

This section evaluates the ability of the proposed RankNet model to learn scenario-dependent preference patterns derived from the T-spherical fuzzy CRITIC–CoCoSo stage. The objective of this stage is not to replace the MCDM model, but to learn a transferable decision policy from the set of scenario-specific MCDM-derived rankings. The evaluation focuses on pairwise prediction accuracy, ranking consistency, local perturbation robustness, training stability across multiple random seeds, and generalization to unseen scenarios.

The RankNet model was evaluated across ten random seeds: 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100. For each seed, the 50 scenarios were randomly divided into 35 training scenarios, seven validation scenarios, and eight test scenarios. All 28 pairwise observations generated from the same scenario were retained in the same subset. The resulting partitions contained 980 training pairs, 196 validation pairs, and 224 test pairs.

This scenario-group holdout design prevented pairwise observations from the same project context from appearing in more than one subset.

RankNet was implemented as a feed-forward neural scoring model with one hidden layer containing 16 neurons. The model used a rectified linear unit activation function, a dropout rate of 0.10, Adam optimization, a learning rate of 0.003, and a weight-decay coefficient of 1×10^{-4}. The maximum number of epochs was set to 10,000. Validation accuracy was evaluated every 20 epochs, and early stopping was applied using a patience value of 300 epochs and a minimum improvement threshold of 1×10^{-4}. The checkpoint with the highest validation accuracy was restored for final evaluation.

Gaussian perturbation with a standard deviation of 0.03 was applied to the test features to assess local prediction stability. Because both the scenario partition and the model initialization varied across seeds, the multi-seed results reflect variation arising from unseen-scenario composition as well as neural-network initialization.

Fig. 6 presents the learning dynamics of the representative Seed 100 run. Fig. 6a shows the training and validation pairwise accuracies, while Fig. 6b presents the training-loss trajectory. Test accuracy is intentionally excluded from the training curves because test observations were not used for model selection or checkpoint monitoring. For Seed 100, validation performance reached its highest recorded level at epoch 60. Training accuracy and validation accuracy increased rapidly during the initial epochs, indicating that the model learned the principal pairwise preference structure early in the training process. After the selected epoch, training loss continued to decline, whereas validation accuracy showed no sustained improvement. This divergence indicates mild overfitting and supports the use of validation-based early stopping.

Table 7 summarizes the RankNet results across the ten repeated scenario-group holdout experiments. The model achieved a mean training accuracy of 0.9262 ± 0.0215, a mean validation accuracy of 0.9173 ± 0.0208, and a mean test accuracy of 0.9018 ± 0.0234. Test accuracy ranged from 0.8705 to 0.9420 across the ten seeds. The difference between mean training and test accuracy was approximately 0.0244, indicating a moderate but not excessive reduction in performance when the model was applied to unseen scenarios. The lower results relative to a random pair-level split are expected because the test subset contains complete project contexts that are absent from training rather than additional pairs sampled from previously observed scenarios. Ranking agreement was evaluated by comparing the RankNet-predicted strategy order with the CRITIC–CoCoSo reference ranking in each unseen test scenario. The mean test Spearman correlation was 0.8997 ± 0.0308, with values ranging from 0.8601 to 0.9494. These findings indicate that RankNet recovered most pairwise preferences and preserved the overall ranking structure under unseen project conditions. However, the agreement was not perfect, showing that the model learned an approximate transferable preference policy rather than trivially reproducing the reference rankings.

Local robustness was assessed by adding Gaussian noise with a standard deviation of 0.03 to the test features and calculating the Spearman correlation between the original and perturbed RankNet scores. The mean local perturbation correlation was 0.9966 ± 0.0020, ranging from 0.9931 to 0.9985. This result indicates that small changes in the feature values produced only minor changes in the learned preference scores. This analysis evaluates local prediction stability rather than external robustness. It does not establish that the model will remain stable under major distributional shifts, alternative scenario-generation mechanisms, or empirical construction data. Accordingly, broader robustness should be examined in future studies using real project cases and structurally different scenario sets.

The computational requirements of the RankNet stage were limited. The mean training time was 0.8771 ± 0.4170 s, with individual runs ranging from 0.3443 to 1.6104 s. The selected epochs ranged from 60 to 620, with a mean of 230 epochs. These results indicate that the neural ranking stage can be retrained rapidly

when the scenario set or preference observations are updated. All analyses were implemented in Python 3.13.5 using PyTorch 2.11.0 [46], scikit-learn [44], pandas, NumPy, and SHAP under a 64-bit Windows 11 operating system. The experiments were executed on an Intel Core i5-14600KF CPU with 16 GB RAM. Although the computer was equipped with an NVIDIA GeForce RTX 4070 graphics card, the installed PyTorch environment was CPU-only, CUDA was unavailable, and no GPU acceleration was used. The reported computational times therefore represent CPU execution.

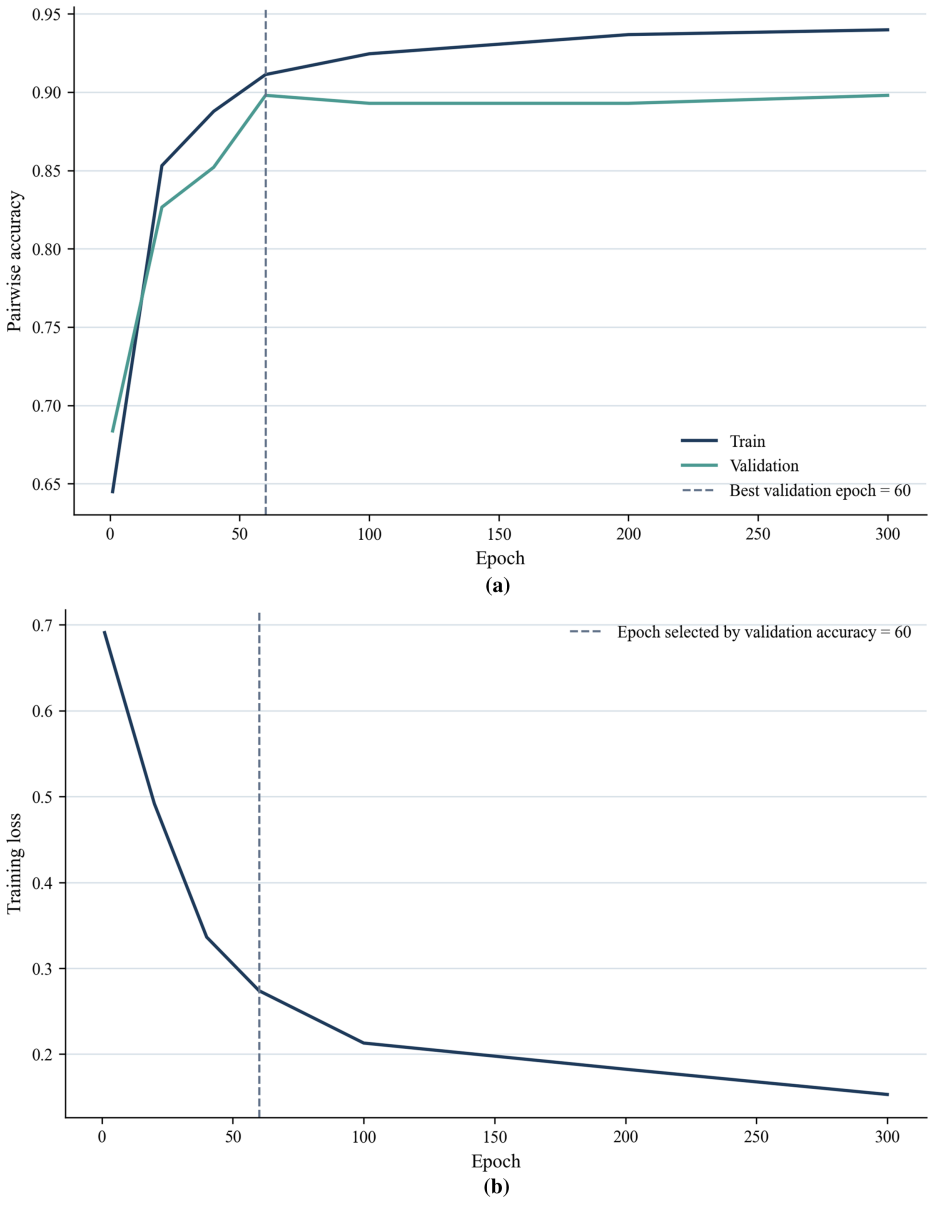

Figure 6: Training dynamics of the RankNet model for Seed 100: (**a**) training and validation pairwise accuracy, with the selected checkpoint indicated at epoch 60; and (**b**) training loss across epochs. Test performance was evaluated only after restoring the best validation checkpoint and is therefore not displayed in the training trajectory.

Table 7: RankNet performance across ten repeated scenario-group holdout experiments.

Metric	Mean	Std	Min	Max
Best epoch	230.0	215.3	60.0	620.0
Final training accuracy	0.9262	0.0215	0.8949	0.9551
Final validation accuracy	0.9173	0.0208	0.8980	0.9592
Final test accuracy	0.9018	0.0234	0.8705	0.9420
Test Spearman correlation: RankNet vs. MCDM	0.8997	0.0308	0.8601	0.9494
Test local perturbation robustness	0.9966	0.0020	0.9931	0.9985
Training time (s)	0.8771	0.4170	0.3443	1.6104

In summary, the scenario-group evaluation provides a more demanding and credible assessment than random pair-level splitting. The results show that RankNet maintains approximately 90% pairwise accuracy and a Spearman ranking agreement of approximately 0.90 when applied to complete project scenarios excluded from training. Together with the high local perturbation correlation and short CPU training time, these findings support the feasibility of learning a transferable and computationally efficient preference policy from scenario-specific MCDM rankings.

3.4 Comparative and Ablation Analysis

This section evaluates whether the RankNet-based learning stage provides additional value beyond reproducing the T-spherical fuzzy CRITIC–CoCoSo reference rankings. High agreement between RankNet and the MCDM-derived rankings indicates that the model has learned the underlying preference structure, but it does not by itself demonstrate that a neural pairwise ranking architecture is superior to simpler learning methods. Therefore, RankNet was compared with several conventional classifiers under the same repeated scenario-group holdout design. A separate feature-ablation analysis was also conducted to examine how direct scenario variables and alternative–scenario interaction features contribute to preference learning.

Four baseline models were considered: logistic regression, random forest, gradient boosting, and a multilayer perceptron classifier. All models were trained and evaluated using the same scenario-level partitions as RankNet, ensuring that the test observations represented scenarios not included in model training. Logistic regression was further evaluated under three feature settings: (i) eight alternative-level features only (ALT only), (ii) alternative-level and six scenario-level features (ALT + SCN), and (iii) the complete feature set containing alternative-level, scenario-level, and 48 interaction features (ALT + SCN + INT). The remaining baseline models and RankNet were evaluated using the complete 62-feature setting.

As shown in Table 8, logistic regression obtained the highest mean test accuracy under the complete feature setting, reaching 0.9027, followed closely by RankNet at 0.9018 and the multilayer perceptron classifier at 0.8978. Their corresponding Spearman correlations were 0.9024, 0.8997, and 0.8967, respectively. These differences are small, indicating that RankNet achieved performance comparable to the strongest linear and neural classification baselines rather than clearly outperforming every competing model. RankNet nevertheless substantially outperformed the two examined tree-based ensemble models in ranking recovery. Random forest achieved a mean test accuracy of 0.8696 and a Spearman correlation of 0.7922, whereas gradient boosting achieved values of 0.8531 and 0.7751, respectively. The larger reduction in Spearman correlation indicates that these models recovered individual pairwise labels more effectively than they preserved the complete strategy-ranking structure within unseen scenarios. The local perturbation correlations were high for RankNet, logistic regression, and the multilayer perceptron classifier, reaching 0.9966, 0.9964, and 0.9967,

respectively. In contrast, random forest and gradient boosting produced lower robustness correlations of 0.9622 and 0.9600. Thus, RankNet demonstrated strong local stability, although it did not uniquely dominate this metric.

Table 8: Comparative and ablation results under repeated scenario-group holdout evaluation.

Model	Feature Setting	Test Accuracy	Spearman Correlation	Robustness Correlation
Logistic Regression	ALT only	0.8969	0.8851	0.9951
Logistic Regression	ALT + SCN	0.8969	0.8851	0.9947
Logistic Regression	ALT + SCN + INT	0.9027	0.9024	0.9964
Random Forest	ALT + SCN + INT	0.8696	0.7922	0.9622
Gradient Boosting	ALT + SCN + INT	0.8531	0.7751	0.9600
MLP Classifier	ALT + SCN + INT	0.8978	0.8967	0.9967
RankNet	**ALT + SCN + INT**	**0.9018**	**0.8997**	**0.9966**

Note: Logistic regression, random forest, gradient boosting, and the multilayer perceptron classifier follow the formulations in [40–43] and were implemented using scikit-learn [44]. RankNet follows the pairwise learning-to-rank formulation in [35]. ALT denotes alternative-level features; SCN denotes direct scenario variables; and INT denotes alternative–scenario interaction features. All numerical values are the authors' results averaged across ten repeated scenario-group holdout experiments.

The ablation results provide further insight into the role of contextual information. Using alternative-level features only, logistic regression achieved a mean test accuracy of 0.8969 and a Spearman correlation of 0.8851. Adding the six direct scenario variables did not change either metric, with the ALT + SCN setting producing the same test accuracy and ranking correlation. This result is consistent with the structure of the pairwise dataset. All strategies compared within the same pair belong to the same scenario and therefore share identical scenario-variable values. In a linear pairwise model, these common scenario components do not directly distinguish one strategy from another. Consequently, merely appending scenario variables to the alternative-level feature vector provides little additional preference information. When the 48 interaction features were included, logistic-regression test accuracy increased from 0.8969 to 0.9027, while the Spearman correlation increased from 0.8851 to 0.9024. The improvement was therefore more evident for ranking-order recovery than for binary pairwise accuracy. This indicates that interaction features help preserve the complete strategy order across unseen scenarios, even when their effect on individual pairwise classifications is relatively modest.

The interaction features convert common contextual conditions into alternative-specific information. For example, labor shortage becomes decision-relevant when combined with the labor requirement of a particular strategy, while project urgency becomes informative when combined with duration and flexibility attributes. Similarly, supply instability and weather uncertainty become more meaningful when interacting with strategy-specific risk and safety profiles. These findings support the contextual design of the proposed framework while also providing a more cautious interpretation of the RankNet component. RankNet does not achieve uniformly superior predictive performance compared with all simpler models. Instead, it provides competitive unseen-scenario performance within a pairwise learning-to-rank architecture that directly assigns strategy scores, reconstructs complete rankings, and supports feature-level interpretation through SHAP. Its contribution therefore lies in combining transferable ranking-policy learning with model explainability, rather than in absolute numerical superiority over every baseline.

3.5 Feature Importance and Explainability

To improve the transparency of the RankNet-based preference policy, SHAP values were calculated for the trained neural scoring model. The analysis used the Seed 100 model as a representative run and explained all 64 scenario–strategy instances belonging to its eight unseen test scenarios. A background dataset of 70 training observations was constructed by selecting two strategy instances from each of the 35 training scenarios. SHAP DeepExplainer was then applied to estimate feature-level contributions to the scalar preference scores produced by RankNet. The maximum observed SHAP additivity error was approximately 4.48×10^{-7}, indicating that the feature attributions reproduced the model outputs with negligible numerical discrepancy. The analysis examined three feature groups: eight alternative-level attributes, six direct scenario variables, and 48 alternative–scenario interaction features.

Unlike the CRITIC–CoCoSo stage, which provides scenario-specific rankings and criteria weights, SHAP explains how the RankNet model uses input features to generate preference scores. Therefore, SHAP does not replace the MCDM interpretation; rather, it complements it by revealing how the AI model learns context-dependent preference structures from the MCDM-derived pairwise data.

3.5.1 Global Feature Importance

Fig. 7 presents the global feature importance measured by the mean absolute SHAP value across the 64 unseen test instances. The most influential feature is the alternative-level flexibility attribute, with a mean absolute SHAP value of 0.1762. It is followed by the alternative-level risk attribute at 0.1475 and the interaction between flexibility and project urgency at 0.1372.

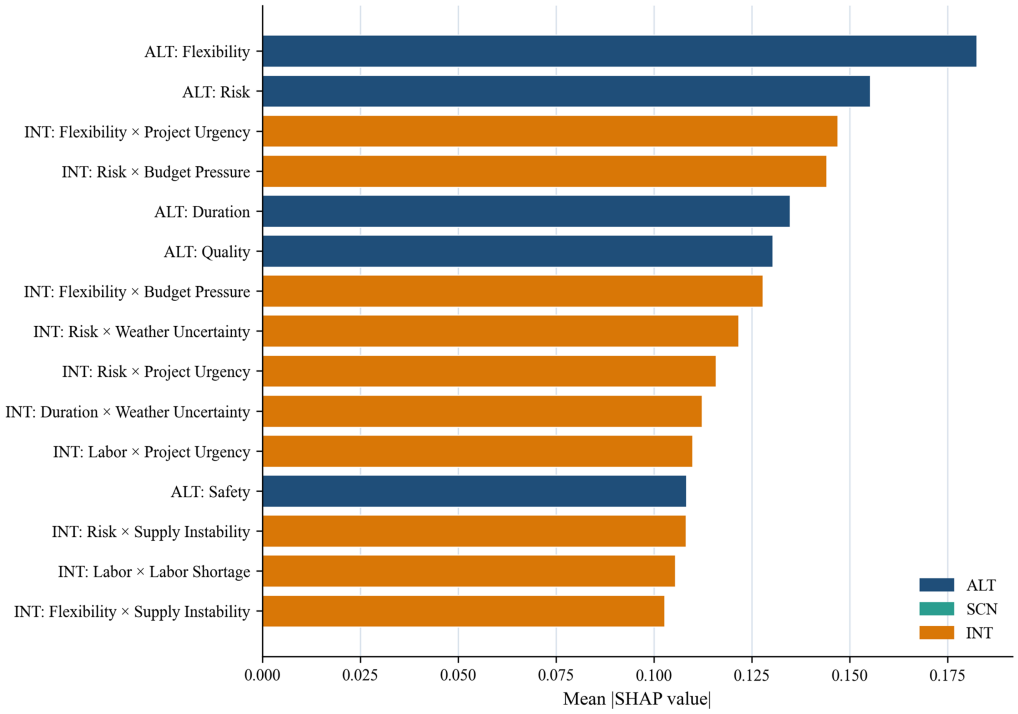

Figure 7: Global SHAP feature importance for the Seed 100 RankNet model, calculated as the mean absolute SHAP value across 64 scenario–strategy observations from eight unseen test scenarios. Alternative-level flexibility and risk attributes, together with flexibility- and risk-related contextual interactions, provide the largest contributions to the learned preference scores.

Several risk-related interaction terms also appear among the leading features. These include Risk × Supply Instability (0.1068), Risk × Budget Pressure (0.0808), Risk × Weather Uncertainty (0.0798), and Risk × Project Urgency (0.0724). Other influential features include Duration × Weather Uncertainty (0.0747), Flexibility × Supply Instability (0.0673), alternative-level safety (0.0584), and alternative-level carbon performance (0.0582). These results indicate that the learned preference policy is driven primarily by alternative-specific flexibility and risk characteristics together with their contextual interactions. Direct scenario variables have comparatively low standalone importance. Their influence is expressed mainly through interaction terms that connect project conditions with the attributes of individual strategies.

The SHAP results should not be interpreted as equivalent to the CRITIC weights. CRITIC weights quantify the contrast and non-redundant information of criteria within each scenario-specific decision matrix, whereas SHAP values explain how the trained RankNet model uses its input features to generate preference scores. Consequently, the criteria with the highest average CRITIC weights do not necessarily correspond directly to the features with the highest SHAP importance.

3.5.2 Distributional Effects of SHAP Values

Fig. 8 presents the SHAP summary plot, which shows both the magnitude and direction of the feature contributions across the 64 unseen test observations. Each point represents one scenario–strategy instance. Its horizontal position indicates the SHAP contribution to the predicted RankNet score, while its color represents the relative magnitude of the corresponding feature value. Features with wider horizontal distributions exert larger or more variable effects on the model output.

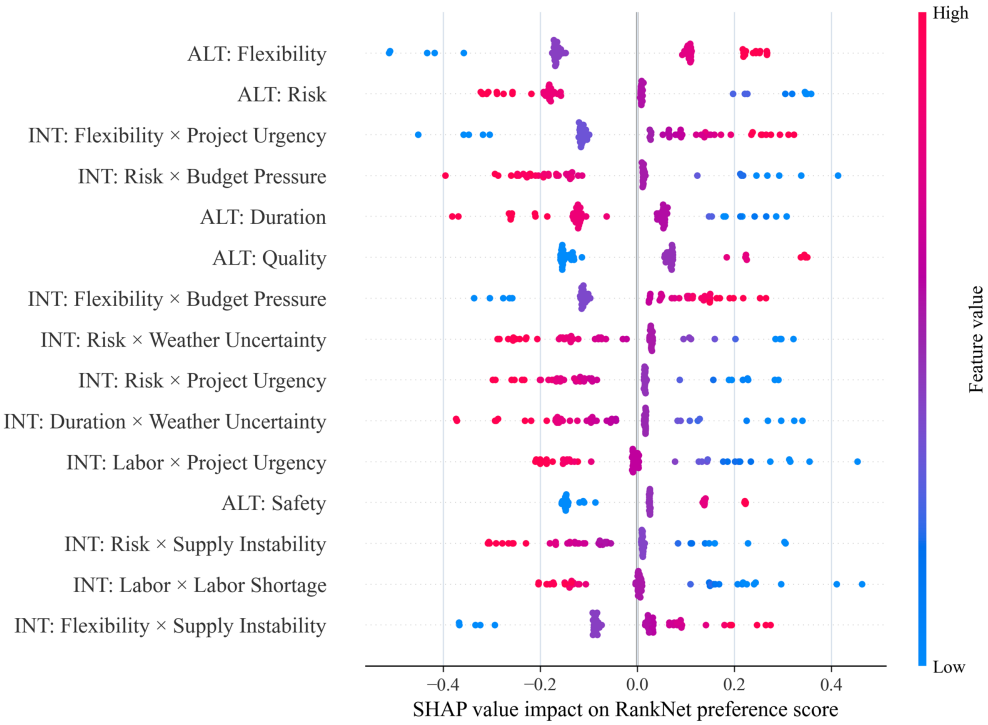

Figure 8: SHAP summary plot for the Seed 100 RankNet model based on 64 unseen test instances. The plot shows the magnitude, direction, and distribution of alternative-level and interaction-feature contributions to the predicted preference scores.

The summary plot confirms that alternative-level flexibility and risk affect preference scores differently across strategies and project conditions. Their SHAP values extend in both positive and negative directions, showing that the same general attribute may increase or decrease a strategy's predicted preference depending on its numerical value and the surrounding interaction structure.

The interaction features display particularly heterogeneous attribution patterns. For example, the contributions of Flexibility × Project Urgency, Risk × Supply Instability, Risk × Budget Pressure, and Risk × Weather Uncertainty vary across the test instances. These distributions indicate that RankNet does not apply one fixed response to urgency, supply disruption, financial pressure, or weather uncertainty. Instead, it evaluates these conditions jointly with the relevant attributes of each candidate strategy.

Direct scenario variables generally show smaller standalone SHAP magnitudes than the leading interaction features. This result is consistent with the pairwise feature structure because all strategies within a scenario share the same direct contextual values. The contextual variables become more discriminative when combined with alternative-level characteristics, allowing the model to distinguish how different strategies respond to the same project environment.

3.5.3 Local Explanation of Individual Decision Behavior

To illustrate how individual feature contributions combine in a specific recommendation, Fig. 9 presents a local SHAP waterfall plot for S4 Risk Averse under the Tight Budget–Weather Risk scenario. In this case, S4 was ranked first by both the T-spherical fuzzy CRITIC–CoCoSo procedure and the RankNet model. The waterfall plot starts from the expected model output and sequentially adds the positive and negative SHAP contributions associated with the selected scenario–strategy observation. Risk-related alternative and interaction features provide important contributions to the predicted score, reflecting the alignment between the risk-oriented profile of S4 and a scenario characterized by simultaneous financial and weather-related pressure. Safety, flexibility, and their contextual relationships also contribute to the final model output.

The local explanation demonstrates that the recommendation is not generated from the strategy label alone. Rather, the predicted preference score emerges from the combined effects of the strategy's fuzzy performance attributes and their interactions with the project conditions. Features that raise the preference score indicate alignment with the scenario, whereas negative SHAP contributions identify attributes that reduce the relative attractiveness of S4. The agreement between the MCDM and RankNet rank in this case provides an interpretable example of successful preference-policy recovery. However, the waterfall plot explains the internal behavior of the trained model for one observation and should not be interpreted as a general causal explanation of why S4 will perform best in every budget- and weather-constrained project.

3.5.4 Interpretation of the Learned Decision Mechanism

Overall, the SHAP analysis indicates that RankNet learns a context-aware preference mechanism based on the combination of alternative-level attributes and alternative–scenario interactions. Alternative-level flexibility and risk are the most influential direct features, while interactions involving urgency, supply instability, budget pressure, and weather uncertainty transmit much of the contextual information into the model. The results are consistent with the comparative and ablation analysis in Section 3.4. Adding direct scenario variables to alternative-level features did not improve logistic-regression performance because the scenario values were common to all strategies compared within the same project context. In contrast, the interaction features translated those shared conditions into alternative-specific information. Their prominence in the SHAP results helps explain the improvement in ranking recovery observed when the complete feature set was used.

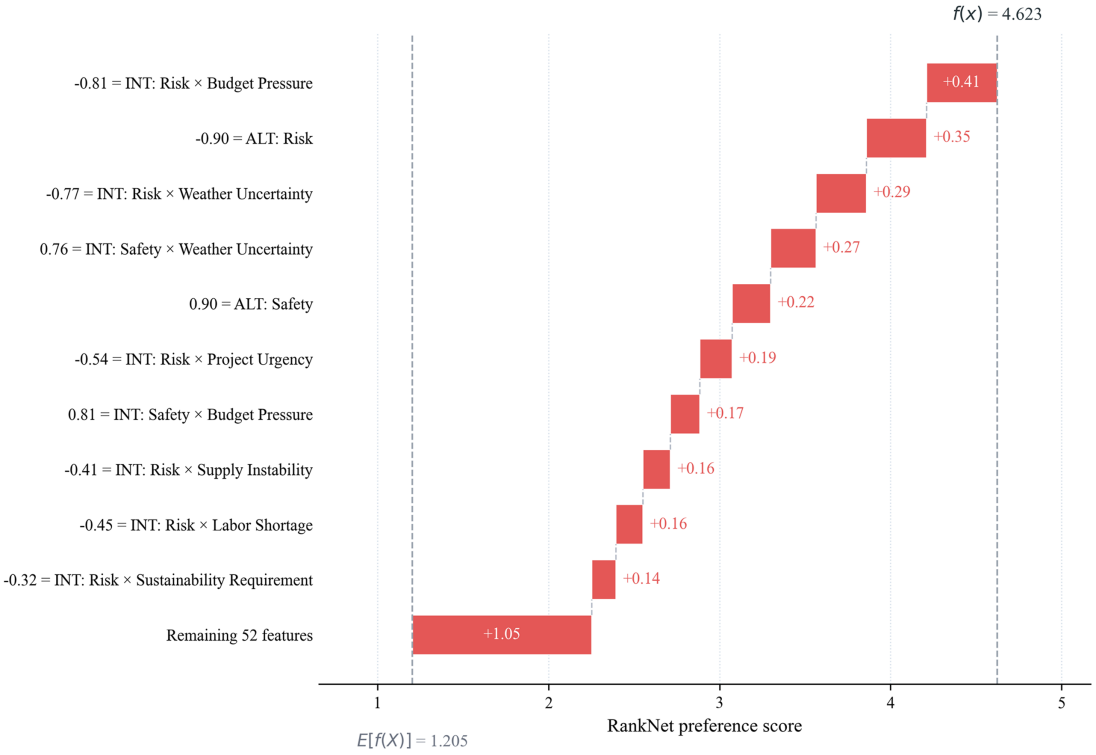

Figure 9: Local SHAP waterfall explanation for S4 Risk Averse under the Tight Budget–Weather Risk scenario. Both the MCDM procedure and RankNet assigned S4 the first rank. The plot decomposes the predicted RankNet preference score into positive and negative feature-level contributions.

The explainability analysis also clarifies the distinction between the MCDM and AI components. The CRITIC–CoCoSo procedure determines reference rankings through fuzzy scoring, objective weighting, and compromise aggregation. RankNet subsequently learns an approximate scoring policy from those reference preferences. SHAP then explains how the trained RankNet model combines its 62 input features when reproducing or transferring that ranking structure to unseen scenarios. SHAP values are model-attribution measures rather than causal effects. They quantify how features contribute to the trained model output relative to the selected background distribution but do not demonstrate that those features causally determine construction-project outcomes. Accordingly, the results support the transparency and internal interpretability of the learned preference policy, while causal and external managerial conclusions require validation using observed project data.

4 Discussion and Managerial Implications

4.1 Discussion of Key Findings

The results demonstrate that construction strategy selection is strongly scenario-dependent and cannot be adequately represented by a single static ranking. The T-spherical fuzzy CRITIC–CoCoSo stage showed that criteria weights and strategy rankings vary across different project conditions. This finding supports the central argument of the study: construction strategies should be evaluated in relation to the project environment in which they are implemented, rather than under fixed assumptions.

The CRITIC results reveal a relatively balanced objective weighting structure rather than the dominance of one criterion. Cost receives the highest mean weight at 0.1326, followed by carbon at 0.1286, labor at 0.1256,

quality at 0.1253, flexibility at 0.1237, risk at 0.1235, duration at 0.1225, and safety at 0.1182. The narrow range among these mean values indicates that construction strategy differentiation arises from multiple concurrent performance dimensions. The mean weights should not be interpreted as fixed managerial priorities because CRITIC recalculates them independently for each scenario. In particular, risk exhibits substantial cross-scenario variation, showing that its informational contribution increases under uncertainty-intensive conditions. The results therefore support a dynamic interpretation of criterion importance: cost has the highest average information contribution, but its advantage is modest, and criteria such as risk, carbon, labor, and flexibility may become more influential under specific project environments.

The CoCoSo results distinguish between scenario-specific dominance and cross-scenario consistency. S4 Risk Averse ranks first in 20 of the 50 scenarios, the highest first-place frequency among the eight strategies. Its strong performance is concentrated in scenarios involving weather disruption, supply instability, disaster response, climate stress, and other uncertainty-intensive conditions. However, S4 also has the highest rank standard deviation at 3.01. It should therefore be interpreted as a highly specialized strategy that can become dominant under severe uncertainty but is not uniformly competitive across all project settings. In contrast, S5 Tech Intensive has the best mean rank of 3.08 and the lowest rank standard deviation of 0.92. It appears among the top three strategies in 36 scenarios but does not rank first in any scenario. This pattern indicates consistent competitiveness rather than scenario-specific dominance. S5 may therefore represent a reliable candidate for shortlisting across heterogeneous project environments, although it is not necessarily the optimal choice under any single condition. S3 Sustainability ranks first in 11 scenarios and becomes particularly competitive under strict environmental, regulatory, and green-development requirements. S8 Balanced ranks first in seven scenarios and performs well under stable or mixed conditions in which no single constraint dominates. S6 Labor Saving also ranks first in seven scenarios, primarily when workforce scarcity is severe. The remaining strategies show narrower forms of specialization. S2 Fast Track ranks first in four scenarios, mainly under critical urgency or combined schedule pressure. S1 Cost Minimization ranks first only in the pure Tight Budget scenario, indicating that a cost-oriented strategy becomes preferable when financial pressure is dominant but loses competitiveness when broader trade-offs are introduced. S7 Quality Priority does not rank first in the scenario set, although it retains moderate performance in several quality-sensitive contexts. These results show that the most frequently optimal strategy is not necessarily the most stable one. S4 provides the strongest scenario-specific dominance, whereas S5 provides the strongest cross-scenario consistency. This distinction is managerially important because a strategy selected for robustness across diverse conditions may differ from one selected for a narrowly defined high-pressure scenario.

A major finding of the learning stage is that RankNet recovered the MCDM-derived preference structure with reasonable consistency under scenario-group holdout evaluation. Across ten random seeds, the model achieved a mean test pairwise accuracy of 0.9018 ± 0.0234 and a mean test Spearman correlation of 0.8997 ± 0.0308. Because complete scenarios rather than individual pairs were withheld from training, these results provide evidence of transferability to unseen simulated project conditions.

The results should nevertheless be interpreted as an approximation of the MCDM-derived preference policy rather than external predictive validation. The target rankings were generated by the proposed T-spherical fuzzy CRITIC–CoCoSo procedure, not observed from completed construction projects. Consequently, RankNet demonstrates the ability to transfer a structured decision rule across the simulated scenario space, but its effectiveness on empirical project data remains to be established.

The comparative results provide a more nuanced assessment of the AI component. Under the complete feature setting, logistic regression achieved the highest mean test accuracy at 0.9027 and the highest Spearman correlation at 0.9024, followed closely by RankNet at 0.9018 and 0.8997, respectively. The multilayer perceptron also produced comparable results, with a test accuracy of 0.8978 and a Spearman correlation

of 0.8967. RankNet therefore does not demonstrate uniform numerical superiority over all simpler models. RankNet nevertheless performs substantially better than random forest and gradient boosting in recovering complete rankings. The Spearman correlations of the two tree-based models were 0.7922 and 0.7751, respectively, despite comparatively moderate pairwise accuracies. This difference suggests that individual pairwise classification performance does not necessarily guarantee faithful reconstruction of the full strategy order within an unseen scenario. The feature-ablation results further show that direct scenario variables alone do not improve logistic-regression performance because all strategies within a scenario share identical contextual values. When alternative–scenario interaction features are included, test accuracy increases modestly from 0.8969 to 0.9027, while the Spearman correlation increases more clearly from 0.8851 to 0.9024. The interaction terms therefore contribute particularly to ranking-order recovery by converting common scenario conditions into alternative-specific information. The value of RankNet should consequently be understood in terms of its pairwise scoring architecture, direct generation of sortable strategy scores, ability to represent nonlinear interactions, and compatibility with SHAP-based interpretation. Its contribution is not based on outperforming every baseline model by a large margin.

The SHAP results provide additional evidence that the learned model relies on alternative-specific contextual relationships. Alternative-level flexibility and risk are the two most influential direct features, with mean absolute SHAP values of 0.1762 and 0.1475, respectively. The interaction between flexibility and project urgency is the most influential interaction feature at 0.1372, followed by Risk × Supply Instability at 0.1068. Other important interactions involve risk with budget pressure, weather uncertainty, and project urgency, as well as duration with weather uncertainty and flexibility with supply instability. By contrast, direct scenario variables have relatively low standalone SHAP importance. The contextual information is therefore transmitted primarily through its interactions with strategy attributes rather than through isolated scenario values. This finding is consistent with the ablation results and reinforces the central modeling argument: project context becomes decision-relevant when it modifies the suitability of a particular strategy profile. SHAP explains the internal attribution structure of RankNet but should not be interpreted as evidence that these variables causally determine project outcomes.

4.2 Theoretical Implications

This study contributes to the literature on fuzzy MCDM by extending its role from static evaluation to decision policy learning. Traditional fuzzy MCDM methods are useful for handling uncertainty and producing rankings, but they usually remain limited to a specific decision matrix. The proposed framework demonstrates that fuzzy MCDM outputs can be transformed into structured pairwise preference data and used to train a learning model. More specifically, 50 scenario-specific rankings are transformed into 1400 non-duplicated unordered pairwise observations, enabling a scoring policy to be evaluated on complete scenarios excluded from model training. This creates a bridge between expert-driven decision modeling and data-driven AI.

The study also contributes to hybrid MCDM–AI research by clarifying the role of AI in the decision process. In the proposed framework, AI is not used as a replacement for MCDM. Instead, MCDM provides structured and interpretable reference rankings, while AI learns an approximate preference policy that can be applied to unseen simulated scenarios. This division of roles helps reduce the risk of treating AI as a black-box substitute for decision analysis. It also creates a more defensible hybrid architecture in which uncertainty modeling, objective weighting, ranking, learning, and explainability are connected sequentially.

A further methodological contribution concerns the unit of data partitioning in pairwise learning-to-rank. Randomly dividing individual pairs can place observations from the same scenario in both training and testing subsets, thereby inflating apparent generalization. By keeping all pairs from a scenario within

one partition, the present study evaluates whether the learned policy transfers across contexts rather than merely across additional comparisons drawn from contexts already observed during training. This principle is relevant to other MCDM–AI applications in which several pairwise observations are generated from the same decision environment.

Another theoretical implication concerns feature interaction modeling. The ablation analysis shows that scenario variables alone may not improve pairwise learning because they do not distinguish alternatives within the same scenario. Interaction features are necessary to translate contextual conditions into alternative-specific preference differences. This finding is important for future AI-based MCDM studies. It suggests that simply adding contextual variables to a model may be insufficient; researchers should explicitly model how context changes the meaning and importance of alternative attributes. The SHAP results provide complementary evidence by showing that several of the most influential inputs are interactions involving flexibility, risk, urgency, supply instability, budget pressure, and weather uncertainty.

The study also contributes to the methodological treatment of T-spherical fuzzy information by applying a neutrality-aware score function consistently throughout the framework. The adopted function includes membership, neutrality or abstinence, and non-membership degrees directly, avoiding the need for an externally specified penalty coefficient. This improves conceptual alignment between the T-spherical fuzzy representation and the numerical scoring stage and prevents the neutrality component from being discarded during defuzzification.

4.3 Managerial Implications

The proposed framework provides several practical implications for construction managers, project planners, and decision-support system developers.

First, construction strategy selection should be treated as a scenario-dependent decision problem. A strategy that performs well under stable conditions may not remain optimal under labor shortages, weather disruptions, urgent deadlines, or budget pressure. Therefore, managers should avoid relying on a single universal ranking of strategies. Instead, strategy selection should be adjusted according to project-specific conditions. The relevant question is not which strategy is universally best, but which strategy profile is most suitable for the current configuration of project constraints.

Second, managers should distinguish between a consistently competitive strategy and a scenario-specific optimal strategy. S5 Tech Intensive records the best mean rank and the lowest ranking variability, making it a comparatively stable candidate across heterogeneous conditions. However, it never ranks first. S4 Risk Averse, by contrast, ranks first most frequently but also exhibits the greatest variability. Thus, S5 may be suitable as a broadly competitive baseline or shortlist option, whereas S4 should be considered when severe uncertainty, disruption, or risk exposure dominates the project environment.

Third, specialized strategies should be selected in alignment with the dominant project constraint. S3 Sustainability becomes attractive under stringent environmental and regulatory requirements; S6 Labor Saving becomes competitive under severe workforce scarcity; S2 Fast Track is favored under critical urgency; and S1 Cost Minimization becomes preferable when tight budget pressure is the primary concern. S8 Balanced performs strongly in stable or mixed environments without a clearly dominant constraint. These findings indicate that specialized strategies should not be judged solely by their average ranking across all scenarios.

Fourth, interaction effects should be explicitly considered in decision support. The results show that project conditions influence strategies differently. Labor shortage does not affect all strategies equally; budget pressure does not have the same implication for cost-minimization and technology-intensive strategies;

and urgency does not affect slow and fast strategies in the same way. Therefore, managers should evaluate not only the severity of project constraints but also how each strategy responds to those constraints. The prominence of flexibility- and risk-related interaction features indicates that adaptability and uncertainty response deserve particular attention when project conditions are unstable.

Fifth, the RankNet component can support rapid scenario screening after it has been trained on a coherent set of MCDM-derived preferences. It assigns scalar scores to candidate strategies and reconstructs their ranking without requiring the entire fuzzy weighting and compromise-ranking procedure to be repeated for each exploratory scenario. This can be useful in early-stage planning, interactive scenario analysis, and preliminary strategy shortlisting. However, the model should be used as a decision-support mechanism rather than an autonomous strategy-selection system. Its policy is learned from simulated scenario definitions and MCDM-derived rankings. Project-specific expert review and empirical information remain necessary before a recommendation is implemented.

Sixth, SHAP-based explanations can improve the transparency of AI-assisted recommendations. Global explanations identify the attributes and interactions that the model uses most strongly, while local waterfall explanations show why a particular strategy receives a high or low score under a specific scenario. For example, the local explanation for S4 Risk Averse under the Tight Budget–Weather Risk scenario demonstrates how risk, safety, flexibility, and contextual interactions combine in one recommendation. These explanations may help managers challenge, validate, or reject a model-generated ranking. Nevertheless, SHAP values describe the behavior of the trained model and should not be interpreted as causal proof that changing a feature will necessarily improve project performance.

4.4 Advantages and Limitations of the Proposed Framework

The proposed framework offers several methodological advantages. First, it integrates uncertainty representation, objective criteria weighting, compromise ranking, preference learning, and model explanation within one sequential structure. The neutrality-aware T-spherical fuzzy score retains membership, neutrality or abstinence, and non-membership information, while CRITIC and CoCoSo generate scenario-specific reference rankings without requiring a fixed subjective weighting structure. Second, the scenario-group evaluation design provides a more rigorous assessment than random pair-level splitting. All pairwise observations generated from the same scenario remain within one data partition, thereby reducing information leakage between training and testing subsets. The exclusion of reverse-pair duplicates also prevents the effective sample size from being artificially inflated. Consequently, model performance reflects transferability across unseen simulated scenarios rather than additional comparisons from previously observed contexts. Third, the framework explicitly represents alternative–scenario interactions. The ablation and SHAP results show that direct scenario variables alone provide limited discrimination among strategies within the same project context, whereas interaction features translate common contextual conditions into alternative-specific information. RankNet further provides directly sortable preference scores, while SHAP improves transparency by identifying the attributes and interactions contributing to individual and global predictions. The low CPU training time also makes the framework suitable for repeated scenario screening.

Several limitations should nevertheless be recognized. First, the scenario dataset is simulation-driven and based on manually calibrated strategy profiles and contextual sensitivities. Although the scenarios were designed to be managerially coherent, they do not constitute observed construction-project outcomes. The results therefore demonstrate internal methodological consistency within the modeled scenario space rather than external predictive validity. Second, RankNet learns from CRITIC–CoCoSo-derived reference rankings and consequently inherits the assumptions embedded in the linguistic scale, scenario definitions, baseline strategy profiles, sensitivity coefficients, normalization rules, criterion-weighting procedure, and

compromise-ranking formulation. Agreement between RankNet and MCDM should therefore be interpreted as successful approximation of the reference preference structure rather than independent evidence of objectively optimal construction decisions.

Third, the scenario-group holdout procedure evaluates transferability only within the range of contextual variables represented by the 50 designed scenarios. It does not establish robustness under major distributional shifts, previously unmodeled constraints, or different construction sectors and geographic settings. Similarly, the Gaussian perturbation analysis assesses only local prediction stability. Finally, SHAP explains how the trained RankNet model uses its inputs but does not establish causal relationships between project conditions, strategy characteristics, and realized construction outcomes. Future studies should therefore validate the framework using expert-elicited cases, observed project records, alternative scenario-generation mechanisms, and out-of-distribution stress tests. Human review should remain part of the final strategy-selection process.

5 Conclusions

This study developed a hybrid T-spherical fuzzy CRITIC–CoCoSo and RankNet framework for scenario-dependent construction strategy selection. The framework transforms scenario-specific fuzzy MCDM rankings into non-duplicated pairwise preferences and learns an interpretable scoring policy using scenario-group data partitioning. This design extends conventional static MCDM analysis by evaluating whether preference patterns can be transferred to complete simulated scenarios excluded from model training.

The findings confirm that construction strategy preference changes substantially across project conditions. The CRITIC results produced a relatively balanced weighting structure, indicating that strategy differentiation depends on multiple criteria rather than one dominant factor. The CoCoSo rankings further distinguished scenario-specific dominance from cross-scenario consistency. S4 Risk Averse achieved the highest first-place frequency but also showed the greatest ranking variability, whereas S5 Tech Intensive achieved the best mean rank and the lowest dispersion without ranking first in any scenario. These contrasting patterns demonstrate that the most frequently optimal strategy is not necessarily the most consistently competitive and that no strategy is universally preferable. Across ten repeated scenario-group holdout experiments, RankNet achieved a mean test accuracy of 0.9018 ± 0.0234, a mean Spearman correlation of 0.8997 ± 0.0308, and a local perturbation robustness correlation of 0.9966 ± 0.0020. RankNet performed comparably to logistic regression and the multilayer perceptron classifier and preserved ranking structures more effectively than the examined tree-based ensembles. Its contribution therefore lies not in uniformly outperforming every baseline, but in providing a pairwise scoring architecture that generates sortable strategy scores, represents nonlinear contextual relationships, and supports SHAP-based interpretation. The ablation and explainability analyses showed that direct scenario variables alone contributed limited alternative discrimination because all strategies within the same scenario shared identical contextual values. Alternative–scenario interaction features were more informative because they translated common project conditions into strategy-specific preference effects. SHAP further identified flexibility, risk, and their interactions with urgency, supply instability, budget pressure, and weather uncertainty as important drivers of the learned policy. All data, computational results, and reproduction code are available in the supplementary material.

The principal limitations arise from the simulation-driven dataset, the dependence of RankNet on MCDM-derived reference rankings, the limited scenario space modeled in this study, and the use of local perturbation tests rather than broader distributional robustness assessments. Accordingly, the results

demonstrate internal methodological consistency and transferability within the designed scenario environment, but not external predictive validity for observed construction projects.

Future research should validate the framework using empirical project records and structured expert elicitation, compare predicted rankings with realized project outcomes, and examine alternative scenario-generation and learning-to-rank approaches. Additional developments may include probabilistic scenario modeling, out-of-distribution stress testing, uncertainty quantification, online model updating, and human-in-the-loop interfaces. Overall, the framework should be viewed as a transparent decision-support mechanism that combines fuzzy uncertainty representation, scenario-dependent ranking, preference learning, and explainability while retaining professional judgment as the final basis for strategy selection.

Acknowledgement: Not applicable.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Thi-Hien Dao; methodology, Thi-Hien Dao; software, Thi-Hien Dao; validation, Thi-Hien Dao; formal analysis, Thi-Hien Dao; investigation, Yih-Tzoo Chen and Thi-Hien Dao; resources, Yih-Tzoo Chen and Thi-Hien Dao; data curation, Yih-Tzoo Chen and Thi-Hien Dao; writing—original draft preparation, Thi-Hien Dao; writing—review and editing, Thi-Hien Dao; visualization, Thi-Hien Dao; supervision, Yih-Tzoo Chen; project administration, Yih-Tzoo Chen. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The supplementary material associated with this article is publicly available on Zenodo at https://doi.org/10.5281/zenodo.21485452.

Ethics Approval: Not applicable.

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

References

1. Zhang L, Pan Y, Wu X, Skibniewski MJ. Artificial intelligence in construction engineering and management. Singapore: Springer; 2021. doi:10.1007/978-981-16-2842-9.
2. Abioye SO, Oyedele LO, Akanbi L, Ajayi A, Davila Delgado JM, Bilal M, et al. Artificial intelligence in the construction industry: a review of present status, opportunities and future challenges. J Build Eng. 2021;44(5):103299. doi:10.1016/j.jobe.2021.103299.
3. Li Y, Song H, Sang P, Chen PH, Liu X. Review of Critical Success Factors (CSFs) for green building projects. Build Environ. 2019;158(2):182–91. doi:10.1016/j.buildenv.2019.05.020.
4. Mardani A, Jusoh A, Zavadskas EK. Fuzzy multiple criteria decision-making techniques and applications–two decades review from 1994 to 2014. Expert Syst Appl. 2015;42(8):4126–48. doi:10.1016/j.eswa.2015.01.003.
5. Zavadskas EK, Antucheviciene J, Vilutiene T, Adeli H. Sustainable decision-making in civil engineering, construction and building technology. Sustainability. 2017;10(1):1–21. doi:10.3390/su10010014.
6. Zavadskas EK, Turskis Z, Kildienė S. State of art surveys of overviews on MCDM/MADM methods. Technol Econ Dev Econ. 2014;20(1):165–79. doi:10.3846/20294913.2014.892037.
7. Chang L, Nordin N, Zhao S, Gu X, Zhao Y. TOPSIS prefabricated building construction evaluation based on interval-valued Pythagorean fuzzy numbers based on prospect theory. Sci Rep. 2025;15(1):2913. doi:10.1038/s41598-025-85729-1.
8. Ali J, Naeem M. r, s, t-spherical fuzzy VIKOR method and its application in multiple criteria group decision making. IEEE Access. 2023;11(1):46454–75. doi:10.1109/ACCESS.2023.3271141.
9. Behzadian M, Kazemzadeh RB, Albadvi A, Aghdasi M. PROMETHEE: a comprehensive literature review on methodologies and applications. Eur J Oper Res. 2010;200(1):198–215. doi:10.1016/j.ejor.2009.01.021.

10. Diakoulaki D, Mavrotas G, Papayannakis L. Determining objective weights in multiple criteria problems: the critic method. Comput Oper Res. 1995;22(7):763–70. doi:10.1016/0305-0548(94)00059-H.
11. Yazdani M, Zarate P, Kazimieras Zavadskas E, Turskis Z. A combined compromise solution (CoCoSo) method for multi-criteria decision-making problems. Manag Decis. 2019;57(9):2501–19. doi:10.1108/MD-05-2017-0458.
12. Wang CN, Pham TT, Nhieu NL. A composited regret-theory-based spherical fuzzy prioritization approach for moving high-tech manufacturing in Southeast Asia. Appl Sci. 2023;13(2):688. doi:10.3390/app13020688.
13. Farman S, Khan FM, Bibi N. T-Spherical fuzzy soft rough aggregation operators and their applications in multi-criteria group decision-making. Granul Comput. 2023;9(1):6. doi:10.1007/s41066-023-00437-3.
14. Razzaq A, Riaz M. Picture fuzzy soft-max Einstein interactive weighted aggregation operators with applications. Comput Appl Math. 2024;43(2):90. doi:10.1007/s40314-024-02609-6.
15. Gong K, Ma W, Zhang H, Goh M. Heterogeneous multi-attribute large-scale group decision-making considering individual concerns and information credibility. Group Decis Negot. 2023;32(6):1315–49. doi:10.1007/s10726-023-09845-x.
16. Li Y, Liu M, Cao J, Wang X, Zhang N. Multi-attribute group decision-making considering opinion dynamics. Expert Syst Appl. 2021;184(5):115479. doi:10.1016/j.eswa.2021.115479.
17. Wan SP, Wu H, Dong JY. An integrated method for complex heterogeneous multi-attribute group decision-making and application to photovoltaic power station site selection. Expert Syst Appl. 2024;242:122456. doi:10.1016/j.eswa.2023.122456.
18. Zhou H, Wang JQ, Zhang HY, Chen XH. Linguistic hesitant fuzzy multi-criteria decision-making method based on evidential reasoning. Int J Syst Sci. 2016;47(2):314–27. doi:10.1080/00207721.2015.1042089.
19. Sahoo SK, Goswami SS. A comprehensive review of multiple criteria decision-making (MCDM) methods: advancements, applications, and future directions. Decis Mak Adv. 2023;1(1):25–48. doi:10.31181/dma1120237.
20. Avramova T, Peneva T, Ivanov A. Overview of existing multi-criteria decision-making (MCDM) methods used in industrial environments. Technologies. 2025;13(10):444. doi:10.3390/technologies13100444.
21. Ghorbany S, Hu M, Nouri A. Commercial buildings decarbonization: benchmarks and strategies for the United States sustainable commercial construction. Sustain Cities Soc. 2025;124(10):106324. doi:10.1016/j.scs.2025.106324.
22. Chen YT, Dao TH, Nhieu NL. Decoding efficiency in construction: a behavioral DEA approach to BRICS economies. IEEE Access. 2024;12:113165–77. doi:10.1109/ACCESS.2024.3443925.
23. Soto Ortiz S, Hubbard B, Kang K, Besiktepe D. Managing black swan event risks in the construction supply chain: a literature review. Intell Infrastruct Constr. 2026;2(1):3. doi:10.3390/iic2010003.
24. Chen YT, Dao TH, Chiu IH. A spherical fuzzy objective weighting framework for factors affecting bidding decisions for construction projects: a case study of Taiwan. Heliyon. 2024;10(16):e36420. doi:10.1016/j.heliyon.2024.e36420.
25. Nguyen TMH, Nguyen VP, Nguyen DT. A new hybrid Pythagorean fuzzy AHP and COCOSO MCDM based approach by adopting artificial intelligence technologies. J Exp Theor Artif Intell. 2024;36(7):1279–305. doi:10.1080/0952813X.2022.2143908.
26. Naz S, Tasawar A, Fatima A, Butt SA, Gonzalez ZC. An efficient 2-tuple linguistic cubic q-rung orthopair fuzzy CILOS-TOPSIS method: evaluating the hydrological geographical regions for watershed management in Pakistan. J Supercomput. 2024;81(1):103. doi:10.1007/s11227-024-06505-y.
27. Nhieu NL. The T-spherical fuzzy Einstein interaction operation matrix energy decision-making approach: the context of Vietnam offshore wind energy storage technologies assessment. Mathematics. 2024;12(16):2498. doi:10.3390/math12162498.
28. Munir M, Mahmood T, Hussain A. Algorithm for T-spherical fuzzy MADM based on associated immediate probability interactive geometric aggregation operators. Artif Intell Rev. 2021;54(8):6033–61. doi:10.1007/s10462-021-09959-1.
29. Salhi A, Alshamrani R, Althbiti A, Ismail A, Abd-ElRahman M, Hassan BM. Optimizing high dimensional data classification with a hybrid AI driven feature selection framework and machine learning schema. Sci Rep. 2025;15(1):35038. doi:10.1038/s41598-025-08699-4.

30. Ahmed SF, Bin Alam MS, Hassan M, Rozbu MR, Ishtiak T, Rafa N, et al. Deep learning modelling techniques: current progress, applications, advantages, and challenges. Artif Intell Rev. 2023;56(11):13521–617. doi:10.1007/s10462-023-10466-8.
31. Rane N. Integrating building information modelling (BIM) and artificial intelligence (AI) for smart construction schedule, cost, quality, and safety management: challenges and opportunities. SSRN J. 2023;4(12):72. doi:10.2139/ssrn.4616055.
32. Datta SD, Islam M, Rahman Sobuz MH, Ahmed S, Kar M. Artificial intelligence and machine learning applications in the project lifecycle of the construction industry: a comprehensive review. Heliyon. 2024;10(5):e26888. doi:10.1016/j.heliyon.2024.e26888.
33. Hallak J, Karah Bash A, Jarallah SK. Hybrid integration of TOPSIS and machine learning models in a scenario-aware framework for shelter prioritization under conflict conditions. J King Saud Univ Comput Inf Sci. 2026;38(5):285. doi:10.1007/s44443-026-00697-4.
34. Köppel M, Segner A, Wagener M, Pensel L, Karwath A, Kramer S. Pairwise learning to rank by neural networks revisited: reconstruction, theoretical analysis and practical performance. In: Machine learning and knowledge discovery in databases. Cham, Switzerland: Springer; 2020. p. 237–52. doi:10.1007/978-3-030-46133-1_15.
35. Burges C, Shaked T, Renshaw E, Lazier A, Deeds M, Hamilton N, et al. Learning to rank using gradient descent. In: Proceedings of the 22nd International Conference on Machine Learning; 2005 Aug 7–11; Bonn, Germany. p. 89–96. doi:10.1145/1102351.1102363.
36. Nwokoro CO, Ejegwa PA. A decade of picture fuzzy sets in multi-criteria decision-making: a comprehensive review of trends, gaps, and future directions. Knowl Decis Syst Appl. 2025;1:145–64. doi:10.59543/kadsa.v1i.14051.
37. Kocaman H, Asan U. Integration modes between MCDM methods and machine learning algorithms: a structured approach for framework development. Mathematics. 2026;14(1):33. doi:10.3390/math14010033.
38. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;30:4768–77. doi:10.48550/arxiv.1705.07874.
39. Mahmood T, Ullah K, Khan Q, Jan N. An approach toward decision-making and medical diagnosis problems using the concept of spherical fuzzy sets. Neural Comput Appl. 2019;31(11):7041–53. doi:10.1007/s00521-018-3521-2.
40. Cox DR. The regression analysis of binary sequences. J R Stat Soc Ser B Stat Methodol. 1958;20(2):215–32. doi:10.1111/j.2517-6161.1958.tb00292.x.
41. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32. doi:10.1023/A:1010933404324.
42. Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat. 2001;29(5):1189–232. doi:10.1214/aos/1013203451.
43. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature. 1986;323(6088):533–6. doi:10.1038/323533a0.
44. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825–30.
45. Shrikumar A, Greenside P, Kundaje A. Learning important features through propagating activation differences. In: Proceedings of the 34th International Conference on Machine Learning; 2017 Aug 6–11; Sydney, Australia. p. 3145–53.
46. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. Pytorch: an imperative style, high-performance deep learning library. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems; 2019 Dec 8–14; Vancouver, BC, Canada. p. 7994–8005.