A Tri-Objective Optimization Framework for Selecting Infrastructure Construction Safety Interventions Balancing Risk, Complexity, and Generalizability

Amr A. Mohy

Construction and Building Engineering Department, Arab Academy for Science and Technology and Maritime Transport, Egypt, Email: A.el-deen5146@student.aast.edu

Abstract

The selection of safety interventions in infrastructure projects is a complex decision-making process involving multiple competing objectives. Existing quantitative models often focus on a bi-objective trade-off, such as risk versus cost, neglecting the strategic value of an intervention's applicability across various scenarios. This paper introduces a novel tri-objective optimization framework to address this gap, using data from OSHA accident narratives in heavy construction. The framework balances three objectives for each intervention: (1) Risk, quantified by a severity score derived using a Large Language Model (LLM); (2) Complexity, a metric termed "Semantic Friction," defined as the cosine distance between the textual embeddings of an unsafe action and its safe alternative; and (3) Generalizability, a metric representing an intervention's broad applicability, calculated as its average semantic similarity to all other safe alternatives. Interventions were grouped into homogenous hazard clusters using k-means clustering. The NSGA-III multi-objective evolutionary algorithm was then employed to identify the 3D Pareto-optimal front for each cluster. The framework's metrics were validated internally, with the generalizability metric demonstrating a statistically significant correlation with linguistic features indicative of broad applicability. This research contributes a third quantifiable criterion, generalizability, to construction safety optimization, providing managers with a nuanced portfolio of solutions balancing risk, implementation complexity, and cross-hazard applicability.

Keywords: Construction Safety, Multi-objective Optimization, NSGA-III, Natural Language Processing, Risk Management, Semantic Similarity, Decision Support Systems.

1. Introduction

The construction industry is characterized by a high incidence of occupational injuries and fatalities, presenting persistent challenges to safety management professionals (Hu et al., 2025; Zermane et al., 2023). Effective safety management relies on the judicious selection and implementation of interventions designed to mitigate identified risks. This selection process constitutes a complex, multi-criteria decision-making problem, where project managers must balance competing objectives such as risk reduction, cost, and operational feasibility (Alkaissy et al., 2022). Traditional approaches often depend on qualitative expert judgment or simplified quantitative models, which may not fully capture the intricate trade-offs involved in creating a resilient safety program (Nyqvist et al., 2024). The increasing availability of large datasets, such as accident investigation narratives, offers an opportunity to develop more sophisticated, data-driven decision-support systems that can systematically evaluate potential safety interventions (Shehadeh and Alshboul, 2025).

Despite advancements in applying artificial intelligence (AI) and machine learning (ML) to construction safety, significant gaps remain in the formulation of decision-support frameworks for intervention selection (Akinosho et al., 2020; Chenya et al., 2022). Existing research has predominantly focused on bi-objective optimization problems, typically modeling the trade-off between risk and cost or time (Cheng et al., 2025; Luong et al., 2019). While valuable, this two-dimensional perspective neglects other critical attributes of an effective intervention. Specifically, a quantifiable, data-driven metric to assess an intervention's broad applicability, or "generalizability," across different hazard contexts is absent from the current body of knowledge. Furthermore, the complexity or cognitive effort required to transition from an unsafe action to a safe alternative is seldom quantified as a distinct objective function, with most models subsuming this aspect under generalized cost metrics (Wen and Hu, 2025). This limitation hinders the ability to prioritize interventions that are not only low-risk but also simple to implement and widely applicable.

This paper aims to develop and validate a tri-objective optimization framework for selecting construction safety interventions that simultaneously balances risk, complexity, and generalizability. The primary objectives are to: (1) systematically structure a large corpus of unstructured accident narratives from the Occupational Safety and Health Administration (OSHA) into a machine-readable format using a large language model (LLM); (2) formulate three competing, quantifiable objectives representing Risk (severity), Complexity (semantic friction between unsafe and safe actions), and Generalizability (semantic similarity to other interventions); (3) apply a multi-objective evolutionary algorithm, NSGA-III, to identify Pareto-optimal portfolios of interventions within distinct hazard clusters; and (4) validate the proposed metrics and framework using internal, data-driven methods, including hypervolume analysis and statistical correlation.

The main contribution of this research is the introduction and formalization of "Generalizability" as a quantifiable third criterion in the safety intervention selection process. This moves beyond conventional bi-objective models to provide a more nuanced decision-making landscape. By optimizing for generalizability, the framework identifies foundational safety measures that can address a wide spectrum of hazards, offering strategic value for training and policy development. A secondary contribution is the formulation of "Semantic Friction," a novel proxy for intervention complexity derived from natural language processing (NLP), which provides a non-monetary measure of implementation difficulty. The resulting framework furnishes safety managers with a portfolio of non-dominated solutions, enabling them to make strategic selections that align with specific organizational priorities, whether they be minimizing immediate risk, reducing implementation complexity, or adopting broadly applicable best practices.

2. Literature Review

2.1 Construction Safety Management and Risk Assessment

The management of occupational health and safety in the construction industry remains a primary area of concern due to the sector's persistently high rates of accidents and injuries (Hu et al., 2025; Zermane et al., 2023). Traditional risk management frameworks often rely on a combination of regulatory compliance, historical accident data analysis, and expert judgment to identify and mitigate hazards (Zou et al., 2007; Khodabakhshian et al., 2023). These approaches, while foundational, can be reactive and may not adequately capture the dynamic and complex interactions of risk factors on a modern construction site (Alkaissy et al., 2020). The limitations of experience-based or static models have motivated a shift toward more dynamic and data-driven methods for risk assessment and decision-making (Aladağ, 2023).

The increasing availability of large-scale datasets, such as those from the Occupational Safety and Health Administration (OSHA), has enabled the application of quantitative and computational techniques to safety management (Shehadeh and Alshboul, 2025). This data-rich environment has facilitated the development of predictive models that use machine learning algorithms to forecast accident likelihood, severity, and type based on project characteristics and historical safety performance data (Yoon et al., 2024; Koc et al., 2023). Such models represent a move toward proactive safety management by providing a basis for resource allocation and targeted inspections. However, while predictive analytics can identify high-risk scenarios, there is a subsequent need for structured methodologies to guide the selection of the most effective interventions from a range of possible alternatives, which remains an underexplored area.

2.2 Artificial Intelligence and NLP in Construction Safety

The application of Artificial Intelligence (AI) has expanded significantly within construction management, offering tools for enhancing productivity, monitoring quality, and improving safety (Akinosho et al., 2020). In safety management, AI is applied in various forms, including computer vision for real-time monitoring of personal protective equipment (PPE) compliance and unsafe behaviors (Liu et al., 2025; Wang, 2025), analysis of sensor data from wearables to assess ergonomic risks (Acharya et al., 2025), and machine learning models for accident prediction (Gondia et al., 2020). These technologies automate data collection and analysis, providing continuous and objective inputs for safety management systems.

A significant portion of safety knowledge, particularly from accident investigations and daily reports, is stored in unstructured text formats. Natural Language Processing (NLP) has emerged as a critical technique for extracting actionable insights from these data sources (Ballal et al., 2024; Toğan et al., 2025). Studies have employed NLP for topic modeling to identify recurring themes in accident narratives (Gadekar and Bugalia, 2023), sentiment analysis of safety reports, and automated classification of incident types (Shuang et al., 2024). More recently, the advent of Large Language Models (LLMs) like Generative Pre-trained Transformers (GPT) has advanced these capabilities, enabling more sophisticated tasks such as automated root cause analysis, generation of safety guidelines, and interactive risk assessment (Martin et al., 2025; Katooziani et al., 2025). This study utilizes an LLM in its initial phase to structure raw OSHA narratives into a consistent format, preparing the data for subsequent quantitative optimization.

2.3 Multi-Objective Optimization in Construction Management

Decision-making in construction management frequently involves balancing multiple, often conflicting, objectives such as time, cost, quality, and safety (Pham and Khoi, 2025). Multi-objective optimization (MOO) provides a formal methodology for addressing such trade-off problems by identifying a set of non-dominated solutions, known as the Pareto front, rather than a single optimal solution (Kaveh et al., 2022). Evolutionary algorithms, including Genetic Algorithms (GA) and Particle Swarm Optimization (PSO), are commonly used to solve these complex, non-linear problems (Afshar and Fathi, 2009). Applications in construction include optimizing project schedules against cost and quality constraints (Nguyen et al., 2023) and resource allocation problems (Kaveh et al., 2021).

Within the domain of safety management, optimization research has typically been limited to bi-objective models, such as balancing safety investment against risk reduction or project duration against safety levels (Zhang et al., 2022). While these models offer valuable insights, they may oversimplify the decision-making process by omitting other pertinent criteria that influence the practical effectiveness of an intervention. The selection of a safety strategy is not merely a function of its risk-mitigation potential and cost; factors such as ease of implementation and its applicability to a range of situations are also important considerations for practitioners. This research addresses this gap by extending the optimization framework to three objectives, utilizing an algorithm well-suited for higher-dimensional problems, NSGA-III, to provide a more comprehensive set of solutions.

2.4 Semantic Analysis and Vector Embeddings

Semantic analysis is a subfield of NLP focused on extracting meaning from text. A foundational technique in modern semantic analysis is the use of vector embeddings, which represent words, sentences, or documents as dense numerical vectors in a high-dimensional space. Models such as Word2Vec, and more recently transformer-based architectures like BERT (Bidirectional Encoder Representations from Transformers), generate embeddings where the geometric relationships between vectors correspond to semantic relationships in the original text (Moon et al., 2022). This technology has been applied in construction research for tasks such as automated review of contractual documents and classification of accident narratives (Shuang et al., 2024).

The quantitative nature of vector embeddings allows for mathematical operations to measure semantic relatedness. Cosine similarity, which measures the cosine of the angle between two vectors, is a standard metric for quantifying how similar two pieces of text are in meaning. A similarity score close to 1 indicates high semantic overlap, while a score near 0 indicates semantic dissimilarity. This study leverages this capability to formulate two of its core objectives. The "Complexity" of an intervention is conceptualized as the semantic distance (1 minus cosine similarity) between the description of an unsafe action and its safe alternative, acting as a proxy for the cognitive or procedural shift required. The "Generalizability" metric is formulated by calculating the average cosine similarity of a given safe alternative against all other safe alternatives in the dataset, providing a data-driven measure of its conceptual breadth. This application of semantic analysis provides the quantitative foundation for the novel dimensions of the proposed optimization framework.

3. Research Methodology

3.1 Methodological Framework

The research methodology is executed in five sequential phases, as depicted in Figure 1. The process begins with structuring raw textual data using a large language model (LLM). This is followed by the formulation of three distinct objectives—Risk, Complexity, and Generalizability—derived from the structured data and its semantic properties. Subsequently, the dataset is partitioned into homogenous hazard clusters. A multi-objective evolutionary algorithm is then applied to each cluster to identify Pareto-optimal sets of interventions. The final phase involves a data-driven validation of the framework's outputs and novel metrics to confirm their internal consistency and analytical utility.

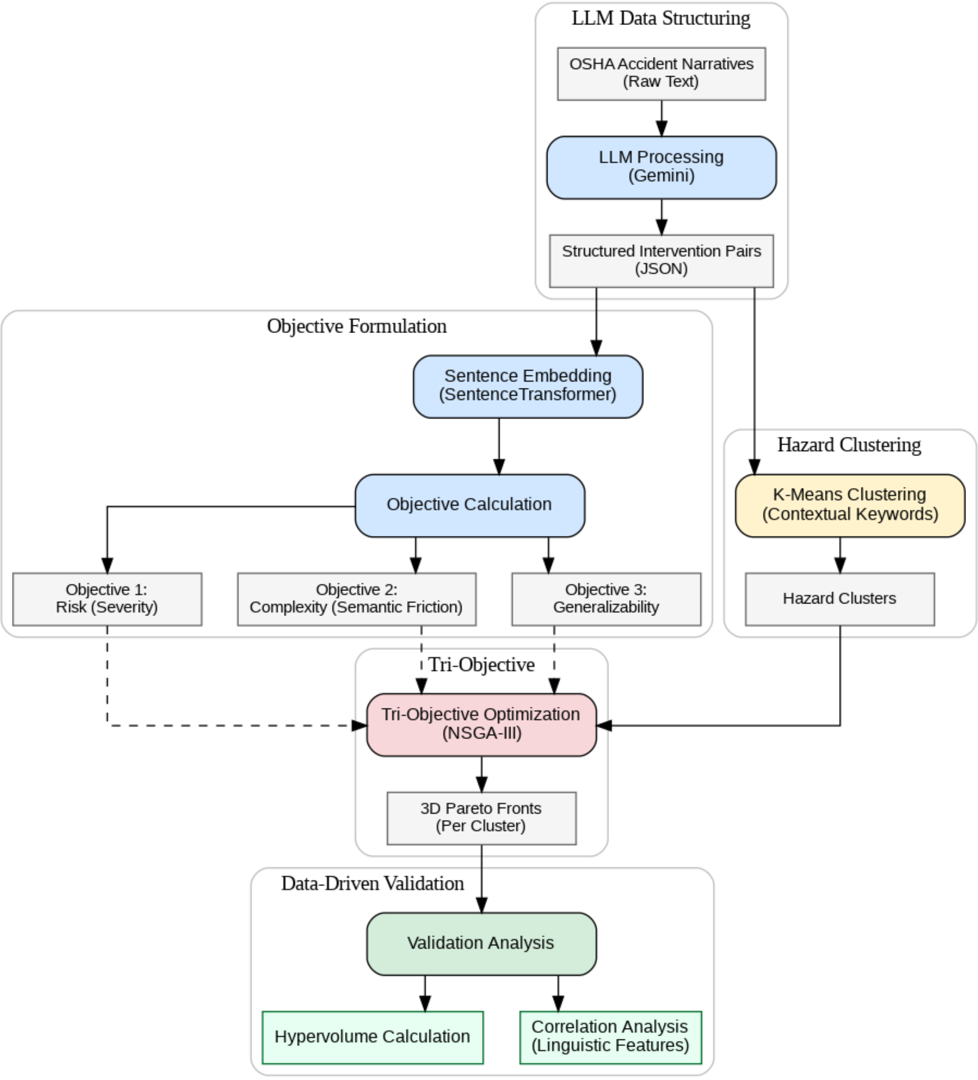


Figure 1: Methodological Framework of the Tri-Objective Optimization Approach.

3.2 LLM-Based Data Structuring

The initial phase of the methodology addresses the transformation of unstructured textual data into a structured format amenable to quantitative analysis. The source data comprises a large corpus of accident investigation narratives obtained from the Occupational Safety and Health Administration (OSHA). Due to their narrative format, these records are not directly suitable for the systematic extraction of features required for optimization modeling. To overcome this limitation, a generative pre-trained transformer model, a type of large language model (LLM), was employed to parse and structure the information contained within each accident report (Martin et al., 2025). This approach leverages the advanced contextual understanding and information extraction capabilities of modern AI to automate a task that would otherwise require extensive manual coding and interpretation (Katooziani et al., 2025).

The process involved iterating through each accident narrative and submitting the text to the LLM via a structured prompt. The prompt was engineered to direct the model to identify and extract four distinct fields of information. The first field, state keywords, consists of a list of terms describing the physical environment and situational context immediately prior to the incident. The second and third fields, unsafe action and safe alternative, capture the core intervention pair by providing concise descriptions of the hazardous action or condition and the corresponding correct procedure or safe state, respectively. The final field, severity, is an integer score assigned by the model on a scale of 1 (minor) to 10 (fatal), representing the outcome of the incident. The model was configured to return its output in a standardized JSON format to ensure machine readability and facilitate straightforward integration into the subsequent analysis pipeline.

To ensure data quality, a filtering protocol was implemented. Narratives with insufficient textual content (fewer than 50 characters) were excluded from processing, as they were unlikely to contain adequate context for reliable information extraction. Additionally, a mechanism was included to handle instances where the model failed to generate a valid JSON response, thereby preventing corrupted data from entering the final dataset. The successful execution of this phase resulted in a structured dataset where each record represents a distinct safety intervention, complete with a quantifiable severity score and contextual keywords, forming the foundation for the objective formulation and clustering stages of the framework.

3.3 Phase 2: Objective Formulation

Following the data structuring process, this phase focuses on the formulation of three competing objectives for each intervention pair. This requires converting the textual descriptions of the unsafe action and safe alternative into numerical representations that can be processed by an optimization algorithm. To achieve this, a pre-trained sentence transformer model (all-MiniLM-L12-v2) was employed to generate high-dimensional vector embeddings for each textual description. This technique maps sentences into a vector space where semantic similarity corresponds to the geometric proximity of their vectors, a method increasingly used for understanding textual data in construction management (Moon et al., 2022). These numerical vectors form the basis for calculating the complexity and generalizability metrics.

The first objective, Risk (O_R), provides a direct measure of the potential harm associated with an unsafe action. It is defined as the integer severity score extracted by the LLM during the data structuring phase, as shown in Equation 1. The goal for this objective within the optimization framework is minimization, prioritizing interventions that address the most severe potential outcomes.

$$O_{R(i)} = Severity_i \qquad \text{Equation 1}$$

The second objective, Complexity (O_C), introduces a novel metric termed "Semantic Friction" to quantify the conceptual distance between an unsafe action and its corresponding safe alternative. This metric serves as a proxy for the cognitive and operational effort required to implement the intervention. It is calculated as the cosine distance (defined as 1 minus the cosine similarity) between the respective sentence embeddings of the unsafe action and safe alternative texts, as formulated in Equation 2. A larger distance signifies a greater semantic disparity, suggesting a more substantial procedural or behavioral change is needed. The optimization goal is to minimize this complexity, favoring interventions that represent a more intuitive or direct correction.

$$O_{C(i)} = 1 - \frac{E_{safe_i} * E_{safe_j}}{|E_{safe_i}||E_{safe_j}|} \qquad \text{Equation 2}$$

where E_x represents the sentence embedding for the textual description x.

The third objective, Generalizability (O_G), is a data-driven metric developed to measure the broad applicability of a given safe alternative. For each safe intervention i, its generalizability score is calculated as the average cosine similarity between its vector embedding and the embeddings of all other safe alternatives in the dataset, as shown in Equation 3. An intervention with a high average similarity, such as "use fall protection," is considered general-purpose because it is semantically related to a wide range of corrective actions. Conversely, a low-scoring intervention, such as "ensure cars are chocked," is considered highly specific to a narrow set of circumstances. The optimization goal for this objective is maximization, which identifies interventions that represent foundational safety principles with wide-ranging relevance.

$$O_G(i) = \frac{1}{N-1} \sum_{j=1, j \neq i}^{N} \frac{E_{safe_i} * E_{safe_j}}{|E_{safe_i}||E_{safe_j}|} \qquad \text{Equation 3}$$

where N is the total number of safe alternatives in the dataset.

3.4 Hazard Clustering

To ensure that the optimization process yields contextually relevant solutions, the dataset of safety interventions was partitioned into homogenous groups based on the type of hazard scenario. A heterogeneous dataset containing a wide variety of accident types would produce a generalized Pareto front that may not be actionable for specific on-site conditions.

By clustering interventions, the subsequent optimization can be performed independently on each distinct hazard group, resulting in tailored sets of optimal solutions. The k-means clustering algorithm, an unsupervised learning technique frequently used for grouping data based on similarity, was selected for this purpose (Oh et al., 2025).

The input features for the clustering algorithm were derived from the state keywords extracted during the LLM data structuring phase. These keywords, which describe the environmental and situational context of each accident, were converted into high-dimensional numerical vectors using a sentence embedding model. The k-means algorithm was then applied to these vectors to partition the interventions into a predefined number of clusters, k, where each cluster contains interventions that occurred in similar settings.

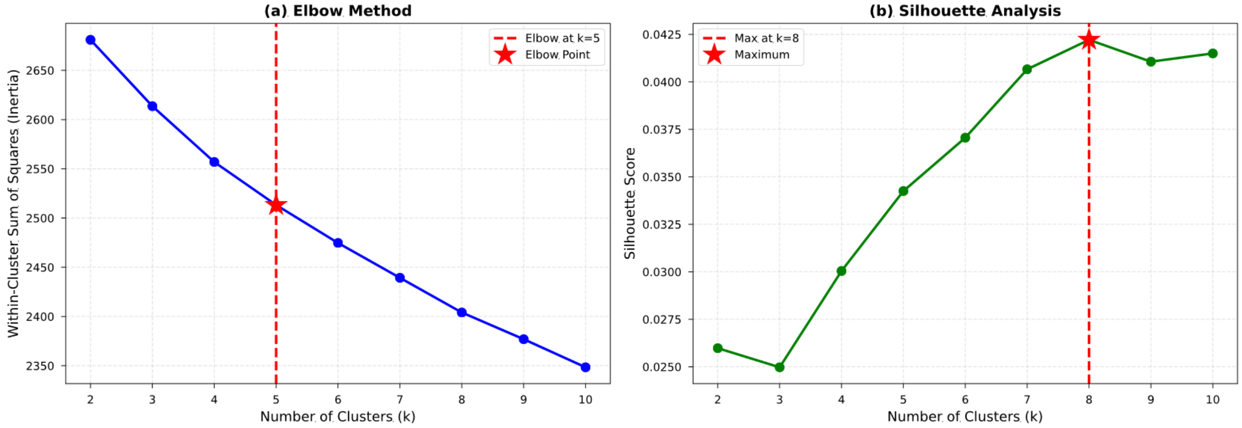


Figure 2: Determination of Optimal Cluster Number using (a) Elbow Method and (b) Silhouette Analysis.

A critical step in this phase was the determination of the optimal number of clusters, k. A value of k that is too small may group dissimilar hazards together, while a value that is too large may create overly specific and sparsely populated clusters. To identify a statistically appropriate value, two common techniques were employed: the Elbow Method and Silhouette Analysis. The Elbow Method involves plotting the within-cluster sum of squares (inertia) for a range of k values and identifying the "elbow" point where the rate of decrease in inertia sharply diminishes. Silhouette Analysis measures how well-separated the resulting clusters are by calculating a score for each data point based on its cohesion within its cluster and its separation from other clusters. The value of k that maximizes the average silhouette score is considered optimal. The graphical results of these analyses, presented in Figure 2, provide a basis for selecting the final number of clusters used in the subsequent optimization phase. The outcome of this process is a dataset segmented into distinct hazard clusters, with each intervention assigned to a group based on its situational context.

3.5 Tri-Objective Optimization

Following the partitioning of interventions into distinct hazard clusters, a tri-objective optimization process was executed for each cluster. The objective of this phase was to identify portfolios of safety interventions that represent the optimal trade-offs between the three competing criteria: minimizing risk, minimizing complexity, and maximizing generalizability. The inherent conflicts among these objectives—for instance, a highly generalizable intervention may not address the highest-risk scenarios, or a low-complexity solution might offer only marginal risk reduction—define this as a multi-objective optimization problem (MOOP). The goal is not to find a single "best" solution but to identify the set of non-dominated solutions, known as the Pareto front, which provides decision-makers with a range of optimal choices (Kaveh et al., 2021).

The Non-dominated Sorting Genetic Algorithm III (NSGA-III) was selected to solve this optimization problem. NSGA-III is a multi-objective evolutionary algorithm particularly well-suited for problems with three or more objectives, often referred to as many-objective optimization (Pham and Khoi, 2025). Unlike its predecessor, NSGA-II, which can struggle with maintaining population diversity in higher-dimensional objective spaces, NSGA-III employs a reference-point-based selection mechanism. This approach ensures that solutions are well-distributed across the entire Pareto surface, providing a comprehensive representation of the available trade-offs. The algorithm was configured with a population size of 100 and set to run for 100 generations to ensure convergence toward a stable Pareto front.

Within the optimization framework, a multi-objective problem was defined where each potential solution in the population represents a portfolio of selected interventions. A solution is represented as a binary vector, with each element corresponding to an intervention within the specific hazard cluster being analyzed; a value of '1' indicates selection, and '0' indicates non-selection. The evaluation of each solution involved calculating the aggregate performance across the three objectives for the selected portfolio. The mean values of obj_{risk} and $obj_{complexity}$ for the chosen interventions were calculated as the first two objective scores to be minimized. To align with the minimization framework of the algorithm, the third objective was the negative of the mean $obj_{generalizability}$, effectively maximizing the generalizability score. In cases where a solution selected no interventions, a penalty was applied by assigning worst-case values for each objective to guide the algorithm away from empty solution sets.

This optimization procedure was performed independently for each of the hazard clusters. The final output for each cluster is a set of non-dominated solution portfolios and their corresponding objective values, which collectively form the 3D Pareto front. Each point on this front represents an optimal portfolio where no single objective can be improved without degrading at least one of the other two objectives. This set of solutions provides a quantitative basis for strategic decision-making in safety management, allowing practitioners to select an intervention portfolio that best aligns with their specific constraints and priorities.

3.6 Data-Driven Validation

The final phase of the methodology was designed to validate the framework's outputs and the integrity of the novel metrics using internal, data-driven techniques, thereby avoiding reliance on subjective expert evaluation. This phase consisted of two primary components: an assessment of the quality of the generated Pareto fronts and a statistical validation of the Generalizability metric.

First, the quality of the solution space for each hazard cluster was quantified by calculating the hypervolume indicator for its respective Pareto front. The hypervolume is a standard performance metric in multi-objective optimization used to measure the size of the objective space dominated by a set of non-dominated solutions relative to a defined reference point (Pham and Khoi, 2025). A larger hypervolume signifies a superior Pareto front, indicating that the solutions offer a better and more diverse range of trade-offs. For this analysis, the reference point was established using the least desirable values for each objective (maximum risk, maximum complexity, and minimum generalizability). This calculation provides a quantitative and objective basis for comparing the quality of the decision-making landscapes across different hazard clusters.

Second, the construct validity of the Generalizability metric (O_G) was assessed through correlation analysis against intrinsic linguistic features of the safe alternative descriptions. This process aimed to confirm that the data-driven semantic metric corresponds to observable linguistic properties, a method used to validate proxies derived from textual data (Jallan and Ashuri, 2020). The guiding hypothesis was that a more general safety intervention would be articulated using more concise and common vocabulary than a highly specific one. To test this hypothesis, two linguistic features were extracted: (1) word count, to measure the conciseness of the intervention text, and (2) vocabulary commonality, a score derived from term frequency analysis to quantify the usage of high-frequency words from the corpus.

Statistical methods were then applied to measure the association between the Generalizability scores and these linguistic features. The Spearman's rank correlation coefficient (ρ) was used to assess the relationship between generalizability and word count, as it is not constrained by assumptions of linearity. A significant negative correlation was hypothesized, suggesting that higher generalizability scores are associated with shorter descriptions. The Pearson correlation coefficient (r) was employed to evaluate the linear relationship between generalizability and vocabulary commonality, with a significant positive correlation expected. The establishment of these statistical relationships provides empirical, data-driven evidence supporting the validity of the Generalizability metric as a proxy for the broad applicability of an intervention.

4. Results and Analysis

4.1 Descriptive Statistics of the Dataset

Following the automated data structuring and filtering protocols, the final dataset for analysis consisted of 3,426 valid intervention pairs derived from OSHA accident narratives. For each intervention, the three primary objectives—Risk, Complexity, and Generalizability—were quantified based on the methodological framework. A statistical summary of these objectives is presented in Table 1, providing a comprehensive overview of the dataset's characteristics before optimization.

Table 1: Summary Statistics of the Intervention Dataset and Calculated Objectives.

Metric	Value
Total Interventions	3,426
Risk (Severity)	
Mean (SD)	6.37 (1.26)
Range	[3, 10]
Complexity (Semantic Friction)	
Mean (SD)	0.691 (0.190)
Range	[0.067, 1.166]
Generalizability	
Mean (SD)	0.167 (0.046)
Range	[0.025, 0.284]
Optimization Outputs	
Number of Clusters Identified	5
Avg. Pareto Solutions per Cluster (SD)	31.6 (11.5)

The Risk objective, which corresponds to the LLM-assigned severity score, ranged from 3 to 10. The mean severity was 6.37 with a standard deviation of 1.26, indicating that the dataset is predominantly composed of incidents with moderate to high severity outcomes, reflecting the nature of incidents typically warranting formal investigation. The Complexity objective, calculated as the semantic friction between unsafe action and safe alternative pairs, had a mean value of 0.691 and a standard deviation of 0.190. This suggests a notable conceptual distance between the average unsafe state and its corresponding safe procedure, implying that many interventions require a non-trivial procedural or behavioral adjustment. The

Generalizability scores, representing the average semantic similarity of a safe alternative to all others, were concentrated in a relatively narrow band from 0.025 to 0.284, with a mean of 0.167 and a standard deviation of 0.046. This indicates a spectrum from highly specific to more broadly applicable interventions.

Figure 3 illustrates the distributions and pairwise correlations of the three objectives. The histogram for the Risk objective (Figure 3a) is left-skewed, confirming the dataset's weighting towards more severe incidents. The distributions for Complexity (Figure 3b) and Generalizability (Figure 3c) approximate normal distributions. The scatter plots in Figure 3(d-f) reveal the relationships between the objectives. There is a negligible correlation between Risk and Complexity (r = 0.089), indicating that higher severity incidents do not necessarily require more complex interventions. Similarly, the correlation between Risk and Generalizability (r = -0.016) is insignificant. A weak positive correlation is observed between Complexity and Generalizability (r = 0.126). The absence of strong correlations among the three objectives confirms their distinctness as independent criteria. This lack of multicollinearity validates their suitability for a multi-objective optimization framework, where the goal is to explore the trade-offs between competing, rather than redundant, performance dimensions.

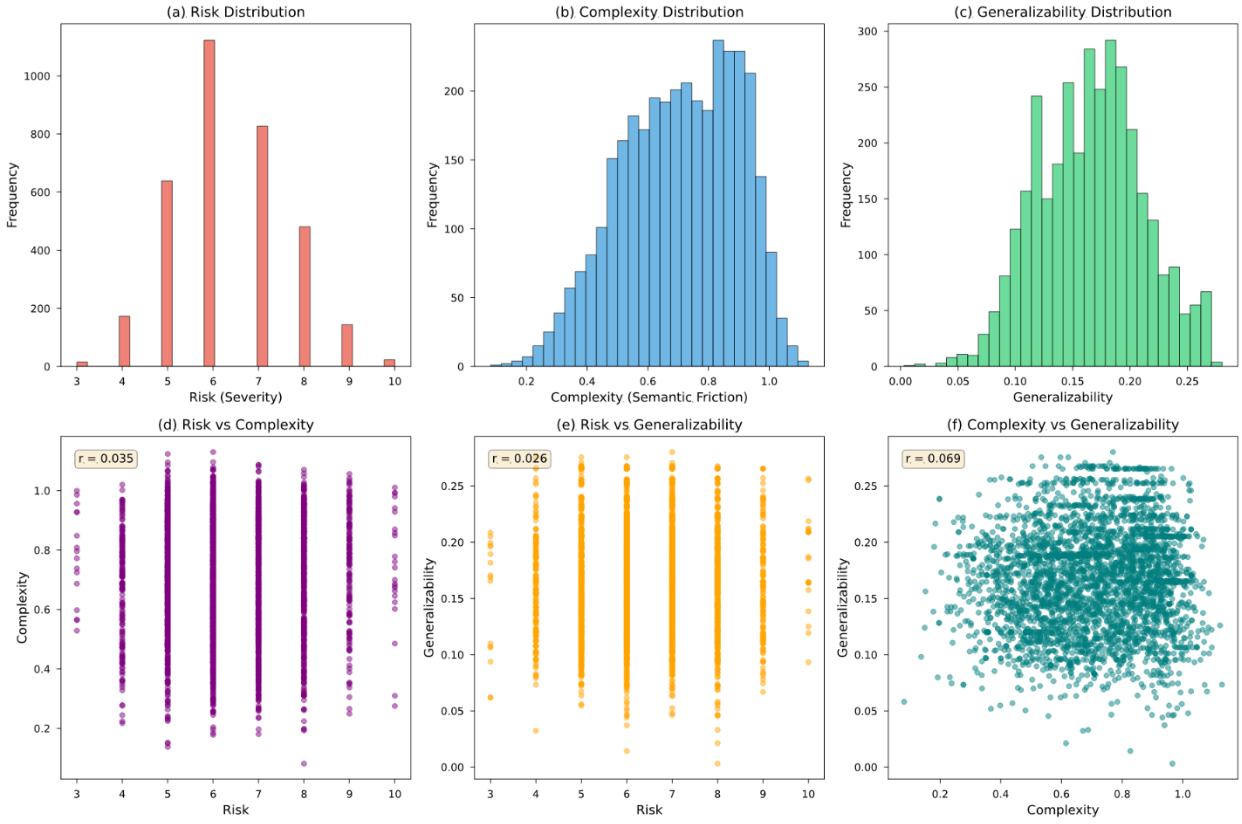


Figure 3: Distributions and Pairwise Correlations of the Three Objectives.

4.2 Hazard Clustering Results

To facilitate a context-specific optimization, the dataset of safety interventions was partitioned into homogenous groups using the k-means clustering algorithm. This unsupervised learning technique was applied to the sentence embeddings of the state keywords associated with each incident, thereby grouping interventions based on the similarity of their situational and environmental contexts (Oh et al., 2025). The process ensures that the subsequent optimization is performed on coherent subsets of data, leading to more tailored and actionable results.

The determination of the optimal number of clusters, k, was guided by the Elbow Method and Silhouette Analysis, as detailed in the methodology (Section 3.4 and Figure 2). The analysis indicated that k=5 provided a robust balance between intra-cluster cohesion and inter-cluster separation. The dataset of 3,426 interventions was subsequently partitioned into five distinct hazard clusters.

To interpret the thematic content of each cluster, the top 10 most frequent keywords from the unsafe action and safe alternative descriptions within each group were extracted. Table 2 details the distribution of interventions, the identified semantic theme, and the mean objective values for each cluster. The analysis reveals distinct and coherent hazard categories:

- **Cluster 0 (n=444): Environmental and Electrical Hazards.** Dominated by terms such as 'heat exhaustion', 'arc flash', and 'dehydration', this cluster represents incidents related to thermal stress and electrical events. It is the smallest cluster and exhibits the lowest average Generalizability (0.129).
- **Cluster 1 (n=966): Machinery-Related Caught-in/between Hazards.** This is the largest cluster, characterized by keywords like 'finger was caught', 'excavator', 'circular saw', and 'crane'. It represents incidents where workers are caught in or between moving equipment parts. Notably, it has the lowest average Risk (6.14) and the highest average Generalizability (0.175), suggesting these are frequent but less severe incidents with broadly applicable solutions.
- **Cluster 2 (n=662): Slips, Trips, and Falls (STFs).** With keywords such as 'slipped and fell', 'tripped', and 'ladder', this cluster clearly represents fall-related incidents from the same level or from low heights (e.g., ladders).
- **Cluster 3 (n=752): Struck-by Hazards in Traffic/Work Zones.** This cluster is defined by terms like 'dump truck', 'work zone', and 'struck by a vehicle'. It encapsulates incidents involving moving vehicles and traffic management, and is associated with

the highest average Risk (6.60) and Complexity (0.737), indicating these are severe incidents requiring significant procedural changes.

- **Cluster 4 (n=602): Excavation and Trenching Hazards.** Characterized by keywords such as 'excavator bucket', 'trench box', and 'excavation', this cluster pertains to risks associated with subterranean work and heavy earthmoving equipment.

These distinct thematic profiles confirm that the clustering process successfully segmented the dataset into meaningful hazard categories. This structured segmentation provides a robust basis for the tailored, cluster-specific optimization performed in the subsequent phase.

Table 2: Hazard Cluster Analysis Results, including Semantic Theme, Cluster Size, and Mean Objective Values.

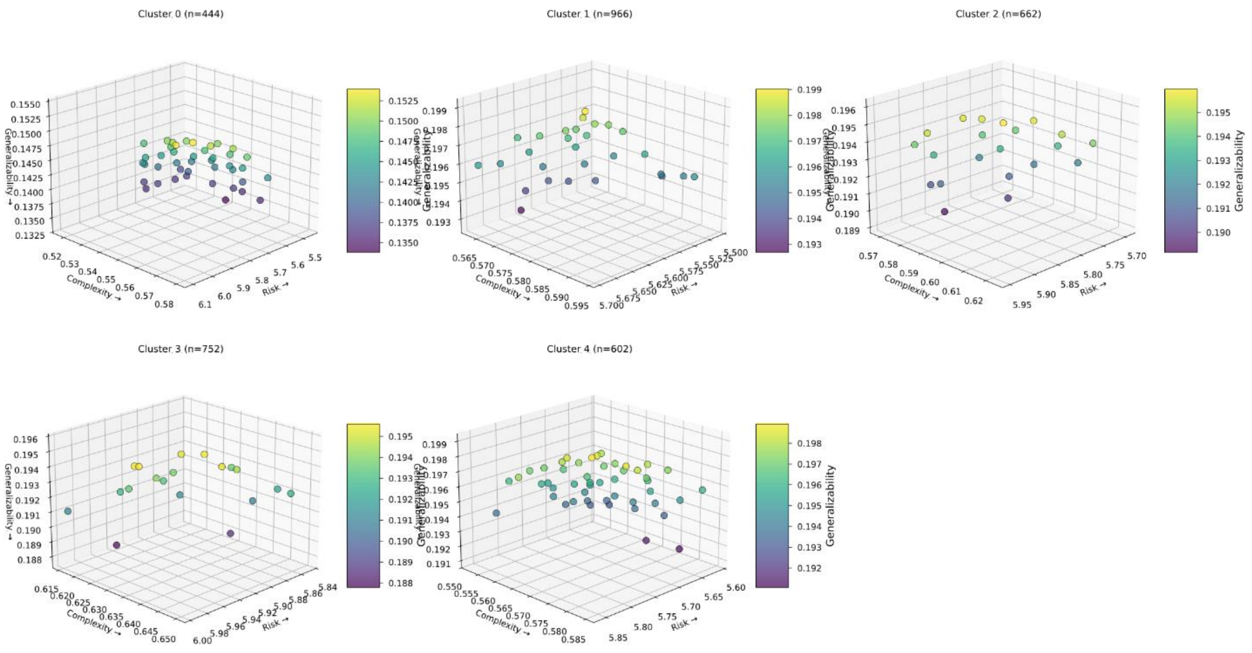
Hazard Cluster	Semantic Theme	Number of Interventions (n)	% of Total	Mean Risk (Severity)	Mean Complexity (Semantic Friction)	Mean Generalizability
0	Environmental & Electrical Hazards	444	13.0%	6.45	0.670	0.129
1	Machinery-Related Caught-in/between Hazards	966	28.2%	6.14	0.672	0.175
2	Slips, Trips, and Falls (STFs)	662	19.3%	6.43	0.693	0.172
3	Struck-by Hazards in Traffic/Work Zones	752	22.0%	6.60	0.737	0.169
4	Excavation & Trenching Hazards	602	17.5%	6.34	0.679	0.172
Total		**3,426**	**100%**	**6.37**	**0.691**	**0.167**

4.3 Optimization and Pareto Fronts

Upon partitioning the dataset into five distinct hazard clusters, the NSGA-III algorithm was executed independently for each cluster. The objective was to identify the set of non-dominated portfolios of safety interventions, which collectively form the Pareto front for each specific hazard context. These fronts represent the optimal trade-offs among the three competing objectives: minimizing average portfolio risk, minimizing average portfolio complexity, and maximizing average portfolio generalizability.

The resulting 3D Pareto fronts for all five hazard clusters are visualized in Figure 4. Each point on a given surface represents an optimal intervention portfolio where no single objective can be improved without degrading at least one of the other objectives. The visual morphology of the Pareto fronts varies across the clusters, indicating that the decision-making landscape is context-dependent. For instance, the surfaces for Clusters 1 and 4 appear more expansive and evenly distributed compared to the more compressed surface of Cluster 3. This suggests that the range and diversity of optimal solutions differ significantly depending on the nature of the hazard group being analyzed. The number of non-dominated solutions found for each cluster ranged from 19 (Cluster 3) to 48 (Cluster 4), providing a varied set of optimal choices for decision-makers.

To illustrate the practical application of these results, a detailed analysis of representative solutions from the Pareto front of a selected cluster is presented in Table 3. Four distinct portfolios are highlighted: one that prioritizes minimizing risk, one that prioritizes minimizing complexity, one that maximizes generalizability, and a 'Balanced' solution. The balanced solution, also known as the knee point, is identified as the portfolio on the Pareto front that is geometrically closest to the ideal point (minimum risk, minimum complexity, maximum generalizability) in the normalized objective space.



Cluster 0 (n=444)

Cluster 1 (n=966)

Cluster 2 (n=662)

Cluster 3 (n=752)

Cluster 4 (n=602)

Figure 4: 3D Pareto Fronts of Optimal Safety Interventions for Each Hazard Cluster.

For example, the portfolio focused on 'Min Risk' achieves the lowest possible average risk score for that cluster but may compromise on complexity or generalizability. Conversely, the 'Max Generalizability' portfolio identifies a set of interventions with the broadest applicability, which could be valuable for developing foundational safety training, but this portfolio may not address the most severe or complex hazards within the cluster. The 'Balanced' portfolio represents a compromise solution, offering a moderate trade-off across all three objectives. This analysis demonstrates how the Pareto front provides a spectrum of quantitatively distinct, optimal strategies. Instead of prescribing a single "best" answer, the framework equips safety managers with a portfolio of choices, allowing them to select an intervention strategy that aligns with their specific organizational priorities, resource constraints, and risk appetite (Ghanbari et al., 2024).

Table 3: Representative Pareto-Optimal Solutions for Cluster 1

Solution Type	Risk	Complexity	Generalizability
Minimum Risk	5.510	0.582	0.194
Minimum Complexity	5.659	0.563	0.196
Maximum Generalizability	5.619	0.579	0.199
Balanced (Knee)	5.510	0.582	0.194

4.4 Comparative Analysis of Clusters

To quantitatively assess and compare the quality of the solution spaces generated for each hazard cluster, the hypervolume indicator was calculated for each of the five Pareto fronts. The hypervolume metric provides a comprehensive measure of a Pareto front's performance by quantifying the volume of the objective space that is dominated by the set of non-dominated solutions, relative to a pre-defined nadir or worst-case reference point (Pham and Khoi, 2025). A larger hypervolume indicates a superior set of solutions, suggesting that the Pareto front offers a wider and more dominant range of trade-offs. This analysis allows for an objective comparison of the decision-making landscapes across different hazard contexts.

The results of the hypervolume calculations are summarized in Table 4. The analysis reveals notable differences in the quality of the Pareto fronts across the five clusters. Clusters 1 and 4 achieved the highest hypervolume values, both at 0.390, indicating that the intervention sets within these hazard groups yield the most favorable and diverse range of optimal solutions. This suggests that for these types of hazards, decision-makers have access to a richer set of trade-off opportunities. Following closely, Cluster 2 had a hypervolume of 0.363, and Cluster 0 had a value of 0.332.

Table 4: Optimization Summary by Hazard Cluster

Hazard Cluster	**Interventions (n)**	**Pareto Front Size**	**Hypervolume**	**Mean Risk**	**Mean Complexity**	**Mean Generalizability**
0	444	42	0.332	5.72	0.549	0.144
1	966	28	0.390	5.59	0.574	0.196
2	662	21	0.363	5.82	0.592	0.193
3	752	19	0.311	5.93	0.634	0.193
4	602	48	0.390	5.71	0.567	0.196

In contrast, Cluster 3 exhibited the lowest hypervolume of 0.311. This finding is consistent with the initial characterization of this cluster in Table 2, where it was identified as having the highest average Risk (6.60) and Complexity (0.737). The lower hypervolume for Cluster 3 suggests a more constrained decision-making environment. The optimal solutions for this hazard group are less dominant in the objective space, meaning the trade-offs are more severe, and there are fewer high-performing options available. This quantitative comparison underscores the value of the clustering phase, as it reveals that the nature and quality of optimal safety strategies are not uniform but are instead dependent on the specific hazard context being addressed.

4.5 Validation of the Generalizability Metric

A crucial component of this research is the introduction of Generalizability O_G as a quantifiable objective. As this is a novel, data-driven proxy, its validity must be established to ensure its meaningfulness within the optimization framework. This was accomplished through an internal validation process that assessed the correlation between the calculated Generalizability scores and intrinsic linguistic features of the safe_alternative texts. This approach is grounded in the hypothesis that a more general intervention, applicable across a wider range of contexts, will be described using more common and fundamental language than a highly specific, context-dependent action (Jallan and Ashuri, 2020).

The statistical validation was performed by correlating the Generalizability score of each intervention with two extracted linguistic features: word count and vocabulary commonality. The results of this analysis are presented in Figure 5. The relationship between Generalizability and word count was examined using Spearman's rank correlation, which yielded a weak but statistically significant positive correlation ($\rho = 0.204$, $p < 0.0001$). This suggests that interventions with higher generalizability scores tend to be described using slightly more words, potentially because foundational safety principles require more descriptive language than concise, specific directives. More revealing was the relationship with vocabulary commonality, which was assessed using the Pearson correlation coefficient. This test revealed a moderate and statistically significant positive correlation ($r = 0.437$, $p < 0.0001$). This result strongly supports the underlying hypothesis, indicating that interventions with higher Generalizability scores are composed of words that appear more frequently throughout the entire corpus of safety actions.

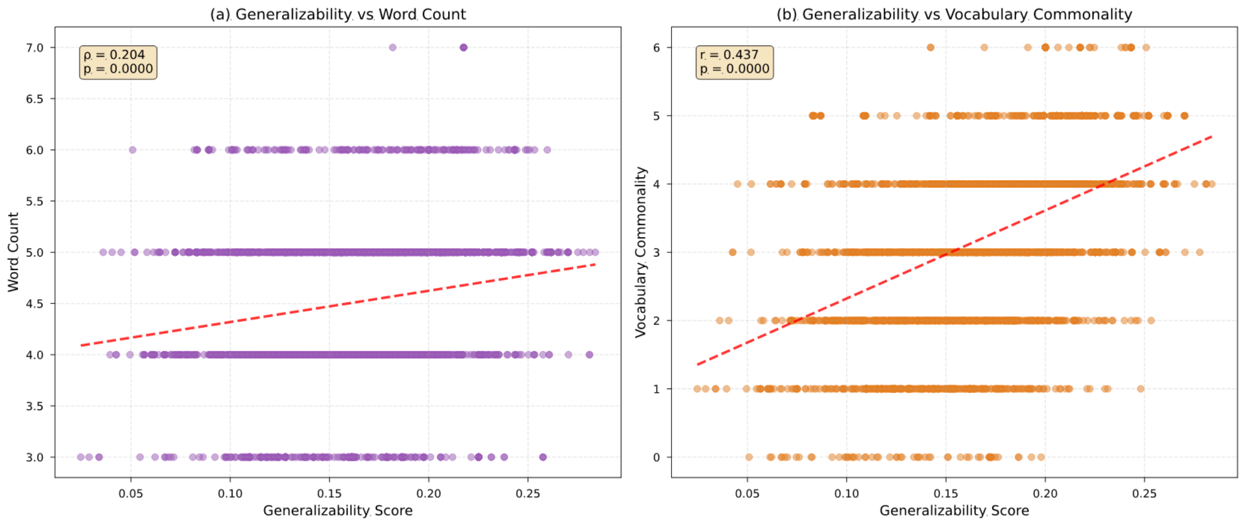


Figure 5: Validation of the Generalizability Metric via Correlation with (a) Word Count and (b) Vocabulary Commonality.

This quantitative finding is further supported by a qualitative and comparative analysis of interventions from the extreme quartiles of the Generalizability distribution, summarized in Table 5. Interventions in the highest quartile (Q3) of generalizability scores have a significantly higher average vocabulary commonality score (4.17) compared to those in the lowest quartile (Q1) (2.53). Examples of high-generalizability interventions include broad safety principles such as "Keep hands off moving equipment" and "Maintain safe clearance from machinery." In contrast, low-generalizability interventions are typically highly specific and relational, such as "Communicate with coworker" and "Coordinate with operator." These examples are semantically distinct and their applicability is confined to very particular scenarios involving interpersonal coordination.

Table 5: Low vs High Generalizability Interventions

Metric	**Low Generalizability (Q1)**	**High Generalizability (Q3)**
Count	857	871
Avg. Risk (Severity)	6.26	6.48
Avg. Complexity	0.680	0.707
Avg. Word Count	4.39	4.81
Avg. Vocab Commonality	2.53	4.17

The consistent and statistically significant relationship between the Generalizability metric and these linguistic properties, combined with the clear conceptual differences observed in the examples, provides robust, data-driven evidence for the construct validity of the metric. This validation supports its use as a reliable proxy for an intervention's breadth of applicability, confirming its suitability as a distinct and meaningful third objective in the optimization framework.

5. Discussion

5.1 Implications of the Tri-Objective Framework for Safety Management

The primary implication of this research is the shift from a single-point or bi-objective decision-making paradigm to a more holistic, portfolio-based approach for selecting safety interventions. The framework's ability to automatically group interventions into thematically coherent hazard clusters—such as 'Slips, Trips, and Falls' (Cluster 2) and 'Struck-by Hazards in Traffic/Work Zones' (Cluster 3)—provides a context-specific foundation for decision-making. Traditional risk management often prioritizes interventions that address the most severe or frequent hazards (Aladağ, 2023). This framework expands that perspective by revealing that the optimal intervention portfolio is not a single point but a spectrum of choices along a context-specific Pareto front, each with a unique balance of risk mitigation, implementation complexity, and broad applicability.

This multi-faceted view empowers safety managers to make more strategic, context-aware decisions. For example, a company specializing in roadwork would focus on the Pareto front for Cluster 3. The finding that this cluster has the highest average Risk and Complexity, along with the lowest hypervolume, quantitatively confirms that it represents a challenging decision environment with more severe trade-offs. Conversely, a general contractor might focus on interventions from Cluster 1 ('Machinery-Related Caught-in/between Hazards'), which, having the lowest average risk and highest generalizability, represents a domain with more foundational and broadly applicable safety solutions. This allows organizations to tailor their safety strategies based on their specific operational focus, moving beyond generic, one-size-fits-all safety plans (Nyqvist et al., 2024).

5.2 Interpretation of the Novel Metrics

A core contribution of this work is the operationalization of two novel metrics, Complexity and Generalizability, using semantic analysis to enrich the decision-making process for safety interventions. The first metric, Complexity, is modeled as "Semantic Friction," which is the cosine distance between the textual embeddings of an unsafe action and its corresponding safe alternative. This metric serves as a non-monetary proxy for the cognitive and operational effort required to implement a change. A high semantic friction score suggests that the safe alternative is conceptually distant from the unsafe action. For example, the transition from an unsafe action like "standing on guardrail" to a safe alternative such as "using fall protection system" would likely have a high semantic friction, as it involves a fundamental shift in procedure and the introduction of new equipment. In contrast, a transition from "guardrail missing" to "install guardrail" represents a direct correction and would have a low semantic friction. This metric provides a method for quantifying the often-overlooked implementation difficulties and "soft costs" associated with training and procedural change, which can be critical for assessing the feasibility of an intervention, particularly in high-pressure or resource-constrained environments (Wen and Hu, 2025).

The second metric, Generalizability, is designed to quantify the extent to which a safe intervention represents a broad, foundational safety principle rather than a narrow, context-specific directive. It is calculated as the average cosine similarity of a given safe alternative's embedding to the embeddings of all other safe alternatives in the dataset. The validation analysis in Section 4.5 confirmed the construct validity of this metric, demonstrating a significant positive correlation with the use of common vocabulary. For instance, an intervention like "Maintain safe clearance from machinery" has high generalizability because it is a principle applicable to numerous types of equipment and work scenarios. Conversely, an intervention such as "Communicate with coworker" has low generalizability as its relevance is confined to specific, multi-person tasks. The results show that Cluster 1 ('Machinery-Related Caught-in/between Hazards') has the highest average generalizability, suggesting that this hazard category is governed by a set of foundational, transferable safety principles. In contrast, Cluster 0 ('Environmental & Electrical Hazards') has the lowest, implying its interventions are more specialized. This distinction is critical for designing effective safety management systems, allowing for a hierarchical approach where general principles are reinforced through company-wide training, while specific directives are reserved for targeted toolbox talks or job hazard analyses.

5.3 Practical Application of Pareto-Optimal Solutions

The 3D Pareto fronts generated through the optimization process provide a comprehensive, multi-dimensional view of the optimal trade-offs among risk, complexity, and generalizability. While these 3D visualizations (Figure 4) are effective for illustrating the overall structure of the solution space, their direct application in a practical decision-making context can be challenging. To enhance the usability of the framework's outputs for safety managers, the 3D fronts can be projected into a series of 2D scatter plots, as exemplified in Figure 6. These plots offer an intuitive interface for navigating the complex trade-offs inherent in intervention selection.

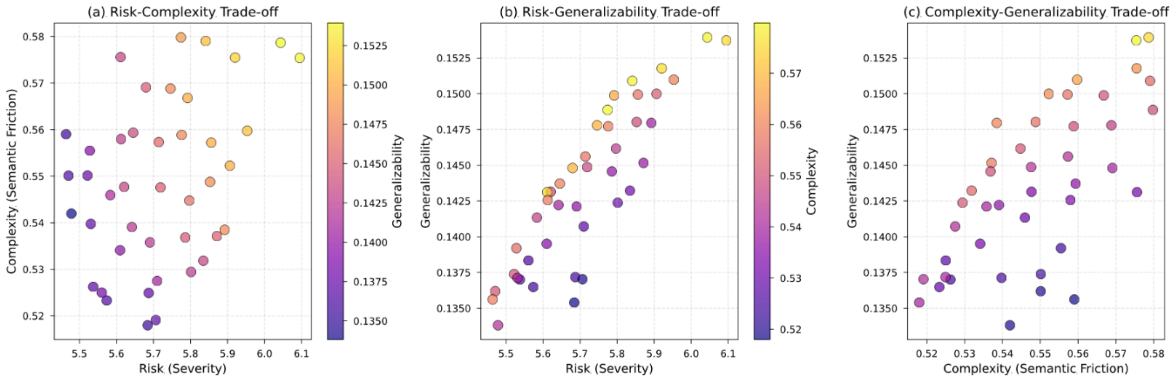


Figure 6: 2D Projections of Pareto Front Trade-offs for a Representative Cluster.

Each 2D projection visualizes the direct relationship between two of the objectives, while using a color gradient to represent the value of the third. This allows for a more targeted analysis based on specific organizational priorities. For instance, consider a safety manager for a company specializing in road construction, whose primary concerns align with the thematic content of Cluster 3 ('Struck-by Hazards in Traffic/Work Zones'). The results from the optimization indicate that this is a high-risk, high-complexity domain with a constrained solution space, as evidenced by its low hypervolume.

When this manager examines the Risk-Complexity trade-off plot for Cluster 3 (Figure 6a), they can visually assess the direct cost in terms of implementation difficulty for each unit of risk reduction. Solutions located on the "knee" of this curve would represent a balanced compromise between the two plotted objectives. The color of each point, indicating its Generalizability score, adds another layer of critical information. This enables the decision-maker to determine whether a low-risk, low-complexity solution is a specific, tactical fix (low generalizability) or a more broadly applicable principle (high generalizability), which might be a better long-term investment for policy and training.

Alternatively, if the manager's objective is to establish a new set of site-wide safety protocols, the Risk-Generalizability plot (Figure 6b) becomes the primary tool. This visualization would guide them toward solutions in the upper-left quadrant, representing portfolios with high generalizability and low risk, even if they come with a moderate level of complexity (indicated by the color gradient). The ability to visually interrogate these trade-offs transforms the abstract output of the evolutionary algorithm into an interactive decision-support tool. This aligns with the principles of effective project portfolio management, where decision-makers must balance multiple competing objectives to align with strategic organizational goals (Ghanbari et al., 2024). The framework, therefore, does not prescribe a single "correct" answer but rather provides a curated set of optimal, quantitatively-backed options from which an informed, strategic choice can be made.

5.4 Contribution, Limitations, and Practicality

This study makes a significant contribution to the body of knowledge by successfully operationalizing "Generalizability" and "Complexity" as quantifiable objectives in safety management, moving beyond conventional bi-objective models. The framework provides a structured, data-driven methodology for analyzing large volumes of unstructured safety data and translating it into a portfolio of optimized intervention strategies. The use of an LLM for data structuring and advanced NLP techniques for metric formulation demonstrates a novel integration of AI in construction safety analytics.

However, several limitations must be acknowledged. First, the framework is contingent on the quality and content of the source data. The OSHA narratives used in this study primarily document incidents that have already occurred and may not capture the full spectrum of near-misses or proactive safety observations. Second, the metrics for Complexity and Generalizability are proxies derived from semantic properties of text. While validated internally, their correspondence to real-world implementation effort and applicability requires further validation through field studies and expert evaluation. Third, the framework does not currently incorporate financial cost, a critical factor in any real-world decision-making process. The exclusion of cost was a deliberate choice to first establish the validity of the novel semantic dimensions, but its integration is a necessary next step for a comprehensive model. Despite these limitations, the proposed framework demonstrates high practical potential. It serves as a proof-of-concept for a sophisticated decision-support system (DSS) for safety management (Bilgin et al., 2023). Future development could see this framework evolve into a software tool that allows safety managers to upload their own incident reports, automatically generate Pareto-optimal intervention portfolios, and interactively explore the trade-offs to select strategies that best fit their project's unique risk profile and organizational goals.

6. Conclusion

This paper presented a tri-objective optimization framework to support the selection of construction safety interventions. The methodology addressed the limitations of existing bi-objective models by introducing and quantifying two novel criteria, Complexity and Generalizability, alongside the conventional metric of Risk. The framework leveraged a large language model to structure unstructured accident narratives from OSHA into a machine-readable format. This structured data was then used to formulate the three competing objectives. Interventions were subsequently grouped into thematically coherent hazard clusters using unsupervised machine learning before a multi-objective evolutionary algorithm, NSGA-III, was applied to identify Pareto-optimal portfolios of interventions for each cluster.

The analysis of infrastructure projects intervention pairs confirmed that Risk, Complexity, and Generalizability are distinct and weakly correlated objectives, validating the necessity of a multi-objective approach for this problem. The hazard clustering process successfully segmented the dataset into five meaningful categories, including 'Slips, Trips, and Falls' and 'Struck-by Hazards in Traffic/Work Zones,' each exhibiting unique statistical profiles and decision landscapes. The optimization process generated distinct 3D Pareto fronts for each cluster, demonstrating that the nature of optimal trade-offs is context-dependent. The quality of these solution spaces, as quantified by the hypervolume indicator, varied across clusters, with certain hazard types offering a more diverse and dominant set of optimal solutions than others. The internal validation process confirmed the construct validity of the "Generalizability" metric, establishing a significant statistical correlation with linguistic features indicative of broad applicability.

In conclusion, this research successfully developed and applied a framework that advances safety decision-making from a bi-objective to a tri-objective paradigm. The primary contribution is the formalization of "Generalizability" as a quantifiable decision criterion, which enables the prioritization of interventions that represent foundational safety principles. The framework provides a structured, data-driven methodology for transforming raw incident data into a portfolio of non-dominated, strategically diverse safety solutions. The resulting Pareto fronts equip safety managers with a quantitative tool to navigate the complex trade-offs between mitigating severe risks, reducing implementation complexity, and promoting broadly applicable safety practices, thereby supporting more nuanced and informed safety management strategies in the construction industry.

Data availability

The data is available upon request.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This research received no external funding.

References

Acharya, S., Kisi, K., Gautam, S.R., Mahmud, T. and Kayastha, R. (2025) 'A High-Performance Hybrid Transformer–LSTM–XGBoost Model for sEMG-Based Fatigue Detection in Simulated Roofing Postures', *Buildings*, 15(17), art. no. 3005. doi:10.3390/buildings15173005.

Afshar, A. and Fathi, H. (2009) 'Fuzzy multi-objective optimization of finance-based scheduling for construction projects with uncertainties in cost', *Engineering Optimization*, 41(11), pp. 1063-1080. doi:10.1080/03052150902943004.

Akinosho, T.D., Oyedele, L.O., Bilal, M., Ajayi, A.O., Delgado, M.D., Akinade, O.O. and Ahmed, A.A. (2020) 'Deep learning in the construction industry: A review of present status and future innovations', *Journal of Building Engineering*, 32, art. no. 101827. doi:10.1016/j.jobe.2020.101827.

Aladağ, H. (2023) 'Assessing the Accuracy of ChatGPT Use for Risk Management in Construction Projects', *Sustainability (Switzerland)*, 15(22), art. no. 16071. doi:10.3390/su152216071.

Alkaissy, M., Arashpour, M., Ashuri, B., Bai, Y. and Hosseini, M.R. (2020) 'Safety management in construction: 20 years of risk modeling', *Safety Science*, 129, art. no. 104805. doi:10.1016/j.ssci.2020.104805.

Alkaissy, M., Arashpour, M., Li, H., Alaghmand, S. and Nezamian, A. (2022) 'Quantitative Analysis of Safety Risks and Relationship with Delayed Project Completion Times', *Risk Analysis*, 42(3), pp. 580-591. doi:10.1111/risa.13778.

Ballal, S., Patel, K.A. and Patel, D.A. (2024) 'Enhancing Construction Site Safety: Natural Language Processing for Hazards Identification and Prevention', *Journal of Engineering, Project, and Production Management*, 14(2), art. no. 0014. doi:10.32738/JEPPM-2024-0014.

Bilgin, G., Dikmen, I., Birgönül, M.T. and Özorhon, B. (2023) 'A Decision Support System for Project Portfolio Management in Construction Companies', *International Journal of Information Technology and Decision Making*, 22(2), pp. 705-735. doi:10.1142/S0219622022500821.

Chen, W., Li, R. and Yao, Y. (2022) 'Return and Volatility Spillovers among Sector Indexes in Shanghai-Shenzhen-Hong Kong Stock Markets: Evidence from the Time and Frequency Domains', *Emerging Markets Finance and Trade*, 58(13), pp. 3840-3852. doi:10.1080/1540496X.2022.2072204.

Cheng, M.-Y., Sholeh, M.N. and Poetra, B.A. (2025) 'Predicting the construction cost-time tradeoff using optimized hybrid deep learning for risk preference decision making', *International Journal of Construction Management*. doi:10.1080/15623599.2025.2574070.

Chenya, L., Aminudin, E., Mohd, S. and Yap, L.S. (2022) 'Intelligent Risk Management in Construction Projects: Systematic Literature Review', *IEEE Access*, 10, pp. 72936-72954. doi:10.1109/ACCESS.2022.3189157.

Gadekar, H. and Bugalia, N. (2023) 'Automatic classification of construction safety reports using semi-supervised YAKE-Guided LDA approach', *Advanced Engineering Informatics*, 56, art. no. 101929. doi:10.1016/j.aei.2023.101929.

Ghanbari, M., Olaikhan, A.A.J. and Skitmore, M. (2024) 'Enhancing project portfolio selection for construction holding firms: a multi-objective optimization framework with risk analysis', *Engineering, Construction and Architectural Management*. doi:10.1108/ECAM-05-2023-0532.

Gondia, A., Siam, A., El-Dakhakhni, W. and Nassar, A.H. (2020) 'Machine Learning Algorithms for Construction Projects Delay Risk Prediction', *Journal of Construction Engineering and Management*, 146(1), art. no. 04019085. doi:10.1061/(ASCE)CO.1943-7862.0001736.

Hu, Y., Zhang, M., Ren, W., Li, H., Zhang, J., Wang, L. and Shuai, S. (2025) 'Detecting Loss of Balance in Construction Workers Using Millimeter Wave Sensing', *Journal of Construction Engineering and Management*, 151(12), art. no. 04025202. doi:10.1061/JCEMD4.COENG-17273.

Jallan, Y. and Ashuri, B. (2020) 'Text Mining of the Securities and Exchange Commission Financial Filings of Publicly Traded Construction Firms Using Deep Learning to Identify and Assess Risk', *Journal of Construction Engineering and Management*, 146(12), art. no. 04020137. doi:10.1061/(ASCE)CO.1943-7862.0001932.

Katooziani, A., Jeelani, I. and Gheisari, M. (2025) 'GPT Applications for Construction Safety: A Use Case Analysis', *Buildings*, 15(14), art. no. 2410. doi:10.3390/buildings15142410.

Kaveh, A., Rajabi, F. and Mirvalad, S. (2021) 'Many-objective optimization for construction project scheduling using non-dominated sorting differential evolution algorithm based on reference points', *Scientia Iranica*, 28(6 A), pp. 3112-3128. doi:10.24200/sci.2021.58952.5988.

Kaveh, A. and Rajabi, F. (2022) 'Fuzzy-multi-mode Resource-constrained Discrete Time-cost-resource Optimization in Project Scheduling Using ENSCBO', *Periodica Polytechnica Civil Engineering*, 66(1), pp. 50-62. doi:10.3311/PPci.19145.

Khodabakhshian, A., Puolitaival, T. and Kestle, L. (2023) 'Deterministic and Probabilistic Risk Management Approaches in Construction Projects: A Systematic Literature Review and Comparative Analysis', *Buildings*, 13(5), art. no. 1312. doi:10.3390/buildings13051312.

Khodabakhshian, A., Re Cecconi, F.R. and López Droguett, E. (2025) 'PROBABILISTIC RISK IDENTIFICATION AND ASSESSMENT MODEL FOR CONSTRUCTION PROJECTS USING ELICITATION BASED BAYESIAN NETWORK', *Journal of Information Technology in Construction*, 30, pp. 185-212. doi:10.36680/j.itcon.2025.009.

Koc, K., Ekmekcioğlu, Ö. and Gürgün, A.P. (2023) 'Developing a National Data-Driven Construction Safety Management Framework with Interpretable Fatal Accident Prediction', *Journal of Construction Engineering and Management*, 149(4), art. no. 04023010. doi:10.1061/JCEMD4.COENG-12848.

Liu, X., Xu, F., Zhang, Z. and Sun, K. (2025) 'Fall-portent detection for construction sites based on computer vision and machine learning', *Engineering, Construction and Architectural Management*, 32(3), pp. 1499-1521. doi:10.1108/ECAM-05-2023-0458.

Luong, L.D., Tran, D.-H. and Phong, P.T. (2019) 'Hybrid multiple objective evolutionary algorithms for optimising multi-mode time, cost and risk trade-off problem', *International Journal of Computer Applications in Technology*, 60(3), pp. 203-214. doi:10.1504/IJCAT.2019.100299.

Martin, H., James, J. and Chadee, A. (2025) 'Exploring Large Language Model AI tools in Construction Project Risk Assessment: Chat GPT Limitations in Risk Identification, Mitigation Strategies, and User Experience', *Journal of Construction Engineering and Management*, 151(9), art. no. 04025119. doi:10.1061/JCEMD4.COENG-16658.

Moon, S., Lee, G. and Chi, S. (2022) 'Automated system for construction specification review using natural language processing', *Advanced Engineering Informatics*, 51, art. no. 101495. doi:10.1016/j.aei.2021.101495.

Nguyen, D., Nguyen, N.-T., Tran, Q. and Tran, D.-H. (2023) 'Tradeoff different construction project goals in using a novel multi-objective sea horse algorithm', *Alexandria Engineering Journal*, 82, pp. 55-68. doi:10.1016/j.aej.2023.09.059.

Nyqvist, R., Peltokorpi, A. and Seppänen, O. (2024) 'Can ChatGPT exceed humans in construction project risk management?', *Engineering, Construction and Architectural Management*, 31(13), pp. 223-243. doi:10.1108/ECAM-08-2023-0819.

Oh, J., Jeong, J., Jeong, J., Yun, B., Mun, H. and Kumi, L. (2025) 'Establishment of Risk Management Groups in Construction Based on Workers' Age and Accident Probability Using Unsupervised Learning', *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 11(4), art. no. 04025055. doi:10.1061/AJRUA6.RUENG-1539.

Pham, V.H.S. and Khoi, L.N.Q. (2025) 'Balancing the trade-off between quad-factors in construction management: a opposition-based Giant Pacific Octopus optimizer method', *Cluster Computing*, 28(5), art. no. 297. doi:10.1007/s10586-024-04975-6.

Shehadeh, A. and Alshboul, O. (2025) 'Enhancing occupational safety in Construction: Predictive analytics using advanced ensemble machine learning algorithms', *Engineering Applications of Artificial Intelligence*, 159, art. no. 111761. doi:10.1016/j.engappai.2025.111761.

Shuang, Q., Liu, X., Wang, Z. and Xu, X. (2024) 'Automatically Categorizing Construction Accident Narratives Using the Deep-Learning Model with a Class-Imbalance Treatment Technique', *Journal of Construction Engineering and Management*, 150(9), art. no. 04024107. doi:10.1061/JCEMD4.COENG-14515.

Toğan, V., Mostofi, F., Tokdemir, O.B. and Kadioglu, F. (2025) 'Efficient Management of Safety Documents Using Text-Based Analytics to Extract Safety Attributes From Construction Accident Reports', *IEEE Access*, 13, pp. 99758-99777. doi:10.1109/ACCESS.2025.3576442.

Wang, S. (2025) 'Automated non-PPE detection on construction sites using YOLOv10 and transformer architectures for surveillance and body worn cameras with benchmark datasets', *Scientific Reports*, 15(1), art. no. 27043. doi:10.1038/s41598-025-12468-8.

Wen, H. and Hu, P. (2025) 'Architectural Gaps in Generative AI: Quantifying Cognitive Risks for Safety Applications', *AI (Switzerland)*, 6(7), art. no. 138. doi:10.3390/ai6070138.

Yoon, S., Chang, T. and Chi, S. (2024) 'Developing an Integrated Construction Safety Management System for Accident Prevention', *Journal of Management in Engineering*, 40(6), art. no. 04024051. doi:10.1061/JMENEA.MEENG-6074.

Zermane, A., Mohd Tohir, M.Z., Zermane, H., Baharudin, M.R. and Mohamed Yusoff, H. (2023) 'Predicting fatal fall from heights accidents using random forest classification machine learning model', *Safety Science*, 159, art. no. 106023. doi:10.1016/j.ssci.2022.106023.

Zhang, Q., Guo, H., Liao, P.-C., Fang, D. and Fu, M. (2022) 'Optimizing safety-measure combinations to address construction risks', *International Journal of Occupational Safety and Ergonomics*, 28(2), pp. 941-957. doi:10.1080/10803548.2020.1847500.

Zou, P.X.W., Zhang, G.K. and Wang, J. (2007) 'Understanding the key risks in construction projects in China', *International Journal of Project Management*, 25(6), pp. 601-614. doi:10.1016/j.ijproman.2007.03.001.