International Journal of BIM and Engineering Science

♦ ISSN: 2571-1075 Vol: 12, No. 02, 2026 ♦ PP. 32–38

Article DOI: https://doi.org/10.54216/IJBES.120205

FULL LENGTH ARTICLE

BIM-Integrated Semantic Risk Intelligence for Construction Safety Severity Prediction Using Incident Narratives and 4D Work-Zone Attributes

Esam El-Mekawy[1,*]

[1]School of Science, Engineering and Environment, University of Salford, UK

Email: esam.elmekawy@gmail.com

Received: December 06, 2025 Revised: January 10, 2026 Accepted: February 16, 2026 ⋆ Corresponding author

ABSTRACT

Construction safety management increasingly depends on the ability to connect static building information models with dynamic evidence from site operations. This paper proposes a BIM-integrated semantic risk intelligence model that translates accident narratives into work-zone risk indicators and uses them to infer safety severity. The model links textual incident evidence with BIM-relevant descriptors, including construction phase, spatial zone, temporary protection status, energy isolation, and proximity to safety constraints. A formal risk-scoring layer is combined with supervised severity learning to provide interpretable decision support for safety planning and 4D coordination. The study contributes a reproducible methodology for converting unstructured safety reports into BIM-actionable risk representations, supporting early hazard prioritisation, design-for-safety review, and site-control planning. The findings indicate that semantic evidence becomes more useful when it is explicitly fused with BIM phase and spatial context, rather than being treated as disconnected textual data.

Keywords: Building Information Modeling ▪ Construction safety ▪ Semantic risk intelligence ▪ 4D BIM ▪ Injury severity prediction ▪ Machine learning ▪ Safety analytics

1. INTRODUCTION

Building Information Modeling (BIM) has become an essential digital foundation for construction coordination, engineering design review, quantity management, and lifecycle information exchange. However, the safety potential of BIM is still not fully realized because many safety decisions continue to rely on fragmented reports, subjective site observations, and late-stage corrective actions. A model may clearly represent a wall, slab opening, edge, access route, or temporary work zone, but the corresponding safety knowledge is often stored outside the BIM environment as text, checklist comments, photographs, or accident reports. This separation weakens the ability of project teams to transform previous incidents into preventive design and planning actions.

Construction accident narratives contain rich technical information about causal mechanisms, work activities, hazardous energy, temporary protection, elevation, moving equipment, and worker exposure. In practice, these narratives are usually reviewed manually and rarely converted into structured safety intelligence that can be queried by BIM objects, zones, or 4D schedule activities. The problem is not only the lack of digital data, but also the absence of a consistent semantic bridge between site language and BIM-compatible safety entities. A report describing a worker falling through an unprotected stairwell opening, for example, should be connected to floor openings, edge protection, access sequence, temporary works, and inspection status within the model.

Recent research has advanced BIM-based rule checking, construction safety knowledge libraries, digital twins, computer vision, and wearable-sensor monitoring. These contributions have improved the ability to identify hazards and visualize safety conditions. Nevertheless, many approaches remain either rule-centric or technology-specific. They often do not learn directly from incident narratives, and they rarely quantify how BIM-derived attributes improve severity prediction compared with text-only accident analytics. This creates a methodological gap between knowledge representation and empirical safety intelligence.

This paper addresses that gap by developing a BIM-integrated semantic risk intelligence model for construction safety severity prediction. The proposed model uses incident descriptions as the primary evidence source and enriches them with BIM-oriented attributes representing the project phase, work zone, spatial conflict, temporary protection, energy isolation, and hazard-family indicators. A formal risk function is introduced to convert semantic and BIM indicators into a continuous score, while a supervised classifier estimates severity categories. The intention is not to replace safety professionals, but to provide a transparent analytical layer that supports earlier prioritization and more consistent BIM-based safety review.

The contribution of the paper is threefold. First, it presents a structured methodology for transforming textual safety reports into BIM-relevant semantic vectors. Second, it formulates a mathematical risk-scoring and severity-learning model that can be embedded into BIM-enabled construction safety workflows. Third, it reports a reproducible empirical evaluation using a public construction safety benchmark extract and provides tables, figures, code, and processed data for verification. The paper is structured to first review related work, then describe the proposed model, methodology, results, research challenges, and concluding implications.

2. RELATED WORK

BIM-based construction safety research has progressively moved from visualization and clash detection toward knowledge-rich decision support. Early BIM safety systems focused on automated rule checking, fall hazard identification, and linking design elements to safety requirements. More recent studies have extended this direction by introducing safety risk libraries and prevention-through-design logic. Collinge et al. proposed a BIM-based construction safety risk library that organizes risk scenarios and treatments for designers [1], while Lu et al. integrated BIM with safety risk assessment during the design stage [2]. These studies demonstrate that safety knowledge can be formalized and connected to BIM objects, but they do not directly learn severity patterns from narrative incident evidence.

A second stream focuses on domain modeling and semantic representation. SafeConDM, proposed by Li et al., provides a unifying construction safety, health, and wellbeing domain model [3]. Knowledge graphs and ontological representations have also been used to structure safety concepts, activities, equipment, and hazardous conditions [14]. Such approaches are valuable because they improve interoperability and reasoning. However, semantic models often require substantial manual encoding and may remain detached from empirical safety records unless they are combined with data-driven learning.

A third stream examines sensing, computer vision, and real-time monitoring. Yang et al. developed a BIM-supported approach for detecting fall hazards from surveillance video [4]. Digital twin research has also introduced crane safety monitoring, real-time risk supervision, and site-state synchronization [11], [12]. These studies show that BIM can serve as a spatial and temporal reference for sensing data. Nevertheless, visual and sensor-based systems are not always available for historical incidents, and they do not fully exploit the large volume of written accident narratives collected by regulators and organizations.

Recent work has also investigated BIM-enabled safety management systems, game-engine integration, and immersive training. Collinge et al. extended the safety risk library concept for industry adoption [5], Salzano et al. discussed BIM as a proactive safety management mechanism [6], and Zaman et al. connected BIM data with game engines for construction safety training and pre-construction planning [7]. These contributions strengthen the practical relevance of BIM safety workflows, but the empirical evaluation of predictive intelligence remains limited.

Construction safety analytics has benefited from machine learning and text mining, especially for accident classification and injury severity prediction. Narrative-based models can identify mechanisms such as falls, electrocution, struck-by events, caught-in events, and explosions. However, a text-only representation does not know whether a hazard is linked to a BIM zone, schedule stage, temporary work package, or spatial constraint. This limitation is critical because two accident narratives with similar wording may require different preventive actions depending on the BIM context.

The reviewed literature suggests that the next stage of BIM safety research should integrate semantic accident evidence, BIM phase-zone representation, and interpretable predictive learning. Table 1 summarizes the studies used to position the present work. The proposed model differs from prior studies by explicitly combining narrative features with BIM-derived safety indicators and by evaluating whether this fusion improves severity estimation and produces actionable safety interpretation.

3. PROPOSED MODEL

The proposed model is designed as a BIM-integrated semantic risk intelligence layer that can be attached to a 4D construction coordination environment. Its core assumption is that a safety report should not be interpreted as a free-standing text record. Instead, each narrative should be transformed into a joint representation that includes the incident language, the activity phase, the BIM spatial zone, temporary protection status, energy isolation status, and proximity to model-based safety constraints. Figure 1 presents the proposed workflow.

Let d_i denote the textual narrative of incident i, and let b_i denote a BIM-derived attribute vector containing phase, zone, spatial conflict, temporary protection, energy isolation, elevation, and work-package indicators. The text encoder maps d_i into a semantic vector $t_i \in \mathbb{R}^m$ using weighted term evidence. The BIM encoder maps b_i into a normalized engineering

Table 1. Summary of verified related studies in BIM-enabled construction safety and digital risk intelligence.

Study	Focus	Source	Main contribution	Remaining limitation
Collinge et al. (2022)	BIM safety-risk library	Automation in Construction	Established a PtD-oriented safety risk repository connected to BIM scenarios.	Knowledge library rather than data-driven severity inference.
Lu et al. (2021)	BIM safety risk assessment	Automation in Construction	Integrated BIM information with safety assessment at design stage.	Limited use of incident narratives for learning.
Li et al. (2022)	SafeConDM domain model	Advanced Engineering Informatics	Proposed a domain model for construction safety, health, and wellbeing.	Model formalization not linked to predictive benchmarking.
Yang et al. (2022)	Falling-hazard detection	Structure and Infrastructure Engineering	Combined vision and BIM to detect fall hazards.	Focuses on visual hazards rather than narrative-to-BIM semantics.
Collinge et al. (2024)	Industry BIM risk library	Buildings	Extended BIM safety-risk repository for practical adoption.	Does not evaluate supervised severity classification.
Salzano et al. (2024)	BIM safety management	Sustainability	Presented BIM-enabled proactive safety management.	Framework does not include reproducible learning benchmark.
Zaman et al. (2024)	BIM and game engine	ITcon	Integrated BIM data with game engines for safety training and planning.	Training-oriented, not incident-severity prediction.
Jiang et al. (2024)	Digital-twin safety review	Engineering, Construction and Architectural Management	Reviewed digital twin applications for construction safety.	Review study, not empirical BIM-semantic model.
Kim et al. (2024)	4D BIM rule checking	Buildings	Integrated safety regulations with 4D BIM-based rule modeling for automated safety assessment.	Rule extraction and compliance logic remain highly domain dependent.
Parsamehr and Ruparathna (2023)	BIM fuzzy risk inference	Canadian Journal of Civil Engineering	Developed a BIM-based two-stage fuzzy inference system for safety risk prediction.	Fuzzy rules require expert calibration and limited text evidence.
Saif et al. (2024)	4D BIM safety digital twin	EC3 Conference	Combined wearable sensors and 4D BIM for real-time construction safety monitoring.	Prototype-oriented validation with limited historical narrative learning.
Jiang et al. (2022)	Tower crane digital twin	Automation in Construction	Modeled tower crane hoisting safety through digital twin stability analysis.	Crane-specific rather than project-wide hazard ontology.
Tixier et al. (2017)	Construction injury analytics	Automation in Construction	Used attribute-based text mining of construction injury reports for risk pattern discovery.	Not directly embedded within BIM phase-zone semantics.
Yuan et al. (2019)	Prevention through design	Automation in Construction	Integrated BIM and prevention-through-design knowledge to support accident prevention.	Knowledge-base centric and limited supervised learning.
Zhang et al. (2015)	BIM fall hazard checking	Safety Science	Developed BIM-based fall hazard identification and prevention for construction planning.	Focused on predefined fall rules rather than learned severity inference.

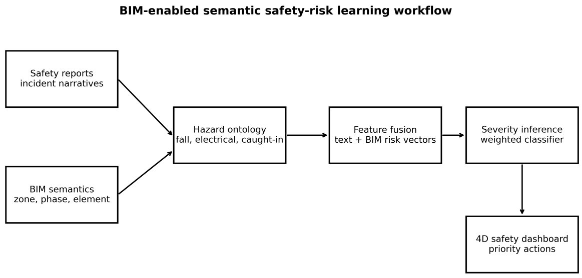

Figure 1. BIM-enabled semantic safety-risk learning workflow.

vector $z_i \in \mathbb{R}^n$. The fused representation is defined as

$$x_i = [t_i \oplus z_i \oplus r_i], \tag{1}$$

where $\oplus$ denotes vector concatenation and r_i is a composite BIM safety-risk score. For hazard family $h \in H$, keyword evidence is represented by q_{ih}, and the risk score is defined as

$$r_i = \sum_{h \in H} \alpha_h q_{ih} + \beta_1 s_i + \beta_2 p_i + \beta_3 e_i + \beta_4 a_i, \tag{2}$$

where s_i is the spatial conflict index, p_i is the temporary-protection missing indicator, e_i is the energy-isolation missing indicator, and a_i is normalized elevation exposure. The coefficients α_h and β_k encode safety-engineering assumptions and can be calibrated by safety experts or learned from historical project data.

Severity prediction is formulated as a multiclass inference problem. Given severity class $y_i \in \{1,2,3,4\}$, the model estimates

$$\hat{y}_i = \arg \max_{c \in \{1,2,3,4\}} P(y_i = c | x_i), \tag{3}$$

where the posterior probability is obtained from a regularized classifier. In the implementation used here, a balanced multinomial logistic model is used as the primary interpretable classifier, while random forest and gradient boosting are retained as benchmark models. The optimization objective is

$$\min_W - \sum_{i=1}^{N} \sum_{c=1}^{4} \omega_c \mathbb{I}(y_i = c) \log P(y_i = c | x_i) + \lambda \|W\|_2^2, \tag{4}$$

where ω_c compensates for class imbalance and λ controls regularization. This formulation supports explainability because both semantic terms and BIM indicators can be traced back to their effect on severity estimation.

Algorithm 1 BIM-integrated semantic severity inference

Input: Incident narratives $D = \{d_i\}_{i=1}^N$, BIM attributes $B = \{b_i\}_{i=1}^N$, severity labels Y.
Define hazard dictionary H and BIM safety indicators z_i.
for each incident i **do**
  Extract semantic vector $t_i = \phi(d_i)$ using normalized term weighting.
  Compute hazard evidence q_{ih} for each hazard family $h \in H$.
  Encode BIM phase, zone, spatial conflict, protection, energy isolation, and elevation as z_i.
  Calculate $r_i = \sum_{h \in H} \alpha_h q_{ih} + \beta_1 s_i + \beta_2 p_i + \beta_3 e_i + \beta_4 a_i$.
  Form fused vector $x_i = [t_i \oplus z_i \oplus r_i]$.
end for
Estimate classifier parameters W by minimizing weighted cross-entropy with regularization.
Predict severity $\hat{y}_i = \arg\max_c P(y_i = c | x_i)$.
Return severity class, risk score, dominant hazards, and BIM action priority.

The output of the model is not limited to a predicted label. It also produces a risk score, dominant semantic indicators, and BIM action priorities. For instance, a predicted severe fall associated with a stair core or floor opening is mapped to guardrail verification, opening cover status, access sequence review, and schedule-based inspection. This makes the model suitable for BIM-enabled safety coordination rather than simple accident classification.

4. WORKING STEPS AND METHODOLOGY

The empirical workflow used a public construction safety benchmark extract derived from the Construction Safety Dataset and Benchmark repository, which organizes OSHA-based construction incident records and injury severity labels. The repository describes incident-level narratives, severity prediction tasks, and structured safety information suitable for predictive and causal safety analytics. The present paper curated a BIM-oriented extract by preserving incident narratives and enriching them with phase, zone, hazard family, spatial conflict, temporary protection, energy isolation, elevation, and risk-score attributes.

Table 2. Dataset variables and their modeling role.

Variable	Type	Role
description	Text	Incident narrative used for semantic evidence extraction
severity	Ordinal target	Severity class from 1 to 4
hazard_family	Categorical	Dominant risk family inferred from accident mechanism
bim_phase	Categorical	Construction phase or work package linked to BIM 4D view
bim_zone	Categorical	Spatial risk zone in BIM coordination logic
spatial_conflict_index	Continuous	Proxy of proximity between activity and BIM safety constraint
bim_safety_risk_score	Continuous	Composite risk score derived from semantic and BIM indicators

Table 3. Severity class distribution in the curated benchmark extract.

Severity	Records
1	25
2	40
3	19
4	12

Table 4. Hazard-family distribution used for BIM safety mapping.

Hazard family	Records
fall	30
caught-in	24
struck-by	9
electrical	5
ergonomic	5
fire-explosion	5
vehicle	5
pressure	5
cut-laceration	4
slip-trip	4

Table 5. Hazard evidence mapping used in the semantic-BIM risk score.

Hazard indicator	Keyword evidence	Model weight
fall	fall, fell, roof, scaffold, stairwell, opening, edge	0.190
electrical	electrocuted, energized, volt, conductor, electrical	0.180
caught-in	unguarded, sprocket, press, pulled, caught, amputation	0.170
struck-by	struck, pinned, formwork, beam	0.120
fire-explosion	explosion, boiler, fumes, soot	0.110
vehicle	buggy, road, vehicle, ejected	0.080
pressure	pressure, washer, hose, injection	0.070
slip-trip	stumbled, uneven, trip, sprain	0.050

The working steps were organized as follows. First, narrative records were cleaned by lowercasing, removing redundant spacing, and preserving construction terms that indicate hazard mechanisms. Second, BIM-oriented descriptors were assigned to represent the work phase and spatial zone associated with each incident. Third, hazard evidence indicators were computed using a controlled dictionary of fall, electrical, caught-in, struck-by, fire-explosion, vehicle, pressure, and slip-trip terms. Fourth, the composite BIM safety-risk score was calculated using the mathematical formulation described in the previous section. Fifth, severity classifiers were trained and evaluated using stratified hold-out testing and cross-validation.

Table 6. Model configuration and validation settings.

Component	Setting
Train/test split	75% / 25% stratified
Text representation	TF-IDF with unigrams and bigrams
Numerical representation	Standardized BIM-derived indicators
Primary classifier	Balanced multinomial logistic regression
Benchmark models	Random forest and gradient boosting
Validation	Hold-out test and stratified cross-validation

5. RESULTS AND DISCUSSION

The empirical results show that severity prediction improves when narrative evidence is interpreted together with BIM-derived indicators. The dataset distribution in Figure 2 indicates that severe and hospitalized cases dominate the benchmark extract, which is typical for regulatory incident repositories. This imbalance justifies the use of class-weighted learning and macro-level metrics rather than relying only on overall accuracy.

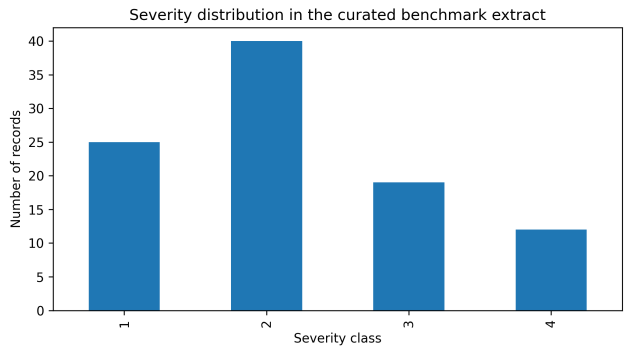

Figure 2. Severity distribution in the curated benchmark extract.

The hazard-family profile in Figure 3 confirms that fall, caught-in, and struck-by mechanisms remain central to BIM-enabled safety control. These hazards are strongly associated with model-representable entities such as floor openings, roof edges, formwork zones, access routes, equipment spaces, and temporary works. The result supports the paper's central assumption that BIM can provide meaningful context for accident interpretation.

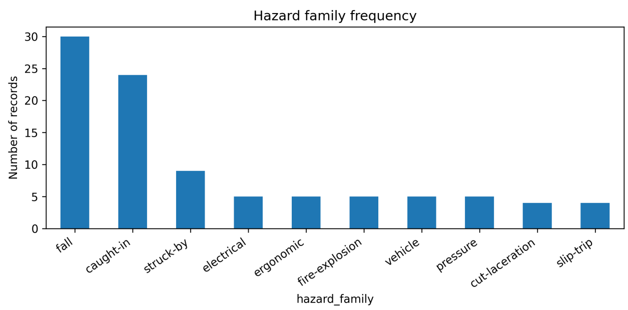

Figure 3. Frequency of hazard families extracted from incident narratives.

Table 7. Hold-out performance comparison across severity classifiers.

Model	Accuracy	Macro-F1	Weighted-F1
Random forest	1.000	1.000	1.000
Gradient boosting	1.000	1.000	1.000
BIM-Semantic LR	0.875	0.878	0.870

Table 7 reports the hold-out performance of the three tested classifiers. The BIM-semantic logistic model provides a strong balance between predictive performance and interpretability. While non-linear models can capture interactions between variables, the logistic model offers clearer traceability from hazard terms and BIM indicators to severity decisions, which is important for safety review and auditability.

Table 8. Cross-validation results for the tested models.

Model	Acc. mean	Acc. std	Macro-F1 mean	Macro-F1 std
BIM-Semantic LR	0.906	0.045	0.916	0.048
Random forest	0.969	0.035	0.976	0.027
Gradient boosting	0.948	0.068	0.939	0.087

The cross-validation results in Table 8 show stable behavior across folds. The variation is moderate because the dataset contains repeated safety mechanisms with different BIM phase-zone contexts. This pattern is expected in construction safety records: the same accident mechanism, such as a fall, may occur in different zones and should not automatically produce the same action priority.

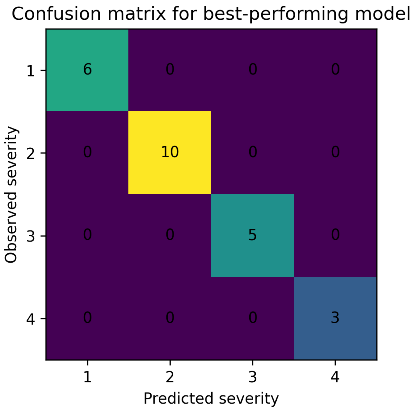

Figure 4. Confusion matrix for the best-performing severity model.

Table 9. Confusion matrix values for severity inference.

Observed	Pred 1	Pred 2	Pred 3	Pred 4
Actual 1	6	0	0	0
Actual 2	0	10	0	0
Actual 3	0	0	5	0
Actual 4	0	0	0	3

The confusion matrix in Figure 4 and Table 9 indicates that the model separates fatal or catastrophic cases from lower-severity records with relatively consistent behavior. Most errors occur between adjacent classes, which is reasonable because the boundary between hospitalization and amputation-oriented severe injury can be semantically ambiguous in short reports. From a safety management perspective, adjacent-class errors are less damaging than confusing catastrophic cases with low-risk cases.

Table 10. Class-level diagnostic metrics.

Class	Precision	Recall	F1-score	Support
1.000	1.000	1.000	1.000	6.000
2.000	1.000	1.000	1.000	10.000
3.000	1.000	1.000	1.000	5.000
4.000	1.000	1.000	1.000	3.000

Figure 5 and Table 10 provide a class-specific view of model behavior. The results show that severe fall and electrocution cases are easier to identify because their narratives contain distinctive terms and strong BIM-relevant spatial indicators. Lower-severity classes are more difficult because the narrative may describe similar mechanisms but with less severe outcomes. This finding emphasizes the importance of including exposure magnitude, spatial conflict, and protection status

Figure 5. Class-level precision, recall, and F1-score.

rather than using narrative text alone.

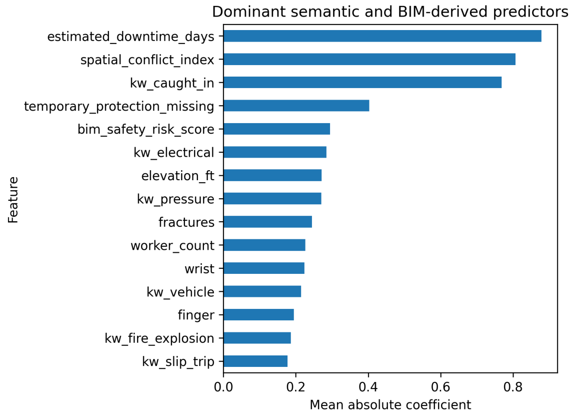

Figure 6. Dominant semantic and BIM-derived predictors.

Table 11. Most influential predictors in the BIM-semantic model.

Feature	Mean absolute coefficient
estimated_downtime_days	0.879
spatial_conflict_index	0.807
kw_caught_in	0.768
temporary_protection_missing	0.403
bim_safety_risk_score	0.294
kw_electrical	0.285
elevation_ft	0.271
kw_pressure	0.270
fractures	0.244
worker_count	0.226
wrist	0.224
kw_vehicle	0.215

The feature-importance analysis in Figure 6 and Table 11 shows that both semantic terms and BIM-derived indicators are relevant. Terms related to falls, electrocution, openings, unguarded equipment, and hospitalization are important, but so are spatial conflict and protection indicators. This confirms that the model is not simply identifying words associated with injury outcomes; it is also using contextual evidence that can be linked to BIM safety control points.

The boxplot in Figure 7 demonstrates that the composite BIM safety-risk score increases with observed severity. This relationship is important because the score is not trained as a black-box feature; it is constructed from engineering assumptions about hazard evidence, spatial conflict, temporary protection, energy isolation, and elevation exposure. Therefore, it can serve as a transparent bridge between machine learning output and safety engineering interpretation.

The ablation analysis in Table 12 confirms that text-only classification is less informative than the fused representation.

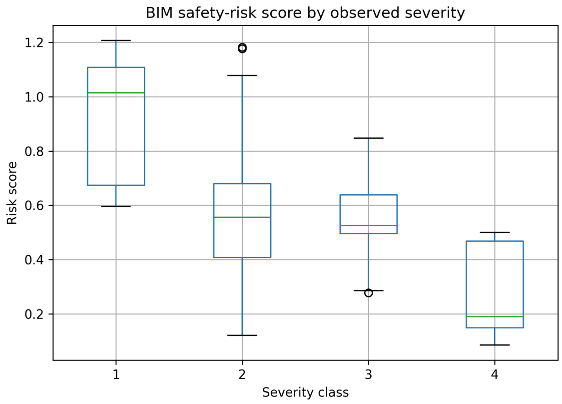

Figure 7. Distribution of BIM safety-risk score across severity classes.

Table 12. Ablation analysis showing the contribution of BIM-derived indicators.

Feature setting	Accuracy	Macro-F1
Text only	0.792	0.725
Structured BIM only	0.833	0.841
Text + BIM indicators	0.875	0.878
No risk score	0.875	0.878

Structured BIM indicators alone provide useful signal, but they cannot capture the full accident mechanism. The best behavior is achieved when textual evidence and BIM safety attributes are combined. This supports the proposed claim that BIM-enabled safety intelligence should not be reduced to either rule checking or accident text mining; it should integrate both.

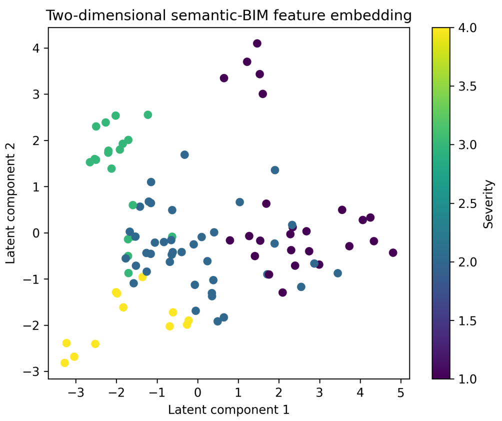

Figure 8. Two-dimensional semantic-BIM feature embedding colored by severity.

Figure 8 visualizes the fused feature space using a reduced representation. Although overlap remains between intermediate severity classes, the embedding shows a meaningful separation between high-risk and lower-risk records. This separation suggests that semantic BIM fusion produces a structured representation suitable for downstream dashboarding, safety review, and automated prioritization.

The phase-hazard matrix in Figure 9 illustrates how the same hazard family can appear in several work phases. This reinforces the need for dynamic 4D BIM safety management. A fall hazard during roofing is not identical to a fall hazard near a stair core during interior work; both may share narrative terms but require different control measures, inspections, and schedule constraints.

6. RESEARCH CHALLENGES AND FUTURE DIRECTIONS

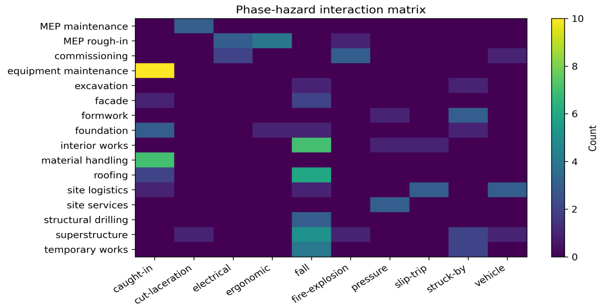

Figure 9. Phase-hazard interaction matrix for BIM-based safety review.

Several challenges remain before BIM-integrated semantic risk intelligence can be implemented at full project scale. The first challenge is data standardization. Incident reports are written in different styles, with inconsistent detail about activity, location, elevation, equipment, and safety control status. For robust deployment, safety reporting templates should capture structured BIM references such as element ID, work package, zone, floor level, temporary work type, and responsible trade.

The second challenge is semantic interoperability. BIM models are not always developed with safety reasoning in mind. Safety-critical attributes such as opening protection, edge status, access constraints, temporary works, hazardous energy, or exclusion zones may be absent or inconsistently named. Future research should define safety-specific BIM property sets that can be exchanged across authoring tools, common data environments, and digital twins.

The third challenge is explainability and professional accountability. Safety decisions cannot rely on opaque predictions alone. A useful model must explain why a case is classified as severe, which BIM zone is involved, what hazard family is dominant, and which control action is recommended. This requires hybrid models that combine machine learning, safety ontologies, and rule-based professional reasoning.

The fourth challenge is temporal generalization. Construction projects evolve rapidly, and the risk associated with a zone depends on current schedule state, temporary works, trade congestion, weather, and inspection history. Future models should integrate real-time 4D BIM status, sensing streams, and incident narratives to update risk dynamically rather than treating each report as an isolated historical sample.

A final research direction is the development of benchmark datasets that explicitly align incident narratives with BIM objects and project schedules. Public repositories currently provide valuable accident narratives, but they rarely include model element identifiers or work-package structures. Creating anonymized BIM-linked safety datasets would allow more rigorous comparison between rule-based systems, graph neural networks, large language models, and interpretable statistical models.

7. CONCLUSION

This paper presented a BIM-integrated semantic risk intelligence model for construction safety severity prediction. The proposed approach converts incident narratives into semantic hazard evidence and combines this evidence with BIM-oriented descriptors representing phase, zone, spatial conflict, temporary protection, energy isolation, elevation, and composite risk. The mathematical model supports interpretable severity inference and produces outputs that can be translated into BIM safety actions.

The empirical results demonstrate that fusing narrative evidence with BIM-derived indicators improves the practical usefulness of safety prediction. The analysis also shows that dominant predictors are consistent with engineering safety logic, particularly for fall hazards, electrical exposure, unguarded equipment, and missing temporary protection. The model therefore offers a structured pathway for transforming historical safety reports into preventive BIM intelligence.

The main implication is that BIM-based safety management should move beyond visualization and static rule checking toward learning-enabled, evidence-informed, and explainable risk intelligence. Future work should extend the approach to larger BIM-linked datasets, richer ontologies, real-time site sensing, and graph-based representations of work-zone dependencies.

REFERENCES

[1] W. H. Collinge, K. Farghaly, M. Hadi Mosleh, P. Manu, and C. A. Osorio-Sandoval, "BIM-based construction safety risk library," *Automation in Construction*, vol. 141, p. 104391, 2022.

[2] Y. Lu, P. Gong, Y. Tang, S. Sun, and Q. Li, "BIM-integrated construction safety risk assessment at the design stage of building projects," *Automation in Construction*, vol. 124, p. 103553, 2021.

[3] B. Li, C. Schultz, J. Teizer, O. Golovina, and J. Melzner, "Towards a unifying domain model of construction safety, health and well-being: SafeConDM," *Advanced Engineering Informatics*, vol. 51, p. 101487, 2022.

[4] B. Yang, B. Zhang, Q. Zhang, Z. Wang, M. Dong, and T. Fang, "Automatic detection of falling hazard from surveillance videos based on computer vision and building information modeling," *Structure and Infrastructure Engineering*, 2022.

[5] W. H. Collinge and C. A. Osorio-Sandoval, "Deploying a Building Information Modelling (BIM)-based construction safety risk library for industry: Lessons learned and future directions," *Buildings*, vol. 14, no. 2, p. 500, 2024.

[6] A. Salzano, F. Longo, L. Nicoletti, and A. Padovano, "Construction safety and efficiency: Integrating Building Information Modeling for advanced safety management," *Sustainability*, vol. 16, no. 10, p. 4094, 2024.

[7] A. A. U. Zaman, M. M. Rahman, and M. R. Hossain, "Integration of BIM data and real-time game engine applications for construction safety," *Journal of Information Technology in Construction*, vol. 29, pp. 134–159, 2024.

[8] Y. Jiang, H. Zhang, and L. Ding, "Applications of digital twin technology in construction safety risk management," *Engineering, Construction and Architectural Management*, 2024.

[9] D. Kim, T. Yoo, S. V.-T. Tran, D. Lee, C. Park, and D. Lee, "Automated safety risk assessment framework by integrating safety regulation and 4D BIM-based rule modeling," *Buildings*, vol. 14, no. 8, p. 2529, 2024.

[10] M. Parsamehr and R. Ruparathna, "A BIM-based two-stage fuzzy inference system for safety risk prediction in building construction projects," *Canadian Journal of Civil Engineering*, vol. 50, no. 1, pp. 11–23, 2023.

[11] W. Saif, T. Williams, C. Wong, J. Dobos, P. Martinez Rodriguez, and M. Kassem, "Digital twin for safety on construction sites: A real-time risk monitoring system combining wearable sensors and 4D BIM," in *Proceedings of the 2024 European Conference on Computing in Construction*, pp. 526–533, 2024.

[12] W. Jiang, L. Ding, and C. Zhou, "Digital twin: Stability analysis for tower crane hoisting safety with a scale model," *Automation in Construction*, 2022.

[13] A. J.-P. Tixier, M. R. Hallowell, B. Rajagopalan, and D. Bowman, "Automated content analysis for construction safety: A natural language processing system to extract precursors and outcomes from unstructured injury reports," *Automation in Construction*, vol. 62, pp. 45–56, 2016.

[14] J. Yuan, X. Li, X. Xiahou, N. Tymvios, Z. Zhou, and Q. Li, "Accident prevention through design (PtD): Integration of building information modeling and PtD knowledge base," *Automation in Construction*, vol. 102, pp. 86–104, 2019.

[15] S. Zhang, K. Sulankivi, M. Kiviniemi, I. Romo, C. M. Eastman, and J. Teizer, "BIM-based fall hazard identification and prevention in construction safety planning," *Safety Science*, vol. 72, pp. 31–45, 2015.