Factors Influencing the Effectiveness of Construction Safety Assessment Models: A Systematic Literature Review

Abdulrahman Mohammed I Shabil[1], Aiman Al-Odaini[2*]

[1]Faculty of Engineering, Lincoln University College, Malaysia, engabuhattan@gmail.com

[2]Faculty of Engineering, Lincoln University College, Malaysia, aiman@lincoln.edu.my

[*]Corresponding Author's E-mail: engabuhattan@gmail.com

Abstract: *The existing body of research on construction safety assessments is fragmented along the lines of technology, safety hazards, and assessment methods. This fragmentation prevents a systematic understanding of the conditions under which safety assessment models become valid, usable, and timely. The purpose of this review is to address this gap, by systematically identifying, assessing, and analyzing the factors that determine the assessment models' effectiveness in construction safety. This review was based on 52 studies published within the predefined review period, drawn from an initial pool of 326 studies, and conducted a thematic synthesis of the studies, which included a synthesis of various assessment criteria, a method of counting studies, and the construction of a binary matrix on the coverage of the studies. According to the studies reviewed, the factors that determine the assessment models' effectiveness can be grouped in the following domains: integrity of the evidence; contextual, and socio-organizational; and digital. The study results are based on the research of construction safety models and provide an auditable taxonomy of determinants, and an integrated interpretation of construction safety models, based on the technical, contextual, digital, and organizational needs, and can support design, validation, governance, and cross-context application of safety models and proactive construction safety governance.*

Keywords: construction safety; assessment models; occupational risk; machine learning; building information modelling; systematic review.

1. Introduction

The construction industry is one of the most perilous fields of work due to the fragmented, temporary, and decentered nature of construction work, as well as the subcontracting of work, the fluidity of crews, equipment, and work sequences, varying environmental conditions, and the interactions amongst various trades. These conditions make construction work difficult to assess for safety and inhibit the usefulness of lagging indicators. Lagging indicators, such as injury statistics, describe the result of harm and do not support injury prevention (X. Wu et al., 2015). Construction safety assessments are designed to account for these factors by rendering information pertaining to hazards, exposure, behaviors, intervention, management, and project conditions into risk levels, alert signs, expected results, or intervention recommendations. Management, indicator systems, multivariate and uncertainty models, and machine-learning (ML) approaches are present in the literature. So are the recent advances in assessment using construction site posture and movement, as well as images, sounds, physiological data, schedules, and project models (Awolusi & Marks, 2017). However, despite advancements in technology, construction safety assessment models designed with these technologies can fail to be effective when conditions for model assessments are not met. For example, a model safety assessment for construction may perform well, yet remain ineffective when data are lacking or justification for model assessments is insufficient, or when logic for model assessments is poorly defined, personnel using the safety assessment model are unqualified, or safety assessment model outputs are poorly aligned with organizational structures.

Model effectiveness results from particular technical, contextual, institutional, human, and organizational interplays. Such determinants include data quality, indicator validity, and predictive performance, among others. The influence of different factors can be grouped, and the boundaries can be respected to demonstrate how they act during different parts of the assessment cycle. However, most previous studies focused on a specific technology, a specific type of hazard, or a particular analytical technique. Studies on virtual and augmented reality focus on visualization and training; in the context of computer vision (CV), studies focus on the recognition of unsafe behavior; and bibliometric studies focus on artificial intelligence (AI), data fusion, and digital modeling. Real-time monitoring studies focus on the choices of sensing, communication, localization, and integration. While these studies are all meaningful, they do not greatly synthesize the factors that determine the modeling of valid, trusted, and useful models and the adoption, maintenance, and use of models for intervention purposes.

In the absence of such synthesis, the empirical foundations remain fragmented with inconsistent terminology. Although the feasibility was shown, field effectiveness was rarely demonstrated. Furthermore, the studies reviewed rarely distinguished between the frequency with which a determinant was examined and the quality with which it was examined. This review attempts to close this gap by identifying, standardizing, organizing in groups, and synthesizing the determinants of effectiveness of construction safety assessment models. The review maps the focus of research, assesses the confidence of the evidence, and creates a lifecycle vision that incorporates data acquisition, confirmation, contextual tailoring, and implementation, as well as organizational monitoring and learning. Thus, the review is meant to serve as a basis for model development, selection, use in practice, and governance.

2. Methodology

A systematic literature review was designed to encapsulate determinants that affect the efficacy of construction safety assessment models. This was done for model families, study designs, and project contexts. A qualitative configurative synthesis was undertaken for this review, and was complemented by a counts based and binary coverage matrix. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) reporting framework was followed for the stages of identification, screening, assessing eligibility and inclusion (Page et al., 2021). This review was said to be PRISMA-informed, rather than fully PRISMA compliant, as the documentation did not include database yields, the final search date or a distinct count of reports not retrieved. No formal protocol registration was documented or claimed. This review was limited to English-language documents published from January 1, two thousand fifteen, through December 31, two thousand twenty-five. The databases used were Scopus, WoS Core, ASCE Library, IEEE Xplore, Science Direct, SpringerLink, Wiley, Taylor and Francis, and MDPI. Google Scholar was used for citation chasing and finding published versions. The search strategy incorporated construction environments, safety, assessment or prediction models, and the concepts of effectiveness or validation. Other terms included building information modelling (BIM), ML, CV, Fuzzy Analytic Hierarchy Process (AHP), BN, wearable sensors, real-time monitoring, data fusion, and safety leading indicators. The search strategy was customized to each database; however, the exact strings were not preserved.

Eligible publications were those that described a construction safety assessment model, system of safety assessment indicators, a safety prediction method, a safety risk-evaluation framework, or an enabling technology. Those publications should have included sufficient details on their methodologies to allow for coding. Publications that were not construction-related, had safety references that were incidental, were not in English, were duplicated, were out of scope, were gray literature, and non-construction safety publications were excluded from selection. Initial search results yielded 326 publications. After the duplication removal, 254 publications underwent title and abstract selection. Following this, 167 were removed, and a further 35 were removed after reading the full text, leaving 52 publications to be included in the final analysis. This process is depicted in Figure 1. Number of reviewers, processes when reviewer disagreement occurred, inter-rater statistics, and exclusion count by reason were omitted. The studies were evaluated on the purpose, fit of design, clarity of inputs, sufficiency of data, validation, context, reproducibility, and interpretability. Validation and contextual description were required to meet a minimum score of 10. Data captured included year, context, model type, design, data, validation, and evidence of implementation. Limitations and suggested determinants were also noted. Hybrid deductive / inductive coding was applied to standardize the terminology. Model effectiveness was treated as the primary outcome. Each study was allowed to contribute only one count per determinant. Confidence was rated as moderate, limited to moderate, or limited. Due to the heterogeneity of outcome, design, and validation, a meta-analysis was not warranted.

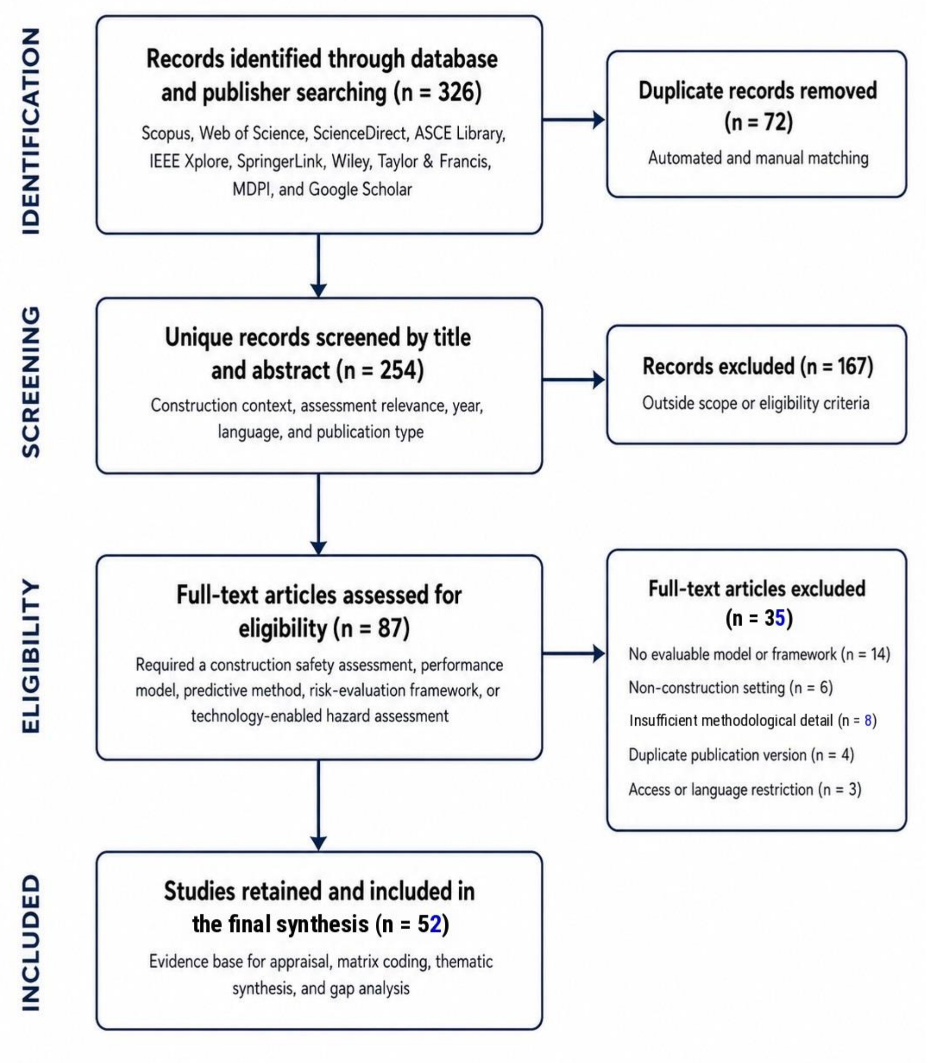

Figure 1: PRISMA-informed study-selection process

3. Construction Safety Assessment Landscape

The final corpus encompasses 52 studies cataloged within the predefined review period. The corpus constitutes a methodologically varied corpus. In terms of evidence, this corpus includes model development, case studies, runs of simulations, survey-based structural modeling, observational sensing, classifications based on ML, critical reviews, and bibliometric mapping. Most of the methods and models used in the studies include multi-criteria models, various approaches to uncertainty, AI, and predictive models, as well as management and indicator systems, sensing and real-time applications, and human and organizational models. Non-linear modeling approaches include structural equation modeling, clustering, neural networks, ensemble learning, graph convolution, fuzzy inference, AHP, cloud models, data fusion, and rule extraction. While this corpus captures several dimensions of safety assessment, it also prevents direct comparison of performance metrics. In predictive studies, the emphasis is on the classification of studies or error measures. In multi-criteria studies, the focus is on weighting and uncertainty, while in studies based on sensing, the emphasis is on detection and timeliness. Management models focus on the behavior and performance of an organization. The evidence is supported by the input-output efficiency as illustrated through the longitudinal analysis of regional safety performance (Qi et al., 2022).

Construction safety assessment models utilize information on hazards, exposure, behaviors, management systems, project features, and work conditions to formulate judgments to aid in prevention. Here, effectiveness is defined as the extent to which a method (1) produces information of sufficient validity and reliability, (2) performs adequately in the actual context of its use, (3) exhibits interpretability, and (4) is timely and makes possible the attribution of responsibility. This definition of effectiveness goes beyond the consideration of predictive accuracy. For example, a model that exhibits predictive accuracy may be ineffective from an operational perspective if it is built on a narrow dataset, generates opaque outputs, provides late warnings of responsibilities, and does so in a manner that is unclear. On the other hand, a model may be quite simple and provide a valid system of indicators if the data on which it draws are routinely collected, thresholds are set and adjusted in a collaborative manner, and there is evidence of active and engaged management. Effectiveness may be distributed across a lifecycle integrating evidence, model logic, context, users, organizational response, and ongoing monitoring and recalibration.

Validation methods are not uniform and are often not rigorous. For example, some models describe internal predictive performance and judgment of experts, and some describe sensitivity analysis, case verification, or practical demonstration. More rarely, validation of model outputs in other contexts, time periods, countries, or contracting systems is reported. The more recent ML models are straining at the boundaries of class imbalance, the quality of labels, and data leakage while multi-criteria models are more concerned with indicator choice, membership of assessment panels, consistency of weights, and assumptions. Finally, a suite of issues is affecting digital systems, such as the calibration of models, the occlusion, and the consistency of data, as well as issues surrounding connectivity, interoperability, privacy, and maintenance.

Table 1: Literature Review Matrix

No.	Study	DQ	IV	PA	CF	MT	UU	RTD	TI	SI	MS	WP	TQ	RF
1	X. Wu et al. (2015)		✓		✓					✓	✓			
2	Saurin et al. (2015)	✓	✓		✓					✓	✓	✓		
3	D. Fang et al. (2015)		✓		✓						✓	✓		
4	Seo et al. (2015)	✓	✓		✓						✓	✓	✓	
5	Teizer and Cheng (2015)	✓		✓	✓		✓	✓	✓					
6	Aryal et al. (2017)	✓		✓	✓			✓	✓			✓		
7	Awolusi and Marks (2017)	✓	✓		✓		✓			✓	✓	✓		
8	J. Chen et al. (2017)	✓		✓	✓			✓	✓					
9	Raviv et al. (2017)		✓		✓	✓				✓	✓			✓
10	Samantra et al. (2017)	✓	✓	✓	✓	✓				✓				
11	Carpio de los Pinos and González García (2017)		✓		✓	✓	✓							✓
12	Kasprowicz (2017)	✓	✓	✓	✓	✓								
13	Tong et al. (2018)		✓		✓	✓	✓			✓	✓	✓		✓
14	Ilbahar et al. (2018)	✓	✓	✓	✓	✓	✓							✓
15	W. Fang et al. (2018)	✓		✓	✓		✓	✓	✓			✓		

No.	Study	DQ	IV	PA	CF	MT	UU	RTD	TI	SI	MS	WP	TQ	RF
16	Poh et al. (2018)	✓	✓	✓	✓	✓		✓	✓		✓			
17	Li et al. (2018)				✓		✓	✓	✓				✓	
18	Yan et al. (2019)	✓	✓	✓	✓	✓				✓	✓			✓
19	H. Chen et al. (2019)	✓		✓	✓			✓	✓			✓		
20	Shen et al. (2020)	✓		✓	✓	✓			✓					
21	Sanni-Anibire et al. (2020)	✓	✓		✓		✓			✓	✓	✓		✓
22	Zhou et al. (2020)	✓	✓	✓	✓	✓		✓	✓	✓				
23	C. J. Lin et al. (2020)	✓	✓	✓	✓	✓			✓	✓				
24	Ayhan and Tokdemir (2020)	✓		✓	✓	✓			✓					
25	Y.-C. Lee et al. (2020)	✓		✓	✓		✓	✓	✓					
26	W. Fang et al. (2020)	✓		✓	✓		✓	✓	✓			✓		
27	Choe and Leite (2020)		✓		✓	✓	✓			✓	✓			✓
28	Trinh and Feng (2020)	✓	✓		✓						✓	✓	✓	
29	Lyu et al. (2020)	✓	✓	✓	✓	✓				✓				✓
30	K. Wu and Z. Wu (2020)	✓	✓	✓	✓	✓	✓			✓		✓		
31	Zhang and Mohandes (2020)	✓	✓	✓	✓	✓				✓				✓
32	Zhao and Obonyo (2020)	✓		✓	✓			✓	✓			✓		
33	Mohandes and Zhang (2021)	✓	✓	✓	✓	✓	✓			✓	✓			✓
34	S.-S. Lin et al. (2021)	✓	✓	✓	✓	✓		✓	✓	✓				
35	Zhu et al. (2021)	✓		✓	✓	✓			✓					
36	Lu et al. (2021)	✓	✓		✓	✓	✓		✓	✓	✓			✓
37	Chellappa et al. (2021)		✓		✓					✓	✓	✓	✓	✓
38	George et al. (2022)	✓		✓	✓	✓			✓					
39	Toğan et al. (2022)	✓		✓	✓	✓			✓					
40	Wang et al. (2022)	✓	✓	✓	✓	✓						✓	✓	
41	Mostofi et al. (2022)	✓		✓	✓	✓			✓					
42	Junjia et al. (2023)	✓	✓		✓	✓	✓		✓	✓				✓
43	Dadashi Haji et al. (2023)	✓	✓		✓		✓	✓	✓	✓	✓			
44	Pereira et al. (2024)	✓	✓	✓	✓		✓	✓	✓			✓	✓	
45	Pereira et al. (2025)	✓	✓	✓	✓		✓	✓	✓			✓	✓	
46	Alipour-Bashary et al. (2021)		✓		✓	✓	✓			✓				
47	Khalid et al. (2021)		✓		✓	✓					✓		✓	✓
48	Qi et al. (2022)	✓	✓		✓	✓								
49	Mohandes et al. (2023)		✓		✓	✓	✓			✓				
50	Mostofi and Toğan (2023)	✓	✓	✓	✓	✓	✓							
51	C. Lee et al. (2023)	✓	✓		✓	✓	✓			✓				
52	Kim et al. (2024)	✓	✓		✓	✓	✓		✓					✓

DQ = data quality; IV = indicator validity; PA = predictive accuracy; CF = context fit; MT = model transparency; UU = user usability; RTD = real-time data; TI = technology integration; SI = stakeholder input; MS = management support; WP = worker participation; TQ = training quality; RF = regulatory fit. A check mark indicates substantive treatment, not a statistically significant effect.

The structured matrix categorizes substantive treatment of thirteen determinants for all fifty-two studies. Context fit emerges in every article, as each study analyzes a construction setting. Thirty studies explicit focus on data quality, thirty-one on indicator validity, twenty-three on model transparency, and twenty-two on predictive accuracy. Stakeholder input and user usability appear in twenty-three studies. The remaining determinants have the following frequencies: management support and worker participation (17), real-time data and regulatory fit (15), and training quality (8). These counts represent the focus of scholarly interests, but do not indicate effect size or rank order of determinants. Analyzing the studies as a whole, there is emphasis on the technical feasibility and measurement and validation of data, and model logic, with a focus on workforce capability, and less on institutional alignment, sustained use and governance. The studies focus on buildings and rail and urban projects, as well as tunnels, excavations, equipment interfaces, sustainable construction, and worker-centred monitoring. The extraction record, however, does not give complete information on geographic focus or sample size. This will be stated honestly, and there won't be unsupported estimates.

4. Discussion: Thematic Synthesis of Factors

Of the 13 domains described in the corpus, Technical Evidence Integrity is the most developed. Data quality, identified in 41 studies, relates to the accuracy, completeness, consistency, and representativeness and labeling of model inputs. In construction modeling, noise and inconsistencies stemming from different contractors, and class and observation imbalances, affect many construction data sets. These modeling issues are particularly relevant and affect the integrity of any analysis, regardless of the sophistication of the modeling approach. Addressed in 37 studies, indicator validity issues arise when measures do not meaningfully represent the hazard, exposure, behavior, or management condition they are purported to describe. While collecting indicators may represent the only feasible option, the use of indicator measures in data modeling for prevention is unlikely to be valid (X. Wu et al., 2015). The concept of predictive accuracy appears in 30 studies. However, the terms accuracy, precision, recall, and calibration, and various error measures relate to different constructs, as do the indicators of predictive performance. Data governance, construct validity, verification, and thresholding are evidenced from the studies. However, external, temporal, and anticipatory validity are addressed to a much lesser extent. Technical performance, therefore, should be considered a necessary, but insufficient, condition for demonstrating safety and effectiveness.

The contextual and institutional relevance of a model is established when the model is able to retain its meaning in multiple environments. There is inherent variation in the complexity of construction projects, the profile of risk and hazards, the sequence of work, the procurement and contracting structure, the workforce, the maturity of technology, the climate, and the regulatory environment. All these variations influence the meaning of an indicator, the availability of data, the risk thresholds, the validation process, and the applicable control actions. Context validity was noted in all 52 studies. Every publication included in the studies refers to construction, but this widespread applicability should not be interpreted to mean a causal precedence. Trinh and Feng (2020) found that the complex nature of a construction project can have negative impacts on safety performance. However, a positive safety culture may help to strengthen this impact. Regulatory fit is noted infrequently and is mentioned in 15 studies. The various jurisdictions define legal responsibilities, reporting, acceptable risk, and liability; however, where these regulatory frameworks exist, they are often embedded implicitly within the evaluation criteria or the weights assigned by experts. The absence of these frameworks restricts the transfer and application. Future evaluations should record the assumptions underlying each legal framework, facilitate the flexibility of the framework, and clarify the addressees of the legal outputs to ensure that adequate preventive actions are taken.

The link between data outputs and human inputs is made possible by the interpretability and usability of decisions. Model transparency is captured by thirty-two studies and entails visibility of criteria and weights, influential variables, uncertainty, and limitations, as well as the justification of outputs. While some multi-criteria methods can reveal reasoning of experts, certain ML systems are incapable of explaining why a particular hazard classification or recommendation was made. Transparency, while important, is not the same as simplicity. In outputs of a system, there must be a sufficient amount of information that allows the user to assess the credibility of the output and the recommendation. Usability is covered by twenty-three studies and encompasses data-entry burden and understanding, as well as a model's fit with the user's workflow, and the prioritization and actionability of its outputs. A model may be technically sound, but if the interface of the model places a burden on the user, or the alerts are ambiguous or the outputs of the model do not assign responsibility, the model has no value. Outputs must be clear and ordered to be actualized and connected to possible mitigative actions.

With timely data and integrated systems, assessments move beyond retrospective reviews. The use of real-time data in the 15 studies referenced here, assists in the identification of hazards, as well as communicating and implementing responses in a more timely manner. Numerous modalities can provide the required data, including wearable sensors, CV, location, sound, and physiological data. However, the timely provision of data, in the presence of excessive false alerts and limited capacity to respond, is of little use. The fusion of position and posture, along with other data, improves the assessment of the situation (Awolusi & Marks, 2017). The concern for integration is illustrated in 25 studies, and is mainly the interoperability of sensors, BIM, databases and mobile tools, schedules, and safety-management systems. The BIM-based assessment helps to promote safety in construction by linking risks with design elements, activities, areas, and times (Kim et al., 2024; Lu et al., 2021). Integration, however, brings about interdependencies related to the completeness of models, data standards, interoperability, calibration, privacy, and cybersecurity and sustainment concerns. The evidence presented favors the digitization of processes where integration is advanced along the value chain of information and response, compared to previous systems that mainly focused on the addition of devices or increased analytical capabilities.

Socio-organisational enactment shapes the acceptance, resourcing, and conversion of findings into control action. Twenty-three studies integrate stakeholder input, including owners, designers, supervisors, specialists, workers, and regulators, in the selection, weighting, validation, and interpretation of indicators. Participation can strengthen the legitimacy, and relevance of indicators, but consensus is not a substitute for measurement validation. Management presence is noted in 17 studies, and enables participation, resource allocation, and prioritization and sustained work on control action. Without these conditions, alerts may be identified, but are likely to remain unresolved. Worker participation is noted in 17 studies, and reinforces the identification, feedback, and reporting on hazards, as well as the acceptance and correction of the system. Safety behavior is embedded in supervisory and organizational conditions and is not an isolated individual choice (Seo et al., 2015). The assessment of training draws the least, ranking lowest, in the studies, but users should have a clear understanding of definitions, observation frameworks, and protocols of uncertainty, as well as escalation procedures and the safety and protection control line. Of the protective control actions, the training of staff is highly supportive, but technology instruction alone is insufficient. Implementing protective control actions requires risk literacy and the ability and readiness to mitigate risks through different protective actions. The framework provides support for participation and management presence, but the longitudinal effects of co-design, acceptance, and training are yet to be explored.

Cautious interpretation of frequency patterns is warranted due to the nature of coverage, which is neither an effect size, importance ranking, or count of significant relationships. Several papers focus on development of models, simulations, applications, or reviews. They do not examine the relationship between a determinant and an outcome. The conditions that enable data quality and the recurrent elements of indicator validity, predictive/calculative accuracy, model transparency, fit contextual, and integrated technology, possess an evidential quality that is heterogeneous. Real-time data captured and used in accordance with regulatory frameworks and within management systems, as well as supportive participation and training, are considered conditional. Their implementation benefits primarily depend on the position, competence, and authority of the systems-and-service manager. The 13 determinants do not provide evidence for causation. Technical conditions are the most moderately confident, while indirect ceases of limited to moderately confident, or are also very limited. The diversity of constructs, samples, risks, validation methods, projects, and the purpose of models, helps to account for the variability. In the absence of evidential quality, the asserts distinguishes research focus from the quality of research, as opposed to the positive versus negative vote.

Synthesized findings lend credence to understanding effectiveness in relation to configurable and lifecycle-based models vis-a-vis dynamic project phases and varying organizational contexts. Quality data can yield little value if evaluative indicators are not defined, if outputs are not interpretable, if managers have no authority to act, and if regulatory frameworks do not define accountability. Transparent models do not make up for the lack of poor calibration, participatory engagement does not address unreliable evidence, and predictive precision will not bridge gaps in integration or sustain activities. A coherent system will bind together evidence collection, construct validation, model building, contextual calibration, clear communication, active implementation, control verification, and sustained monitoring and recalibration. The taxonomy offers a new lens wherein technical integrity, contextual and institutional fit, interpretability, usability, digital timeliness and integration, and socio-organizational enactment are enveloped in a single, auditable structure. For the developer this means starting with the decision , risk, the empowered user, and feasible controls, while tracing and documenting the data along with transparency concerning provenance, uncertainty, validation, framing, and ongoing efforts to mitigate risks and recalibrate. For the practitioner, model selection should consider project fit and the burden of work, training, privacy, and the model's maintainability and management, along with the model's capability to respond, rather than accuracy alone. Owners and regulators should mandate transparent frameworks that support reporting, participatory evaluation, integration, remote monitoring, and continual governance. The framework is a relational model of enablement that is evidence-based and does not purport to be a statistically validated model of causality.

5. Research Gaps and Future Research Directions

The first unresolved issue is concerning the need for greater conceptual clarity. Validity, reliability, accuracy, usability, transparency, and effectiveness are all used interchangeably in the literature and at times describe the same ideas. Future research should identify the outcome being focused on, and separate model performance from the performance and safety impact of its implementation. There is also a lack of integration of the theory. Quality of information, validity of measurement, contingency, socio-technical systems, safety climate, and technology acceptance, are cited frequently but are bound together to form a disparate understanding of how evaluation influences choice. Researchers should identify the lack of clarity in relation to workers, tasks, projects, firms, and regulations and not blame system failures on the actions of individuals. Developing clearer ideas and models will enhance the comparison and replication of studies, and aid in refinement of theory.

The second gap revolves around validation, implementation, and governance. Models should undergo prospective validation across a range of projects, contractors, phases, jurisdictions and periods. Studies should report calibration, discrimination, and performance. Missing data, class imbalances, and performance under dataset shifts should be accounted for. Comparative studies should evaluate full systems against one another and consider the data burden of the system, the speed of the decision, the comprehensibility of the decision, the false alarm burden, the quality of the interventions, and the verifiability of control and risk. Longitudinal and mixed method studies should be implemented to address the sustained use, adaptation, bypassing, or abandonment of systems following a pilot study. Future assessments should include a more robust coding of geography, project type, and firm size, as well as workforce and regulatory characteristics. Research on digital surveillance should address the privacy, consent, and security of the data and consider representativeness, fairness, data ownership, and an access and accountability system for alerts, as well as the means to scaffold workers' voice in challenging and amending the systems. In adopting these priorities, the model-development research would be aligned with the implementation of prevention research.

6. Conclusion

This systematic literature review focuses on studies published within the predefined review period and identifies 13 key determinants of construction safety assessment model effectiveness. The 13 determinants were grouped into 5 categories: the integrity of technical evidence, contextual and institutional fit, interpretability and decision usability, digital timeliness and integration, and socio-organizational enactment. The literature focuses on the fit and quality of context, data and indicators, as well as model transparency, predicted accuracy and the integration of technology, while the quality of training, regulatory fit, and the provision of real-time data seem to be less emphasized. These frequencies indicate the extent of research and not the causal significance. Across assessment models, precision is necessary but not sufficient. Efficacious models also require valid and reliable constructs, appropriate and sufficient control and assessment, implementable and understandable systems, integrated and competent users, and a responsive management and workforce who have been engaged and trained according to the regulatory frameworks.

The review creates an auditable determinant taxonomy and interprets a lifecycle that connects various stages of the evidence process, encompassing model construction, context specification, and the steps of communication, implementation, and subsequent monitoring and recalibration. These findings offer researchers a way to structurally devise comparative studies, and differentiate the technical aspects from the implementation and safety effects, focusing instead on external, prospective, cross-project, and cross-jurisdictional validations. For model developers, the synthesis is a clear call for articulated assumptions, documented evidence and uncertainty, user-oriented outputs, and designed recalibration. For practitioners, the findings suggest that a simpler model that incorporates legitimate indicators and defined roles and that organizationally follows-through may be better than a system that is technically advanced, opaque, and poorly maintained. For owners and regulators, the findings suggest a baseline for reporting purpose, data integrity, validation, and human oversight, as well as evidence of (de)classification of privacy and accountability. Heterogeneity among outcomes, methods, contexts, and validation metrics presents the principal limitation of this body of work. Future studies should aim to integrate technical advances with field-based implementation, governance, and organizational learning that focus on prevention within diverse construction contexts.

References

Alipour-Bashary, M., Ravanshadnia, M., Abbasianjahromi, H., & Asnaashari, E. (2021). A hybrid fuzzy risk assessment framework for determining building demolition safety index. KSCE Journal of Civil Engineering, 25(4), 1144–1162. https://doi.org/10.1007/s12205-021-0812-4

Aryal, A., Ghahramani, A., & Becerik-Gerber, B. (2017). Monitoring fatigue in construction workers using physiological measurements. Automation in Construction, 82, 154–165. https://doi.org/10.1016/j.autcon.2017.03.003

Awolusi, I. G., & Marks, E. D. (2017). Safety activity analysis framework to evaluate safety performance in construction. Journal of Construction Engineering and Management, 143(3), 05016022. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001265

Ayhan, B. U., & Tokdemir, O. B. (2020). Accident analysis for construction safety using latent class clustering and artificial neural networks. Journal of Construction Engineering and Management, 146(3), 04019114. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001762

Carpio de los Pinos, A. J., & González García, M. N. (2017). Critical analysis of risk assessment methods applied to construction works. Revista de la Construcción, 16(1), 104–114. https://doi.org/10.7764/RDLC.16.1.104

Chellappa, V., Srivastava, V., & Salve, U. R. (2021). A systematic review of construction workers' health and safety research in India. Journal of Engineering, Design and Technology, 19(6), 1488–1504. https://doi.org/10.1108/JEDT-08-2020-0345

Chen, H., Luo, X., Zheng, Z., & Ke, J. (2019). A proactive workers' safety risk evaluation framework based on position and posture data fusion. Automation in Construction, 98, 275–288. https://doi.org/10.1016/j.autcon.2018.11.026

Chen, J., Qiu, J., & Ahn, C. (2017). Construction worker's awkward posture recognition through supervised motion tensor decomposition. Automation in Construction, 77, 67–81. https://doi.org/10.1016/j.autcon.2017.01.020

Choe, S., & Leite, F. (2020). Transforming inherent safety risk in the construction industry: A safety risk generation and control model. Safety Science, 124, 104594. https://doi.org/10.1016/j.ssci.2019.104594

Dadashi Haji, M., Behnam, B., Sebt, M. H., Ardeshir, A., & Katooziani, A. (2023). BIM-based safety leading indicators measurement tool for construction sites. International Journal of Civil Engineering, 21, 265–282. https://doi.org/10.1007/s40999-022-00754-9

Fang, D., Wu, C., & Wu, H. (2015). Impact of the supervisor on worker safety behavior in construction projects. Journal of Management in Engineering, 31(6), 04015001. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000355

Fang, W., Ding, L., Love, P. E. D., Luo, H., Li, H., Peña-Mora, F., Zhong, B., & Zhou, C. (2020). Computer vision applications in construction safety assurance. Automation in Construction, 110, 103013. https://doi.org/10.1016/j.autcon.2019.103013

Fang, W., Ding, L., Luo, H., & Love, P. E. D. (2018). Falls from heights: A computer vision-based approach for safety harness detection. Automation in Construction, 91, 53–61. https://doi.org/10.1016/j.autcon.2018.02.018

George, M. R., Nalluri, M. R., & Anand, K. B. (2022). Application of ensemble machine learning for construction safety risk assessment. Journal of the Institution of Engineers (India): Series A, 103, 989–1003. https://doi.org/10.1007/s40030-022-00690-w

Ilbahar, E., Karaşan, A., Cebi, S., & Kahraman, C. (2018). A novel approach to risk assessment for occupational health and safety using Pythagorean fuzzy AHP and fuzzy inference system. Safety Science, 103, 124–136. https://doi.org/10.1016/j.ssci.2017.10.025

Junjia, Y., Alias, A. H., Haron, N. A., & Abu Bakar, N. (2023). A bibliometric review on safety risk assessment of construction based on CiteSpace software and WoS database. Sustainability, 15(15), 11803. https://doi.org/10.3390/su151511803

Kasprowicz, T. (2017). Quantitative assessment of construction risk. Archives of Civil Engineering, 63(2), 55–66. https://doi.org/10.1515/ace-2017-0016

Khalid, U., Sagoo, A., & Benachir, M. (2021). Safety management system (SMS) framework development—Mitigating the critical safety factors affecting health and safety performance in construction projects. Safety Science, 143, 105402. https://doi.org/10.1016/j.ssci.2021.105402

Kim, D., Yoo, T., Tran, S. V.-T., Lee, D., Park, C., & Lee, D. (2024). Automated safety risk assessment framework by integrating safety regulation and 4D BIM-based rule modeling. Buildings, 14(8), 2529. https://doi.org/10.3390/buildings14082529

Lee, C., Kim, M., Ahn, H., Yoo, W. S., Cho, H., & Choi, Y. (2023). Developing a field-oriented analytical safety assessment model for construction workers. KSCE Journal of Civil Engineering, 27(10), 4085–4100. https://doi.org/10.1007/s12205-023-1978-8

Lee, Y.-C., Shariatfar, M., Rashidi, A., & Lee, H. W. (2020). Evidence-driven sound detection for prenotification and identification of construction safety hazards and accidents. Automation in Construction, 113, 103127. https://doi.org/10.1016/j.autcon.2020.103127

Li, X., Yi, W., Chi, H.-L., Wang, X., & Chan, A. P. C. (2018). A critical review of virtual and augmented reality applications in construction safety. Automation in Construction, 86, 150–162. https://doi.org/10.1016/j.autcon.2017.11.003

Lin, C. J., Zhang, M., Li, L. P., Zhou, Z. Q., Liu, S., Liu, C., & Li, T. (2020). Risk assessment of tunnel construction based on improved cloud model. Journal of Performance of Constructed Facilities, 34(3), 04020028. https://doi.org/10.1061/(ASCE)CF.1943-5509.0001421

Lin, S.-S., Shen, S.-L., Zhou, A., & Xu, Y.-S. (2021). Risk assessment and management of excavation system based on fuzzy set theory and machine learning methods. Automation in Construction, 122, 103490. https://doi.org/10.1016/j.autcon.2020.103490

Lu, Y., Gong, P., Tang, Y., Sun, S., & Li, Q. (2021). BIM-integrated construction safety risk assessment at the design stage of building projects. Automation in Construction, 124, 103553. https://doi.org/10.1016/j.autcon.2021.103553

Lyu, H.-M., Sun, W.-J., Shen, S.-L., & Zhou, A.-N. (2020). Risk assessment using a new consulting process in fuzzy AHP. Journal of Construction Engineering and Management, 146(3), 04019112. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001757

Mohandes, S. R., Durdyev, S., Sadeghi, H., Mahdiyar, A., Hosseini, M. R., Banihashemi, S., & Martek, I. (2023). Towards enhancement in reliability and safety of construction projects: Developing a hybrid multi-dimensional fuzzy-based approach. Engineering, Construction and Architectural Management, 30(6), 2255–2279. https://doi.org/10.1108/ECAM-09-2021-0817

Mohandes, S. R., & Zhang, X. (2021). Developing a holistic occupational health and safety risk assessment model: An application to a case of sustainable construction project. Journal of Cleaner Production, 291, 125934. https://doi.org/10.1016/j.jclepro.2021.125934

Mostofi, F., & Toğan, V. (2023). A data-driven recommendation system for construction safety risk assessment. Journal of Construction Engineering and Management, 149(12), 04023139. https://doi.org/10.1061/JCEMD4.COENG-13437

Mostofi, F., Toğan, V., Ayözen, Y. E., & Tokdemir, O. B. (2022). Construction safety risk model with construction accident network: A graph convolutional network approach. Sustainability, 14(23), 15906. https://doi.org/10.3390/su142315906

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

Pereira, F., González García, M. d. l. N., & Poças Martins, J. (2024). An evaluation of the technologies used for the real-time monitoring of the risk of falling from height in construction—Systematic review. Buildings, 14(9), 2879. https://doi.org/10.3390/buildings14092879

Pereira, F., González García, M. d. l. N., & Poças Martins, J. (2025). BIM for safety: Applying real-time monitoring technologies to prevent falls from height in construction. Applied Sciences, 15(4), 2218. https://doi.org/10.3390/app15042218

Poh, C. Q. X., Ubeynarayana, C. U., & Goh, Y. M. (2018). Safety leading indicators for construction sites: A machine learning approach. Automation in Construction, 93, 375–386. https://doi.org/10.1016/j.autcon.2018.03.022

Qi, H., Zhou, Z., Li, N., & Zhang, C. (2022). Construction safety performance evaluation based on data envelopment analysis (DEA) from a hybrid perspective of cross-sectional and longitudinal. Safety Science, 146, 105532. https://doi.org/10.1016/j.ssci.2021.105532

Raviv, G., Shapira, A., & Fishbain, B. (2017). AHP-based analysis of the risk potential of safety incidents: Case study of cranes in the construction industry. Safety Science, 91, 298–309. https://doi.org/10.1016/j.ssci.2016.08.027

Samantra, C., Datta, S., & Mahapatra, S. S. (2017). Fuzzy based risk assessment module for metropolitan construction project: An empirical study. Engineering Applications of Artificial Intelligence, 65, 449–464. https://doi.org/10.1016/j.engappai.2017.04.019

Sanni-Anibire, M. O., Mahmoud, A. S., Hassanain, M. A., & Salami, B. A. (2020). A risk assessment approach for enhancing construction safety performance. Safety Science, 121, 15–29. https://doi.org/10.1016/j.ssci.2019.08.044

Saurin, T. A., Formoso, C. T., Reck, R., Beck da Silva Etges, B. M., & Ribeiro, J. L. D. (2015). Findings from analysis of incident-reporting systems of construction companies. Journal of Construction Engineering and Management, 141(9), 05015007. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000988

Seo, H.-C., Lee, Y.-S., Kim, J.-J., & Jee, N.-Y. (2015). Analyzing safety behaviors of temporary construction workers using structural equation modeling. Safety Science, 77, 160–168. https://doi.org/10.1016/j.ssci.2015.03.010

Shen, T., Nagai, Y., & Gao, C. (2020). Design of building construction safety prediction model based on optimized BP neural network algorithm. Soft Computing, 24(11), 7839–7850. https://doi.org/10.1007/s00500-019-03917-4

Teizer, J., & Cheng, T. (2015). Proximity hazard indicator for workers-on-foot near miss interactions with construction equipment and geo-referenced hazard areas. Automation in Construction, 60, 58–73. https://doi.org/10.1016/j.autcon.2015.09.003

Toğan, V., Mostofi, F., Ayözen, Y. E., & Tokdemir, O. B. (2022). Customized AutoML: An automated machine learning system for predicting severity of construction accidents. Buildings, 12(11), 1933. https://doi.org/10.3390/buildings12111933

Tong, R., Wu, C., Li, Y., & Fang, D. (2018). An assessment model of owner safety management and its application to real estate projects. KSCE Journal of Civil Engineering, 22(5), 1557–1571. https://doi.org/10.1007/s12205-017-1740-1

Trinh, M. T., & Feng, Y. (2020). Impact of project complexity on construction safety performance: Moderating role of resilient safety culture. Journal of Construction Engineering and Management, 146(2), 04019103. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001758

Wang, H.-H., Chen, J.-H., Arifai, A. M., & Gheisari, M. (2022). Exploring empirical rules for construction accident prevention based on unsafe behaviors. Sustainability, 14(7), 4058. https://doi.org/10.3390/su14074058

Wu, K., & Wu, Z. (2020). A human-centered risk model for the construction safety. IEEE Access, 8, 154072–154086. https://doi.org/10.1109/ACCESS.2020.3017772

Wu, X., Liu, Q., Zhang, L., Skibniewski, M. J., & Wang, Y. (2015). Prospective safety performance evaluation on construction sites. Accident Analysis & Prevention, 78, 58–72. https://doi.org/10.1016/j.aap.2015.02.003

Yan, H., Gao, C., Elzarka, H., Mostafa, K., & Tang, W. (2019). Risk assessment for construction of urban rail transit projects. Safety Science, 118, 583–594. https://doi.org/10.1016/j.ssci.2019.05.042

Zhang, X., & Mohandes, S. R. (2020). Occupational health and safety in green building construction projects: A holistic Z-numbers-based risk management framework. Journal of Cleaner Production, 275, 122788. https://doi.org/10.1016/j.jclepro.2020.122788

Zhao, J., & Obonyo, E. (2020). Convolutional long short-term memory model for recognizing construction workers' postures from wearable inertial measurement units. Advanced Engineering Informatics, 46, 101177. https://doi.org/10.1016/j.aei.2020.101177

Zhou, H., Zhao, Y., Shen, Q., Yang, L., & Cai, H. (2020). Risk assessment and management via multi-source information fusion for undersea tunnel construction. Automation in Construction, 111, 103050. https://doi.org/10.1016/j.autcon.2019.103050

Zhu, R., Hu, X., Hou, J., & Li, X. (2021). Application of machine learning techniques for predicting the consequences of construction accidents in China. Process Safety and Environmental Protection, 145, 293–302. https://doi.org/10.1016/j.psep.2020.08.006