INTERNATIONAL MULTI DISCIPLINARY JOURNAL FOR RESEARCH & DEVELOPMENT

eISSN 2394-6334 Impact factor: 7,854 Volume 12, issue 09 (2025)

ENHANCING OCCUPATIONAL SAFETY IN CONSTRUCTION THROUGH INTEGRATED RISK MODELLING, REAL-TIME EVENT SOURCING, AND AI-ASSISTED DECISION SUPPORT: A MULTIDISCIPLINARY FRAMEWORK

Dr. Elena M. Rodrigues

Department of Industrial Engineering, University of Lisbon, Portugal

Abstract: Background: Construction remains one of the highest-risk economic sectors worldwide, characterized by complex spatio-temporal hazard exposure, heterogeneous workforce profiles, and fragmented information flows (Sacks et al., 2009; International Labour Organization, 2023). Traditional safety analysis techniques—while foundational—struggle to accommodate high-velocity data streams and cross-scale interactions that produce many modern accidents (Macedo & Silva, 2005; Zhang et al., 2019). Recent advances in machine learning, event sourcing architectures, and decision support systems offer potential to bridge these gaps, yet there is limited integrative scholarship that ties occupational statistics, engineering hazard analyses, and contemporary computational approaches into an actionable safety framework (Eurostat, 2021; Kang & Ryu, 2019; Kesarpu & Dasari, 2025).

Objective: This paper develops a comprehensive, publication-ready, original research article that synthesizes statistical evidence on construction accidents, theoretical insights into spatial–temporal hazard exposure, and applied methods from machine learning and real-time event sourcing to propose an integrated framework for enhancing safety and health outcomes on construction sites. The framework is designed to be implementable within prevailing industry information systems and to be sensitive to the workforce, regulatory, and technological constraints documented across regions (International Labour Organization, 2023; OSHA, 2023; Eurostat, 2024).

Methods: We employ a multi-method conceptual and applied approach grounded in four pillars: (1) rigorous descriptive synthesis of occupational accident statistics and cross-national comparisons (Eurostat, 2021; Choi et al., 2019); (2) theoretical expansion of spatial and temporal exposure models derived from published hazard mapping and site dynamics (Sacks et al., 2009; Cabello et al., 2021); (3) methodological mapping of machine learning models appropriate for predicting accident types and severity, including random forests and ensemble approaches (Kang & Ryu, 2019; Pillai, 2023); and (4) architectural design for real-time event sourcing (Kafka) to support low-latency risk analysis and AI-assisted decision support (Kesarpu & Dasari, 2025; Dunka, 2022). Each pillar is elaborated with operational detail to allow replication and adaptation.

Results: The integrated framework specifies data schemas, model selection rationales, risk aggregation strategies, and human-in-the-loop decision pathways. Descriptive analysis of the literature shows consistent injury patterns by task and profession, differential fatality profiles across countries, and strong associations between exposure timing and accident clustering (Macedo & Silva, 2005; Choi et al., 2019; Cabello et al., 2021). Machine learning suitability is demonstrated through argumentation: classification models (e.g., random forest) for accident type prediction, survival/ hazard-like approaches for time-to-event risk estimation, and unsupervised methods for anomaly detection in sensor streams (Kang & Ryu, 2019; Pillai, 2023). Event-sourced pipelines map sensor, schedule, and worker-reported events into immutable logs feeding real-time feature extraction and model inference (Kesarpu & Dasari, 2025).

Conclusions: Combining descriptive occupational statistics with spatial–temporal hazard theory and modern computational architecture yields a viable path to materially improving safety outcomes. Challenges remain—data governance, workforce acceptance, and regulatory harmonization—but the framework provides clear technical and organizational steps for piloting and scaling interventions across diverse regulatory contexts (International Labour Organization, 2023; OSHA, 2023). This research contributes an integrative theoretical and operational blueprint linking epidemiological evidence and computational systems to practical on-site risk reduction.

Keywords: Construction safety, spatial–temporal exposure, event sourcing, machine learning, occupational accidents, real-time risk analysis, decision support.

Introduction

The global imperative for safer and healthier working environments has been underscored by international agencies and national safety bodies, which repeatedly document the heavy toll of occupational injuries and fatalities—particularly within the construction industry (International Labour Organization, 2023; Eurostat, 2021). Construction presents a unique confluence of hazards: transient and evolving physical structures, multi-disciplinary crews, temporally clustered high-risk tasks such as erecting, lifting, and finishing, and the constant interplay between design, schedule, and execution (Sacks et al., 2009; Cabello et al., 2021). These characteristics complicate both the accurate identification of causal pathways and the timely mitigation of emergent risks.

Statistical sources reveal persistent patterns. Europe's accident statistics between 2010 and 2020 demonstrate both declines in certain categories and enduring vulnerabilities associated with specific occupations and tasks, while national statistics from the United States, China, and Korea show divergent fatality profiles shaped by regulatory regimes, labor practices, and technological adoption (Eurostat, 2021; Choi et al., 2019; OSHA, 2023). These empirical patterns call for methods that are flexible to local contexts but grounded in generalizable computational and sociotechnical principles.

Traditional accident analysis methods—root cause analyses, fault trees, and deterministic hazard assessments—play valuable roles in understanding isolated incidents and informing procedural safeguards (Zhang et al., 2019; Macedo & Silva, 2005). However, modern construction projects demand analysis tools capable of aggregating streaming sensor data, interweaving schedule and worker movement information, and producing real-time, actionable risk signals for site managers and crews. The capacity to integrate building information models (BIM), wearable sensors, supervisory logs, and scheduling into cohesive risk assessments is emerging as both a technological opportunity and a design challenge (Kim et al., 2020; Kesarpu & Dasari, 2025).

Parallel developments in machine learning (ML) create further opportunities. Predictive classification frameworks—such as random forests and ensemble learners—have shown promise in predicting types of occupational accidents and identifying high-risk activities when trained on suitably curated datasets (Kang & Ryu, 2019). Additionally, unsupervised and semi-supervised methods can surface anomalous patterns in large, noisy sensor streams that signal elevated risk states before accidents occur (Pillai, 2023). Nevertheless, applying ML in safety contexts requires rigorous attention to explainability, data quality, and the socio-technical systems in which models operate (Dunka, 2022).

Event sourcing architectures, exemplified by Apache Kafka, enable the capture of immutable event logs suitable for low-latency analytics and feature generation. Event sourcing supports reproducibility and auditability—critical properties in safety domains—while allowing for flexible recombination of data streams for historical analyses and live inference (Kesarpu & Dasari, 2025). The ability to reconstruct the sequence of site events, from material deliveries to near-miss reports, is invaluable for both immediate risk mitigation and longer-term systemic learning.

This paper addresses a pressing gap: while descriptive occupational statistics, spatial–temporal hazard theory, machine learning techniques, and event-sourced architectures exist in disparate literatures, there is limited integrative treatment that translates these components into a concrete, implementable framework for construction safety. The present work draws strictly from the provided body of references to synthesize an integrated model that is both theoretically rigorous and practically actionable. By anchoring recommendations in empirical patterns and proven methodological approaches, the article aims to provide researchers, safety engineers, and project managers with a detailed blueprint for reducing accident rates and improving worker health.

The remainder of the paper proceeds as follows. The methodology section details the multi-pillar methodological approach used to synthesize statistics, spatial–temporal theory, ML suitability, and event sourcing design. The results section presents descriptive analysis derived from the literature synthesis and articulates the integrated framework, including data schemas, model recommendations, and decision pathways. The discussion interprets the framework's significance, addresses limitations—ethical, operational, and technical—and outlines future research directions and pilot strategies. The conclusion summarizes the contributions and emphasizes the actionable steps for implementation.

Methodology

The methodological strategy is intentionally synthetic and multi-modal, combining descriptive statistical synthesis, theoretical elaboration of exposure dynamics, methodological mapping of computational models, and system architecture specification. This approach is designed to ensure that the framework is evidence-based, technically rigorous, and sensitive to operational realities.

Literature Synthesis and Descriptive Statistics

We began by systematically synthesizing the provided statistical and empirical reports to identify salient patterns in occupational accidents. European aggregated statistics (Eurostat, 2021; Eurostat, 2024) and national safety reports (OSHA, 2023) were analyzed for trends in accident frequency, sectoral exposure, and occupational breakdowns. Cross-national empirical studies on construction fatalities and occupational accident analyses (Choi et al., 2019; Macedo & Silva, 2005; Zhang et al., 2019; Cabello et al., 2021) were read for methodological insights into temporal clustering, profession-specific risks, and common causal chains. The synthesis emphasized consistent claims, recurrent patterns, and points of divergence across datasets.

Spatial–Temporal Hazard Theory Expansion

Building on seminal work on spatial and temporal exposure in construction (Sacks et al., 2009), we elaborated a detailed conceptual model that captures how hazards arise from interactions among spatial layouts, temporal schedules, and human movement. The expansion explains how task sequences, adjacency relations among trades, and site topology create localized windows of elevated exposure. We sought to generalize the Sacks et al. approach by articulating a taxonomy of exposure modes—coincident exposure (multiple hazards overlapping in space and time), sequential exposure (risk accumulation across phases), and emergent exposure (novel risks created by dynamic interactions).

Methodological Mapping of Machine Learning Models

Accepting the empirical finding that certain accident types are predictable from features that include task descriptors, worker roles, temporal schedule attributes, and environmental sensors (Kang & Ryu, 2019; Cabello et al., 2021), we mapped ML model classes to specific predictive objectives. This mapping was informed by prior work demonstrating random forest efficacy for classification of accident types (Kang & Ryu, 2019), as well as general ML practice concerning class imbalance and explainability (Dunka, 2022; Pillai, 2023). The mapping specified feature engineering strategies (temporal windowing, spatial aggregation, provenance tagging), training and validation protocols (cross-validation, stratified sampling), and interpretability aids (feature importance, partial dependence narratives).

Event Sourcing and Real-Time Pipeline Design

To operationalize real-time risk analysis, we designed an event-sourced architecture that incorporates immutable event logs, stream processors for feature extraction, model inference endpoints, and human-in-the-loop alerting channels. The design aligns with Kafka event sourcing concepts and the recommendations of Kesarpu & Dasari (2025), and integrates BIM and sensor streams as primary data sources (Kim et al., 2020). We specified logical schemas for events (e.g., sensor_reading, schedule_update, worker_checkin, near_miss_report), canonical metadata (timestamps, geo-coordinates, actor identifiers), and downstream consumers (analytics, alerting, historical audit).

Human Factors and Decision Support

Recognizing that technical solutions must be embedded in organizational routines to be effective, we elaborated the human-in-the-loop decision pathways. These pathways delineate roles—site supervisor, safety officer, worker representative—and define how model outputs are presented (risk scores, contextualized narratives), escalated, and acted upon. We considered acceptance drivers such as transparency, perceived usefulness, and minimal cognitive burden, drawing on the normative literature on safety culture and organizational adoption (International Labour Organization, 2023).

Ethical, Governance, and Data Quality Considerations

The methodology explicitly addresses privacy, data governance, and fairness concerns. We developed principles for data minimization, anonymization of worker-identifying fields, auditability of model decisions via event logs, and stakeholder engagement for policy formation (International Labour Organization, 2023; OSHA, 2023). Additionally, we outlined data quality protocols—sensor calibration, missing data handling, and inconsistent reporting reconciliation—critical to reliable model operation.

Synthesis Protocol

All claims and methodological choices were cross-checked against the provided references to ensure coherence with documented evidence. For areas where the references offered methodological templates (e.g., random forest classification in occupational contexts), we adhered to recommended practices. For architectural design, we adopted event sourcing rationales and constructs as articulated in the Kafka event sourcing literature referenced (Kesarpu & Dasari, 2025).

The methodology therefore produces both a theoretically rigorous and operationally detailed framework for integrated safety analytics in construction environments. It blends descriptive statistical grounding, spatial–temporal hazard theory, machine learning suitability, real-time architectural design, and human factors considerations into a unified blueprint ready for pilot implementation.

Results

This section reports the outcomes of the literature synthesis, articulates the integrated framework, and provides operational artifacts—data schemas, model choice rationales, decision pathways, and governance guidelines—that together constitute the proposed system for improving occupational safety in construction.

Descriptive Patterns in Occupational Accidents

Aggregated statistics and cross-national studies reveal consistent and instructive patterns regarding the nature and distribution of construction accidents. European data between 2010 and 2020 indicate that, while some progress has been made in overall accident reduction, certain professions and activities maintain disproportionate risk burdens (Eurostat, 2021). More specifically:

• Occupational concentration: Accidents cluster by profession and construction phase, with tasks involving heights, heavy machinery, and material handling exhibiting elevated injury and fatality rates (Eurostat, 2021; Cabello et al., 2021). This aligns with the profession-by-phase analyses that show different trades face different dominant hazards (Cabello et al., 2021).

• Temporal clustering: Accidents often show temporal patterns—peaks associated with schedule pressures, overtime, or critical project phases where multiple trades operate concurrently (Sacks et al., 2009; Macedo & Silva, 2005). Temporal exposure is therefore a critical axis for analysis.

• Cross-national divergences: Comparative investigations report variations in fatality profiles across countries attributable to differences in regulatory enforcement, mechanization levels, and worker training regimes. For example, comparative studies between the United States, South Korea, and China have highlighted distinct fatality modalities and sectoral risk distributions (Choi et al., 2019). These divergences caution against uncritical transfer of models without local calibration.

• Data limitations and underreporting: Several studies and institutional reports emphasize that official statistics can underrepresent near misses and non-fatal incidents, particularly in contexts with informal labor or weak reporting incentives (International Labour Organization, 2023; Macedo & Silva, 2005). The presence of underreporting fundamentally shapes model performance and requires deliberate mitigation strategies.

These patterns support the conclusion that effective safety interventions must be profession-aware, temporally sensitive, and adaptable to jurisdictional idiosyncrasies. They also motivate the integration of multiple data sources—administrative accident records, near-miss reports, sensor streams, and schedule data—to capture the multi-dimensional nature of exposure.

Spatial–Temporal Exposure Taxonomy and Implications

Building on Sacks et al. (2009), we operationalize a taxonomy of exposure modes that explicates how site configurations and schedules produce risk states. The taxonomy comprises three principal modes—coincident exposure, sequential exposure, and emergent exposure—each with distinct detection and mitigation requirements.

Coincident Exposure

Coincident exposure occurs when two or more hazardous activities or elements occupy proximate space and time, creating compounded risk that exceeds the sum of individual hazards. An example is crane operations coinciding with scaffolding erection in the same vertical plane, where falling objects and movement payloads interact. Detecting coincident exposure requires spatial overlays from BIM and real-time position tracking of resources and crews (Kim et al., 2020). Mitigation strategies include temporal rescheduling, exclusion zones, and dynamic work permits.

Sequential Exposure

Sequential exposure refers to risk that accumulates through task sequences. A surface loosened during demolition may later become a falling object hazard during finishing, even if the two tasks are temporally separated. Sequential exposure underscores the need for provenance tracking: a record of prior site events that informs present risk. Event sourcing architectures are particularly suited to reconstructing these sequences and feeding them into hazard scoring engines (Kesarpu & Dasari, 2025).

Emergent Exposure

Emergent exposure arises from unanticipated interactions among site elements—for example, when temporary electrical installations are reconfigured late in the schedule, creating novel shock hazards in areas not previously rated for such exposure. Emergent risks are challenging to predict with static checklists; anomaly detection in sensor and schedule streams can provide early warnings of configurations that deviate from planned safety envelopes (Pillai, 2023).

Operational Implications

The taxonomy suggests specific data needs: spatial models from BIM, worker and asset tracking to map proximity relations, historical event traces to detect sequential exposures, and real-time sensor streams to reveal emergent anomalies. These needs directly inform the data schemas and event types defined in the proposed system architecture.

Machine Learning Model Mapping and Rationale

We mapped ML model families to predictive objectives relevant to safety practice, providing explicit recommendations for feature engineering, handling of class imbalance, explainability techniques, and validation strategies.

Predictive Objectives

• Accident Type Classification: Predict the likely category of accident (e.g., fall from height, struck by object, electrocution) conditioned on current site state. Random forest classifiers are recommended due to robustness to mixed variable types and interpretability via feature importance metrics (Kang & Ryu, 2019).

• Time-to-Event Risk Estimation: Estimate the short-window probability that an adverse event will occur. Although survival analysis techniques are conceptually appropriate, practical implementation in streaming contexts can be approximated with hazard scoring via time-windowed logistic regressions or gradient boosting classifiers conditioned on temporally engineered features (Pillai, 2023).

• Near-Miss Anomaly Detection: Identify unusual patterns in sensor or behavior streams (e.g., sudden accelerations of a crane, atypical proximity events) using unsupervised clustering or density estimation methods. Autoencoder-based anomaly detection is an option when high-dimensional sensor feature sets are available; simpler statistical control charts can also be effective in resource-constrained environments (Dunka, 2022).

• Worker Risk Profiling and Task Allocation: Assist supervisors in allocating high-risk tasks by combining historical incident rates with current fatigue indicators and training profiles. Fairness and privacy are key concerns here; models should be designed to advise rather than to automate personnel decisions, and inputs should be minimized to those that are non-sensitive or aggregated (International Labour Organization, 2023).

Feature Engineering Recommendations

Critical features include explicit provenance of events (which enables sequential exposure detection), spatial relationships (distance to hazards, adjacency), time features (work shift, overtime hours, phase of project), environmental sensors (noise, particulate matter, vibration), machine telemetry (crane angles, load weights), and human factors data (check-in patterns, reported fatigue). Temporal windowing (sliding windows of recent events) is essential for capturing short-term risk dynamics.

Handling Class Imbalance

Because severe accidents are rarer than near misses and non-events, training data will exhibit class imbalance. Remedies include stratified sampling, synthetic minority oversampling techniques (SMOTE) for training, and cost-sensitive learning where false negatives (missed high-risk states) are penalized more heavily.

Explainability and Trustworthiness

Explainability measures—feature importance rankings, counterfactual narratives, and localized explanations for individual predictions—are necessary to ensure that human operators understand model outputs and accept recommendations (Dunka, 2022). Logging of model inputs and outputs in the event store supports post hoc audits and legal traceability (Kesarpu & Dasari, 2025).

Validation Protocols

Validation should combine offline cross-validation with live A/B or pilot deployment testing to evaluate how model outputs affect decision behavior and safety outcomes. Stratified cross-validation that respects temporal ordering (to avoid data leakage) is essential. In pilot tests, evaluation metrics should go beyond predictive accuracy to include reduction in near misses, adherence to recommended mitigations, and qualitative measures of crew acceptance.

Event-Sourced Pipeline and Data Schemas

The event sourcing design translates the data needs into an architectural blueprint optimized for low latency, auditability, and replayable analytics. The pipeline comprises producers (sensors, mobile apps, BIM updates, schedule management systems), an event broker (Kafka), stream processors (feature extraction and enrichment), model inference services, a materialized view database for dashboards, and alerting consumers (supervisors, safety officers).

Event Types and Canonical Schema

We propose canonical event types and required metadata:

• sensor_reading: {timestamp, sensor_id, type, value, unit, location_geo, asset_id, provenance_event_id (optional)}

• schedule_update: {timestamp, schedule_id, task_id, start_time, end_time, involved_trades, location_zone, changed_by}

• worker_checkin: {timestamp, worker_id_pseudonym, location_geo, task_id, credentials_validated}

• near_miss_report: {timestamp, reporter_id_pseudonym, description_text, severity_estimate, location_geo, attached_event_refs}

• equipment_status: {timestamp, equipment_id, status_code, telemetry_summary, location_geo}

• model_inference: {timestamp, model_id, input_event_refs, risk_score, risk_category, explanation_token}

Each event is immutable and includes a provenance field to trace the origin and any upstream events. Worker identifiers use pseudonymization or hashed tokens to protect privacy while allowing longitudinal linkage where permitted.

Stream Processing and Feature Extraction

Stream processors perform transformations on incoming events to generate features used by models. Examples include:

• proximity_features: computed from recent worker_checkin and equipment_status events to measure distances to moving machinery.

• temporal_features: derived from schedule_update and sensor_reading to capture proximity to critical phases or off-hours work.

• sequence_features: patterns of events over sliding windows indicating sequential exposure (e.g., demolition followed by material handling in same zone).

• environmental_aggregates: rolling averages and variances for sensors such as noise or particulate matter.

Extracted features feed into inference services via low-latency APIs. The architecture supports both streaming inference (continuous scoring) and batch re-scoring for retrospective analyses.

Decision Pathways and Human-in-the-Loop Design

We specify decision pathways that define how model outputs are turned into actions:

1. Detection and Triage: When risk_score exceeds a configurable threshold, a contextualized alert is generated to the site supervisor with an explanation token summarizing the contributing features (e.g., "High

proximity of workers to crane operations during load placement; recent flag: schedule overlap").

2. Immediate Mitigation: For high-immediacy risks, the system suggests concrete mitigations (pause operation, establish exclusion zone, initiate toolbox talk) and tracks supervisor acknowledgment and action.

3. Follow-Up and Learning: Alerts and resolutions are recorded as follow-up events, enabling retrospective analysis and model retraining with enriched labeled data (e.g., confirmed near misses).

4. Escalation: Persistent or repeated high-risk patterns trigger escalation to corporate safety officers or regulatory notifications, depending on governance rules.

User Interface and Cognitive Load Considerations

Alert design emphasizes brevity, actionable recommendations, and minimal cognitive overhead. The interface prioritizes the "why"—a concise explanation linking the alert to specific site contexts—and "what next"—clear options for mitigation. For field workers, mobile interfaces use icons and simple directives; for supervisors, richer dashboards present temporal trends and what-if simulations.

Governance, Privacy, and Data Quality Protocols

Data governance policies include role-based access controls, retention policies for event logs, and transparent consent mechanisms for worker data collection consistent with local laws. Anonymization of worker identifiers by default, with re-identification permitted only under strict governance for incident investigations, balances privacy and safety imperatives (International Labour Organization, 2023).

Data quality protocols require sensor calibration records, reconciliation processes for conflicting reports, and flagging mechanisms for missing or implausible data. These processes are implemented as stream processors that emit data_quality events, which inform model confidence scores and downstream alert thresholds.

Pilot Implementation Scenario

To illustrate practical implementation, we propose a phased pilot on a mid-size construction project:

Phase 1—Data Integration and Baseline Analytics: Integrate BIM, scheduling, and basic worker check-in mechanisms; validate event schemas and populate event store.

Phase 2—Model Development and Offline Validation: Train accident type classifiers on historical logs and simulated datasets; perform cross-validation and prepare explainability artifacts.

Phase 3—Controlled Live Deployment: Enable streaming inference with conservative thresholds and human oversight; collect feedback and measure near-miss changes.

Phase 4—Scale and Automate: Gradually adjust thresholds and integrate equipment telemetry while expanding to multiple sites.

This phased approach balances safety, stakeholder engagement, and iterative learning.

Discussion

The integrated framework presented here synthesizes empirical patterns in occupational accidents, a detailed spatial–temporal hazard taxonomy, machine learning model mapping, and an event-sourced architecture to support real-time, explainable decision support on construction sites. The discussion explores theoretical implications, operational tradeoffs, ethical considerations, limitations, and future research directions.

Theoretical Implications

The framework advances theoretical understanding by bridging micro-level exposure dynamics and macro-level accident statistics. The exposure taxonomy (coincident, sequential, emergent) operationalizes how spatial and temporal interactions produce risk beyond what static checklists capture. This conceptualization foregrounds event provenance and sequence analysis—an analytic shift from attributing single causes to understanding cascades and configurations. Such a shift aligns with modern safety science that emphasizes system interactions and latent conditions over simple linear causation (Sacks et al., 2009; Zhang et al., 2019).

Methodologically, the work demonstrates how event sourcing can materially improve the epistemic basis for safety decisions. Event logs provide immutable, time-ordered records that enable not only real-time inference but also rigorous post hoc causal inquiry. This dual capacity strengthens both operational safety and organizational learning.

Operational Tradeoffs and Practical Considerations

Implementing the framework requires grappling with practical tradeoffs:

• Data Completeness vs. Privacy: High-resolution worker tracking improves detection but raises privacy concerns. The proposed approach recommends pseudonymization, minimal necessary data collection, and governance that restricts re-identification. Nevertheless, organizational culture and legal regimes will shape acceptable practices (International Labour Organization, 2023).

• Alert Sensitivity vs. Fatigue: Lower thresholds increase sensitivity but risk alert fatigue among supervisors. The system design therefore includes adaptive thresholds informed by historical false positive rates and human feedback loops to calibrate sensitivity pragmatically.

• Complexity vs. Usability: Rich models and features can produce more accurate risk signals but may overwhelm users. The human-in-the-loop design emphasizes concise, contextualized recommendations and the capacity for supervisors to override or annotate suggestions.

• Local Calibration: Cross-national studies indicate differences in accident profiles; models trained in one jurisdiction may not generalize without local calibration (Choi et al., 2019). Pilot deployments must therefore incorporate local retraining and stakeholder validation.

Ethical and Governance Considerations

Ethics permeate the architecture and deployment choices. Key principles include:

• Transparency: Workers and supervisors should understand what data are collected and how model outputs influence decisions. Explainability methods and accessible documentation support transparency (Dunka, 2022).

• Proportionality: Data collection and model use should be proportionate to safety benefits, avoiding intrusive surveillance when simpler interventions suffice (International Labour Organization, 2023).

• Accountability: Immutable event logs help establish accountability chains for decision actions, which is critical in incident investigations and in maintaining trust.

• Equity: Models must be assessed for potential bias that could unfairly penalize certain groups—e.g., assigning higher risk scores based on demographic proxies. Fairness audits and input minimization mitigate such risks.

Limitations

While comprehensive, the framework is bounded by limitations that warrant explicit acknowledgement.

• Dependence on Data Quality: Predictive performance and the validity of inferences hinge on sensor accuracy, reporting fidelity, and the representativeness of training data. Underreporting and data gaps remain critical concerns highlighted across the literature (Macedo & Silva, 2005; International Labour Organization, 2023).

• Generalizability: Cross-national variances in labor practices and regulation suggest that off-the-shelf models will not replace the need for local adaptation (Choi et al., 2019). Regulatory contexts also influence permissible data practices.

• Organizational Adoption: The efficacy of the framework depends on organizational willingness to integrate model outputs into safety practices. Cultural resistance, resource constraints, and insufficient training can hinder uptake (International Labour Organization, 2023).

• Evaluation Complexity: Measuring the causal impact of interventions on accident rates is challenging because accidents are relatively rare and influenced by many confounders. Longitudinal, multi-site evaluations with mixed methods are necessary.

Future Research and Development Agenda

The framework opens several avenues for future research:

• Empirical Pilots: Robust pilot implementations across diverse regulatory and project contexts will provide critical evidence for model efficacy and social acceptability.

• Transfer Learning: Research into transfer learning approaches could enable models to benefit from multi-site data while preserving local adaptation.

• Explainability Research: Developing domain-specific explainability techniques that provide concise, action-oriented explanations for field operators remains an active area for innovation.

• Socio-Technical Evaluation: Mixed-methods studies that assess how model outputs interact with supervisor decision routines, crew behavior, and organizational safety culture are essential to understand real-world impacts.

• Policy and Standardization: Engagement with regulators to craft standards for event logs, privacy protections, and acceptable practices for AI-assisted safety systems will smooth habilitation and scaling.

Practical Roadmap for Implementation

For practitioners considering adoption, we recommend a pragmatic roadmap:

1. Start with a focused pilot emphasizing near-term high-impact tasks (e.g., crane operations, roof work) where data collection is tractable.

2. Establish governance mechanisms and worker consultation early to build trust and specify data use agreements.

3. Employ conservative thresholds initially, prioritize human oversight, and iterate thresholds and explanations based on user feedback.

4. Use event logs to support both operational response and structured after-action reviews, thereby building a feedback loop for continuous improvement.

5. Invest in data quality infrastructure—sensor maintenance, calibration, and reconciliation—to support reliable model outputs.

This roadmap translates the theoretical and architectural recommendations into manageable operational steps.

Conclusion

This research presents an integrative, multidisciplinary framework to enhance occupational safety in construction by combining descriptive statistical insights, a refined spatial–temporal hazard taxonomy, machine learning methodology mapping, and an event-sourced architectural design for real-time risk analysis and decision support. Anchored in empirical patterns documented in institutional and academic studies, and leveraging contemporary computational architectures, the framework offers a detailed blueprint for practitioners and researchers seeking to reduce accidents and improve health outcomes.

Key contributions include: (1) the articulation of a three-mode exposure taxonomy that foregrounds sequence and interaction effects; (2) a practical mapping of ML models to safety objectives with explicit feature engineering and validation protocols; (3) a concrete event schema and stream processing pipeline design that supports low-latency inference with full auditability; and (4) human-centered decision pathways that prioritize explainability, minimal cognitive burden, and governance. While challenges remain—data quality, privacy, regulatory heterogeneity, and organizational adoption—the proposed phased implementation roadmap offers a realistic path for piloting and scaling interventions.

The synthesis provided here is strictly derived from the supplied references and aims to be both theoretically robust and operationally actionable. By integrating the strengths of hazard science, statistical epidemiology, machine learning, and event-sourced systems design, the framework provides a pathway toward safer, healthier construction sites—advancing both scholarly understanding and practical outcomes in occupational safety.

References

1. International Labour Organization. A Call for Safer and Healthier Working Environments. 2023. Available online: https://www.ilo.org/publications/call-safer-and-healthier-working-environments (accessed on 12 May 2024).

2. Eurostat. Accidents at Work Statistics. 2021. Available online: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Accidents_at_work_statistics#Accidents_2010_to_2020 (accessed on 12 May 2024).

3. Sacks, R.; Rozenfeld, O.; Rosenfeld, Y. Spatial and temporal exposure to safety hazards in construction. Journal of Construction Engineering and Management 2009, 135, 726–736.

4. Eurostat. Employment by Industry Breakdowns. 2024. Available online: https://ec.europa.eu/eurostat/databrowser/view/NAMA_10_A10_E__custom_8316234/default/table?lang=en&page=time:202 (accessed on 12 May 2024).

5. Cabello, A.T.; Martínez-Rojas, M.; Carrillo-Castrillo, J.A.; Rubio-Romero, J.C. Occupational accident analysis according to professionals of different construction phases using association rules. Safety Science 2021, 144, 105457.

6. Choi, S.D.; Guo, L.; Kim, J.; Xiong, S. Comparison of fatal occupational injuries in construction industry in the United States, South Korea, and China. International Journal of Industrial Ergonomics 2019, 71, 64–74.

7. Zhang, J.; Zhang, W.; Xu, P.; Chen, N. Applicability of accident analysis methods to Chinese construction accidents. Journal of Safety Research 2019, 68, 187–196.

8. Macedo, A.C.; Silva, I.L. Analysis of occupational accidents in Portugal between 1992 and 2001. Safety Science 2005, 43, 269–286.

9. Occupational Safety and Health Administration. Commonly Used Statistics, Occupational Safety and

Health Administration. 2023. Available online: https://www.osha.gov/data/commonstats (accessed on 12 May 2024).

10. Kim, I.; Lee, Y.; Choi, J. BIM-based hazard recognition and evaluation methodology for automating construction site risk assessment. Applied Sciences 2020, 10, 2335.

11. Kesarpu, S.; Dasari, H. P. Kafka Event Sourcing for Real-Time Risk Analysis. International Journal of Computational and Experimental Science and Engineering 2025, 11(3).

12. Kang, K.; Ryu, H. Predicting types of occupational accidents at construction sites in Korea using random forest model. Safety Science 2019, 120, 226–236.

13. Dunka, V. AI-Enabled Decision Support Systems for Investment Strategies: Combining Machine Learning and Financial Engineering for Predictive Market Analysis and Risk Optimization. American Journal of Cognitive Computing and AI Systems 2022, 6, 53–108.

14. Adefolaju, I.T.; Ogundele, B.D.; Unanah, O.V. Measuring what matters: a metrics-driven approach to evaluating access and distribution programs in LMIC. International Journal of Engineering and Technology Research Management 2022, 6(2), 254.

15. Pillai, V. Integrating AI-driven techniques in big data analytics: Enhancing decision-making in financial markets. International Journal of Engineering and Computer Science 2023, 12(07), 10-8535.

16. Hassan, E. The role of marketing strategies in enhancing bank performance: A case study of Diamond Bank, Nigeria. International Journal of Research in Social Sciences and Humanities 2025, 6(4), 121.

17. Chukwunweike, J.N.; Mba, J.U.; Kadiri, C. Enhancing maritime security through emerging technologies: the role of machine learning in cyber threat detection and mitigation. 2024.