



(REVIEW ARTICLE)



INTEGRATING MACHINE LEARNING WITH OCCUPATIONAL INCIDENT ANALYTICS: EMERGING FRAMEWORKS FOR NATIONWIDE WORKPLACE SAFETY ENHANCEMENT

Ayomide Solomon Alao *

Georgia State University.

* Corresponding Author

Magna Scientia Advanced Research and Review, 2026, 17(02), 446–457

Publication history: Received on 08 July 2026; revised on 17 August 2026; accepted on 19 August 2026

Article DOI: <https://doi.org/10.30574/msarr.2026.17.2.0171>

Copyright © 2026 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0

ABSTRACT

The integration of predictive analytics into United States occupational safety and health practice promises to shift workplace risk management from post-incident recordkeeping toward prospective injury prevention. Synthesizing U.S.-focused research published between 2021 and 2026, this narrative review critically evaluates machine learning applications across severe incident classification, narrative text processing, real-time computer vision, return-to-work outcome forecasting, and emerging federal oversight models. While algorithmic capabilities have reached high computational performance, including production-scale transformer deployments for administrative coding and high-accuracy ensembles for accident narrative parsing, the literature remains dominated by retrospective offline experiments. Crucially, empirical evaluations demonstrate a profound gap between model precision and tangible worker safety, as virtually no published studies measure prospective reductions in workplace injury or illness rates. This lack of demonstrated field impact is further complicated by severe systemic data fragmentation across federal enforcement registries, statistical surveys, sector-specific databases, and state-bounded workers' compensation claims. To bridge this divide, this paper articulates an integrated nationwide framework featuring federated data interoperability, risk-calibrated algorithmic hygiene standards, mandatory prospective evaluation protocols, and a phased evolution toward binding administrative regulation. Aligning computational innovation with measurable workplace hazard reduction, rather than further optimizing classification accuracy on historical datasets, represents the essential mandate for the future of occupational safety analytics.

Keywords: Machine Learning, Occupational Health and Safety, Predictive Analytics, Injury Surveillance, Algorithmic Hygiene, Federated Learning

1 INTRODUCTION

Occupational safety and health (OSH) practice in the United States has historically depended on retrospective and compliance-oriented data collection, such as injury logs, post-incident investigations, and periodic survey estimates that describe what has already gone wrong rather than predicting what is about to happen (Jetha et al., 2025). This lagging-indicator paradigm has produced a rich statistical record alongside a very limited capacity for prevention. In 2024 alone, U.S. employers reported 2.5 million nonfatal workplace injuries and illnesses in private industry, representing a rate of 2.3 cases per 100 full-time equivalent workers. During that same period, 5,070 fatal work injuries occurred, yielding a

fatality rate of 3.3 per 100,000 full-time equivalent workers (U.S. Bureau of Labor Statistics, 2026b, 2026c). These figures have remained largely stable for over a decade despite sustained regulatory attention, suggesting that incremental improvements to reactive systems have reached a plateau.

Machine learning (ML) has been proposed as a mechanism to break this plateau by enabling occupational incident analytics to shift from description to prediction. Rather than merely tabulating injuries after they occur, ML models can be trained on structured administrative fields and unstructured narrative text to estimate the likelihood, severity, and causal pathway of future incidents (Khairuddin et al., 2022). Over the past five years, this application has matured considerably. The U.S. Bureau of Labor Statistics itself now operates a production machine-learning system. Evolving from regularized logistic regression models introduced in 2014 to deep neural networks and currently a distilled transformer architecture, this system automatically assigns detailed injury and illness codes to hundreds of thousands of annual survey narratives, replacing a task that previously consumed approximately 25,000 hours of manual labor per year (U.S. Bureau of Labor Statistics, 2026a). This is far from a laboratory proof of concept; it represents a federal statistical agency deploying natural language processing at a national surveillance scale.

At the same time, a comprehensive systematic review found that out of 1,255 published studies on artificial intelligence applications in occupational health and safety, only two reported a measurable effect on worker morbidity or mortality outcomes, and neither was of high methodological quality (Jetha et al., 2025). This gap between computational feasibility and demonstrated protective impact is a central tension that this review takes seriously. The literature is proficient at optimizing classification accuracy on historical incident data, yet it remains surprisingly silent on whether deploying these models actually reduces injuries in a live workplace.

A second, distinctly American complication compounds this tension, namely that no single U.S. dataset offers complete national coverage of the occupational injury landscape. The Occupational Safety and Health Administration (OSHA) enforces recordkeeping requirements directly in only about half of the states, while OSHA-approved State Plans cover public- and private-sector workers elsewhere and remain excluded from federal severe-injury reporting (Occupational Safety and Health Administration, 2026). The BLS Survey of Occupational Injuries and Illnesses (SOII) offers a broader scope, but it explicitly excludes fatalities, self-employed individuals, small farms with ten or fewer employees, and federal workers (U.S. Bureau of Labor Statistics, 2026b). Meanwhile, workers' compensation data are administered independently by each state with no unified national claims repository (National Institute for Occupational Safety and Health, 2024). Any nationwide machine learning framework must therefore be built to integrate fragmented and jurisdictionally partial data sources rather than assuming a single authoritative feed exists.

This narrative synthesis evaluates U.S.-focused literature published between 2021 and 2026, with foundational pre-2021 sources cited explicitly as background, to examine how machine learning has been integrated into occupational incident analytics. Ultimately, this review seeks to address a central governing question regarding how fragmented, jurisdictionally partial U.S. OSH datasets can be integrated via predictive machine learning without losing sight of real-world worker outcomes. To answer this question, the review first grounds the discussion in the fragmented landscape of U.S. incident surveillance and data availability, then systematically evaluates ML applications across four key operational domains: severity prediction, text analytics, computer vision, and return-to-work trajectory modeling. It then considers the regulatory and governance frameworks shaping these technologies, analyzes critical cross-cutting limitations in current literature, and concludes by proposing an integrated nationwide architecture designed to move OSH analytics from retrospective tabulation to actionable and proactive protection.

2 OCCUPATIONAL INCIDENT SURVEILLANCE IN THE UNITED STATES

Before machine learning algorithms can be applied to analyze occupational incidents, it is vital to understand what surveillance systems generate the data for these algorithms to interpret. Compared to countries with a single national registry of injuries, the U.S. has a more fragmented surveillance system, consisting of several federal and state programs that estimate the severity of injuries occurring in the workplace (National Institute for Occupational Safety and Health, 2024). These programs have different scopes and coverage, do not always overlap, and include slightly different subsets of injuries.

At the federal level, OSHA establishes and maintains the Severe Injury Reporting (SIR) registry, which requires employers to report any work-related fatalities within eight hours and any inpatient hospitalizations, amputations, or eye loss within twenty-four hours (Occupational Safety and Health Administration, 2026). At the same time, OSHA's jurisdiction covers only half of the states, as the other half have State Plans, which are at least as protective as OSHA but do not report their data to the main federal database (Occupational Safety and Health Administration, 2026). Thus, SIR only contains a partial view of all severe occupational injuries that occur within the U.S. territory.

Another federal-level surveillance system is the Bureau of Labor Statistics (BLS), which conducts the Survey of Occupational Injuries and Illnesses (SOII) to estimate the number of nonfatal illnesses, injuries, and fatalities at the workplace every year. This survey covers establishments, not individual workers, and has several exclusions, including the self-employed, small-scale farms employing fewer than ten workers, and certain other occupations (U.S. Bureau of Labor Statistics, 2026b). At the same time, the fatalities that the SOII covers are limited to those occurring within the establishments, while deaths outside of the workplace are estimated in a separate survey, the Census of Fatal Occupational Injuries (CFOI). The CFOI compiles data from various sources, including death certificates, workers' compensation claims, OSHA logs, and medical examiner reports, to provide an overview of fatal injuries that occur in the workplace (U.S. Bureau of Labor Statistics, 2026c). Overall, BLS provides broader coverage of occupational injuries than OSHA, but with exceptions for certain populations and the absence of coverage for fatalities outside the workplace.

Finally, a separate surveillance system exists for workers' compensation, which helps reimburse workers for a portion of the losses they experience due to occupational injuries. Since workers' compensation is a jointly funded insurance program administered by the states, its administrative procedures, compensations, claim forms, and requirements for reporting on occupational injuries vary from state to state. There is no single federal database for workers' compensation claims, and even NIOSH's Center for Workers' Compensation Studies serves mainly as an analytical and coordinating entity for state programs rather than a comprehensive repository for workers' compensation data (National Institute for Occupational Safety and Health, 2024).

For a machine learning model to process occupational injury data, there are several options, each of which would result in learning only partial information about occupational injuries that occurs nationwide. An algorithm that learns on the OSHA SIR data will understand inpatient hospitalizations, amputations, and eye-loss injuries reported at the federal level, but not fatalities or other types of injuries that do not meet the severity criteria and occur in jurisdictions with State Plans. An algorithm that learns on the SOII data will know about various nonfatal occupational injuries but have no information about fatalities, especially those occurring outside of the workplace. Finally, an algorithm that learns on the workers' compensation claims data, such as Ohio claims or another state's claims, will know about workers' compensation experiences specific to that state but have no information about the rest of the country. According to Jetha et al. (2025), the fragmentation of the U.S. surveillance system is not unique, but the degree to which the United States decentralized occupational injury surveillance makes the gaps in coverage particularly severe for any single jurisdiction. Therefore, when designing a machine learning system for analyzing occupational injuries that applies to the entire nation, it is necessary to remember that any single data source only represents a partial perspective and plan the analysis accordingly.

3 U.S. OCCUPATIONAL SAFETY DATA SOURCES

Turning to the data sources more directly related to the occupational safety domain, the machine learning literature appears to rely on four major ecosystems of information.

3.1 OSHA Severe Injury Reports

The Severe Injury Reporting (SIR) system is formed by free-text descriptions and standardized codes (industry, injury type, body parts, etc.) of severe injuries reported by employers after a serious work-related incident. The number of cases reported to the federal OSHA in 2024 was 9,034 (Khairuddin et al., 2022; Khalifa, 2025).

3.2 BLS Survey Products and Fatality Census

The Bureau of Labor Statistics official web page lists two survey products: the Survey of Occupational Injuries and Illnesses (SOII), yielding estimates of nonfatal occupational injuries and illnesses (U.S. Bureau of Labor Statistics, 2026b) and the Census of Fatal Occupational Injuries (CFOI), providing mortality statistics (U.S. Bureau of Labor Statistics, 2026c). Unlike OSHA's SIR, the BLS surveys use a combination of sampling and multiple record linkage to compile their estimates and do not have a large corpus of texts mined by external entities for machine learning research. Instead, the BLS itself employs natural language processing to tag injury descriptions extracted from SOII and CFOI records, arguably the most extensive application of this technology in the occupational safety domain (U.S. Bureau of Labor Statistics, 2026a).

3.3 MSHA Accident, Injury, and Illness Reports

The jurisdiction of the Mine Safety and Health Administration (MSHA) covers mining-related employment, a domain not subject to OSHA regulations. It has its own accident, injury, and illness database tapped into for negative outcome prediction and text clustering research. In particular, Danish et al. (2025) used the Sentence-BERT transformer architecture and UMAP dimensionality reduction to find patterns in 247,873 MSHA records, grouping similar accidents without human supervision. Similarly, Kumar et al. (2025) experimented with conformal prediction algorithms on

MSHA data to assess the uncertainty of severity predictions, a crucial step that has been overlooked in much of the severity prediction literature.

3.4 State Workers' Compensation Claims

In the absence of a unified national claims database, machine learning researchers rely on single-state administrative datasets obtained through dedicated research partnerships. The most prominent among these are Ohio, offering over 1.2 million claims spanning 2001 to 2011 (Meyers et al., 2018), and California, featuring 2.6 million lost-time claims spanning 2005 to 2019 (Dworsky, 2025). While these rich longitudinal datasets are invaluable for developing automated coding methodologies and analyzing system-level disparities, their single-state scope inherently restricts the national generalizability of any predictive models built upon them.

Taken together, these four ecosystems illustrate a fundamental data fragmentation challenge. Each source offers a distinct lens, whether through enforcement narratives, statistical survey estimates, sector-specific records, or state-bounded administrative claims, on the same underlying phenomenon of workplace injury. Crucially, few machine learning studies in the current literature draw on more than one of these distinct data ecosystems simultaneously.

4 MACHINE LEARNING METHODS IN OCCUPATIONAL INCIDENT ANALYTICS

Having identified the primary data ecosystems, it is possible to classify machine learning models according to the predicted response variables. Thus, severity prediction, which refers to the estimation of the risk that a workplace incident will result in a specific severe consequence, including but not limited to hospitalization, amputation, or days of lost work, is the most common type of predictive model in American studies. Furthermore, most of these models used federal Severe Injury Reporting records as the basis for processing.

4.1 Evaluation Metrics and Imbalance Dynamics

Because predictive modeling studies across the literature report a range of performance metrics, establishing a consistent analytical baseline is essential. Classification models are evaluated using four foundational metrics derived from true positive, false positive, true negative, and false negative counts. Accuracy measures the proportion of correct predictions across all cases. Precision reflects the proportion of positive predictions that are truly positive, while recall or sensitivity measures the proportion of actual positive cases correctly identified by the model. The F1 score provides the harmonic mean of precision and recall, balancing both metrics into a single performance measure.

Importantly, overall accuracy alone can be deeply misleading when applied to highly imbalanced occupational injury datasets, where non-severe outcomes vastly outnumber severe or fatal cases. Consequently, recall and F1 scores provide far more informative evaluations of model performance in high-stakes safety contexts, a point highlighted explicitly across multiple modeling studies.

4.2 Structured Data and Ensemble Approaches

Initial severity prediction frameworks focused on applying classical machine learning algorithms to structured administrative fields such as injury nature, event type, and affected body part. Utilizing a standardized subset of federal severe injury records, early comparative studies evaluated classifiers including Support Vector Machines, K-Nearest Neighbors, Naive Bayes, Decision Trees, and Random Forests to predict hospitalization versus amputation outcomes (Khairuddin et al., 2022). Following feature optimization, Random Forest models achieved strong classification results, reaching high accuracy levels for both hospitalization and amputation predictions.

Subsequent work applying alternative ensemble techniques, such as AdaBoost classifiers, to the same underlying dataset achieved even higher performance bounds (Khalifa, 2025). The coexistence of these varying results using similar datasets highlights a recurring pattern across the literature, demonstrating that reported classification performance is often as sensitive to algorithm selection and feature engineering as it is to the underlying data quality.

4.3 Text-Based and Deep Learning Approaches

Moving beyond structured fields alone, researchers have increasingly incorporated free-text narrative descriptions to capture richer contextual details surrounding incident dynamics. Incorporating sequential neural networks, such as Bidirectional Long Short-Term Memory (Bi-LSTM) architectures combined with TF-IDF and GloVe word embeddings, enabled models to capture contextual text patterns that non-sequential classifiers missed (Khairuddin et al., 2024). Related multimodal frameworks combining structured categorical fields with narrative text using Bi-LSTM networks demonstrated consistently strong predictive metrics across hospitalization and amputation tasks (Khairuddin et al., 2023). More recently, research has advanced to fine-tuning lightweight large language model ensembles on thousands of construction accident reports (Wang et al., 2025). Achieving an F1 score above 0.92, this ensemble approach

represents one of the highest classification performances reported in recent literature. Notably, it also stands out as one of the rare studies in this domain to report explicit statistical confidence bounds alongside its point estimate.

4.4 Regional and Sector-Specific Subsets

While large national datasets dominate the literature, localized studies demonstrate that sector-specific and geographically restricted models can also yield high predictive performance. Focusing exclusively on construction incidents in the Southeastern United States, regional evaluations confirmed that Random Forest and Decision Tree classifiers outperformed alternative algorithms when distinguishing fatal from non-fatal outcomes (Oladele et al., 2025). This indicates that tailored models can be successfully trained even on relatively small, domain-specific subsets of national safety data.

4.5 Prospective Validation Gaps

Despite the high accuracy figures reported across these diverse algorithmic approaches, a critical methodological gap remains. Performance metrics across the majority of studies in this domain are computed primarily on retrospective, held-out test splits drawn from the same historical dataset used during model training. Few studies validate their models against prospective, real-world deployments. This reliance on retrospective test splits highlights the fundamental gap between computational classification capability and demonstrated, prospective injury reduction on active job sites.

5 NATURAL LANGUAGE PROCESSING AND UNSTRUCTURED TEXT ANALYTICS

Beyond the predictive analytics applications described previously, text-based approaches have seen a steady development in natural language processing across occupational safety literature. Three discernible waves of research utilized either keyword extraction and shallow classifiers, sequence models, or modern contextual embeddings to perform different tasks related to improving workplace safety.

5.1 Keyword and Shallow Classifier Approaches

The first class of methods includes either rule-based or shallow machine learning classifiers. The long-standing Bureau of Labor Statistics automated coding project is an example of such an approach, initially using regularized logistic regression with n-gram features and later incorporating more sophisticated architectures (U.S. Bureau of Labor Statistics, 2026a). A similar application was reported in the large-scale auto-coding of workers' compensation claims in Ohio, where rule-based and shallow classification models were used to extract cause of injury free-text fields and organize them into intervention categories based on the analysis of over 1.2 million records (Meyers et al., 2018). The size of such projects testified to the effectiveness of these approaches on a large scale.

5.2 Sequence Models

The next approaches in natural language processing for occupational safety utilized sequence models to better encode the context and word order, as bag-of-words and TF-IDF representations failed to account for linguistic nuances. Bi-LSTM and GRU architectures were applied to either narratives or cause-of-injury fields to incorporate the contextual information in both directions (Khairuddin et al., 2024). Encodings could be further fine-tuned with domain-specific word embeddings to account for specialized terminology.

5.3 Transformers and Large Language Models

The latest line of research utilizes transformer-based approaches both for mining the text for specific insights and performing end-to-end prediction tasks. Local large language model ensembles were recently implemented to extract relevant information from narrative text (Wang et al., 2025). Alternatively, unsupervised clustering with contextual Sentence-BERT embeddings and subsequent UMAP dimensionality reduction was employed to identify hazards with no predefined categories (Danish et al., 2025). In another example, conformal prediction with hierarchical classification was used in the analysis of agency accident records to report not point estimates but rather confidence regions for severity prediction (Kumar et al., 2025).

Despite the wide variety of methods, the theme of all three waves of natural language processing research in occupational safety is similar: encoding of free-text fields either for improved prediction of injury severity or extraction of relevant insights. In all cases, the narrative text was found to contain valuable information that could not be replaced with simpler categorical variables such as industry or body part. However, the value of such research is currently limited by the lack of implementations in actual production. While the Bureau of Labor Statistics auto-coding project is a federal initiative that processes data regularly, other approaches utilizing complex machine learning and NLP pipelines have limited applicability due to the lack of real-time data streams to process.

6 COMPUTER VISION AND REAL-TIME MONITORING

The use of computer vision for occupational safety in the United States is an area that has been scarcely covered by academic literature, compared to either narrative text analysis or severity prediction, although it seems to be more common in practice than reflected in published research.

6.1 Personal Protective Equipment Detection

Researchers have already applied the computer vision paradigm to occupational safety, primarily through the development of real-time object detection systems, most commonly based on the You Only Look Once architecture, for detecting the presence or absence of personal protective equipment in workers' video images. The computer vision system typically evaluates the video stream in real-time, with a specific emphasis on automation for regulatory requirements, such as OSHA 29 CFR 1910.132. Experimental systems typically operate at high frame rates, but the accuracy of detection is often limited in practice due to poor working conditions, video quality, camera position, and other complicating factors (Nath et al., 2020).

A more recent example applies an improved YOLOv8 architecture to real-time detection of fall-prevention equipment (helmets, harnesses, lanyards) on an edge device, reporting 92.52% mAP50 at 9.11 frames per second (Kim & Xiong, 2025). Notably, an extensive search for comparably recent U.S.-authored peer-reviewed computer vision work in this space did not surface a strong candidate; the most active recent contributions instead originate from institutions in Canada, China, South Korea, and Hong Kong, a pattern that itself reinforces this review's broader observation that domestic U.S. academic output in computer vision for occupational safety lags both international research activity and U.S. industry deployment.

6.2 Industry Benchmarking and Commercial Tools

Beyond the prototype level, computer vision systems have seen varying degrees of adoption in the industry, where they are often used in worksites with continuous video surveillance. These tools are typically evaluated in terms of their ability to detect the use of personal protective equipment, as well as other worksite hazards. National Safety Council Work to Zero benchmark reports (National Safety Council, 2022) contain information that can be used to evaluate commercial vision systems, as they include assessments of multiple work safe vision systems, comparing their performance in terms of different indicators such as PPE detection, physical hazard detection, ergonomic risk identification, and physical security. In general, these reports represent one of the few publicly available resources that allow the evaluation of such computer vision systems.

6.3 Intellectual Property and Implementation Ethics

The growing interest in the field is reflected in the number of patents published recently, most of which concern the use of computer vision in combination with other technologies, such as LiDAR, to create multi-sensor safety monitoring systems, for example, for the automated calculation of spatial risk scores. The prevalence of such patents suggests that the commercialization of automated vision-based workplace safety monitoring tools is likely to increase significantly in the near future. At the same time, the practical adoption of such technologies raises multiple ethical, legal, and privacy concerns, which have been recognized by groups such as the American Society of Safety Professionals (American Society of Safety Professionals, 2024), which has published guidelines for safety professionals concerned with computer vision monitoring. Most such recommendations concern the importance of respecting worker privacy, as well as the implementation of monitoring systems in ways that prevent potential biases.

In general, computer vision seems to epitomize the current state of technology in occupational safety analytics. Although the technology allows for remarkable automation and real-time hazard detection, studies that would directly compare the injury-prevention effectiveness of worksite video monitoring are rare.

7 RETURN-TO-WORK PREDICTION AND REINTEGRATION

Beyond simply classifying incidents and their corresponding severity, machine learning and statistical learning have been used to predict future outcomes following an injury. Although return-to-work analytics plays an important role in this domain, literature specific to predictive machine learning appears limited compared to severity prediction or natural language processing.

7.1 Demographic and Structural Predictors

The research establishing foundations for return-to-work analytics mostly uses pooled demographic and structural risk profiles. For example, by analyzing nationwide surveillance datasets encompassing millions of nonfatal injuries, studies have identified increased incidence rates and longer recovery durations among youths ages 15 to 24 years (Guerin et

al., 2020). At the same time, studies using longitudinal data from the Health and Retirement Study investigated return-to-work trends among aging populations, with the increase in the median age of the workforce being an important contextual factor (Baidwan et al., 2018). Trends in working-age populations and the impact of age-specific physiological limitations on return-to-work outcomes appear topical in this literature.

7.2 Economic Dimensions and Productivity Valuation

Several articles similarly sought to quantify the economic consequences of employee absences due to injuries. In this regard, economic evaluations using nationally representative survey data can identify the distribution of costs among industries using case method analysis for complex musculoskeletal injuries (Hoch et al., 2022). Lost productivity due to injuries for a similarly narrow range of cases can also be estimated using commercial claims databases (Peterson et al., 2021). Overall, such findings may be instrumental in informing return-to-work interventions but represent secondary analyses informing broader evidence-generation efforts.

7.3 Literature Gaps and Future Opportunities

An honest assessment of the literature under consideration can note the limitations in the evidence on return-to-work analytics. In particular, review and survey articles on big data in occupational health can note that return-to-work literature remains descriptive in nature and has limited evidence on the ability of machine learning methods to effectively stratify outcomes for different occupational groups or provide accurate predictions in different jurisdictions (Escorpizo et al., 2024; Vivian et al., 2025). In particular, there is a lack of empirical validation for claims of using deep learning or large language models in this setting. Thus, despite opportunities, return-to-work analytics remains a developing area of research and practice, and more empirical investigations are needed before machine learning methods can be effectively deployed in operational settings in the United States to forecast return-to-work outcomes after injuries.

8 NATIONWIDE GOVERNANCE, ALGORITHMIC HYGIENE, AND REGULATION

As the use of machine learning in occupational safety becomes widespread, the United States' public health authorities and regulators have started to develop frameworks for designing, implementing, and regulating artificial intelligence. Instead of focusing on specific binding regulations, the authorities tend to favor voluntary approaches, analogies to risk control, and strategic planning relevant to algorithmic use in the workplace.

8.1 The Algorithmic Hygiene Framework

One of the first institutional attempts to regulate AI applications in the occupational safety context was the adaptation of industrial hygiene principles to algorithmic features. Researchers and public health officials developed an algorithmic hygiene framework (Sadowski & Howard, 2026), which essentially prioritized the control of the machine learning lifecycle failure modes predicted by traditional industrial hygiene. Treating issues such as training sample bias, concept drift, data leakage, or model non-predictiveness as occupational hazards allowed safety practitioners to use a familiar approach to addressing new challenges presented by the use of predictive systems. By classifying potential failures of machine learning systems as hazards and segmenting them according to the principles of administrative control, design engineering, or training set preparation, the algorithmic hygiene framework provides a standardized audit checklist for each stage before the implementation of an algorithm in the workplace environment.

8.2 Other Approaches to Regulation

On the level of workforce safety, other approaches to regulation focus on problems of literacy specific to artificial intelligence applications. Department of Labor representatives have developed an AI literacy framework (U.S. Department of Labor, 2026), which serves as a guide to the content and principles of instruction necessary to achieve competency in artificial intelligence. It is important to note that the competencies outlined in the framework, as well as the AI literacy guide in general, are non-binding in nature, meaning that they highlight recommendations but cannot be directly converted into legislative acts or regulatory restrictions. Maintaining the differentiation between regulatory guidance and law-creating instruments is crucial to understanding the essence of the U.S. approach to AI regulation in works of public health. Federal authorities utilized strategic planning and guidance documents to secure national preparedness for the implementation of artificial intelligence without binding specific regulations for individual employers.

8.3 Regulatory Strategy and Strategic Planning Horizons

To effectively plan its research and service priorities regarding emerging technologies, NIOSH maintains a multi-year strategic plan spanning fiscal years 2019 to 2026, twice extended from its original FY2019-2023 scope (National Institute for Occupational Safety and Health, 2022). This extended timeframe closely overlaps with the review period examined in this paper, and reflects a broader pattern in which federal occupational safety planning documents

increasingly account for the pace of technological change alongside traditional physical and chemical hazards. The plan itself does not brand this as a formal "horizons" framework; rather, it establishes seven strategic goals across sector and cross-sector programs, periodically revised as priorities shift.

9 CHALLENGES, LIMITATIONS, AND RESEARCH GAPS

Upon review of the readings, it becomes apparent that there exist five interrelated shortcomings related to the topic of occupational safety. Each affects the current research literature in different capacities, stemming from either algorithmic design choices or data-related constraints.

First, the outcomes gap is by far the single largest issue discovered in the literature review of relevant scholarly works. When examining the capabilities of artificial intelligence in occupational safety, most predictive models in published works fail to demonstrate improvements to actual safety outcomes. In over a thousand studies reviewed, only two actually measured whether the implementation of an algorithm reduced injuries or illnesses among workers (Jetha et al., 2025). The vast majority classified accuracy, sensitivity, and precision of held-out test sets, which means that the field has limited evidence of its practical efficacy on real-world safety outcomes.

Second, the lack of nationwide datasets accounts for several discrepancies observed in the works. No study provides a truly nation-sized sample of occupational injuries and illnesses, which produces models with various degrees of applicability. Those trained on the federal Severely Injured Workers Reporting system represent a fraction of incidents that have federal jurisdiction involvement (Occupational Safety and Health Administration, 2026). Similarly, models trained on mining-related incidents have a severely limited domain due to the unique hazards of the industry (Danish et al., 2025). Even in cases of single-state workers' compensation data being used for training, such as administrative databases of Ohio (Meyers et al., 2018) or California (Dworsky, 2025), published works fail to generalize their findings beyond the regional economy of the sample.

Third, most statistical findings on predictive safety models utilize point estimates that do not account for uncertainty introduced during the bootstrapping procedure. The vast majority of severity prediction works stated only the point estimate of model performance, with very few providing confidence intervals (Jetha et al., 2025). While recent advances in the field attempt to utilize conformal prediction frameworks and local large language model ensembling to provide bounded error rates (Kumar et al., 2025; Wang et al., 2025), the field should move away from point estimates of accuracy and performance.

Fourth, findings of most works on predictive safety models are regionally or demographically limited. The review identified several works that analyze specific population groups, such as construction workers in the Southeastern United States or elderly persons undergoing occupational therapy (Baidwan et al., 2018; Oladele et al., 2025). As a whole, surveys of big data applications in occupational safety find that few studies conduct a nationwide analysis (Vivian et al., 2025) comparing findings on different demographics with sufficient statistical power to identify trends. Lastly, a review of the administrative state finds that, while recent regulatory developments promote voluntary adoption of AI/ML-based technologies, current federal guidance fails to provide binding requirements for firms utilizing such systems in occupational safety applications (U.S. Department of Labor, 2026). When taken together, these points illustrate that the primary challenges of implementing machine learning methods in occupational safety stem from a confluence of factors, primarily related to the absence of domain-spanning datasets, limited practical efficacy trials, and considerations of uncertainty, demography, and regulation.

10 PROPOSED INTEGRATED NATIONWIDE FRAMEWORK

To address the issues of fragmentation and analytical limitations discovered in the reviewed literature, this study proposes an integrated nationwide framework for machine learning in occupational safety and health. The framework aims to serve as the conceptual foundation for the development of practical solutions that would employ machine learning to predict and prevent work-related injuries while accounting for the federalist nature of the U.S. surveillance system. The framework consists of four foundational pillars.

10.1 Federated Data Interoperability Layer

An integrated nationwide framework requires, first and foremost, a data interoperability layer. Instead of pursuing the elusive goal of establishing the existence of a single feed of national data or attempting to reconcile the differences between federal statistical and state workers' compensation databases, the data layer in the proposed framework is based on the principle of federation. It defines standardized data exchange formats and privacy-preserving API standards between national statistical databases, federal oversight bodies, sector-specific registries, and state workers' compensation systems. To harmonize the requirements of privacy-preserving machine learning and the federated nature of the U.S. data infrastructure, the data layer makes use of state-of-the-art approaches to privacy-preserving

federated learning and secure multiparty computation. The layer thus enables the development of machine learning models that learn from pooled data across jurisdictions without the need to access raw data directly, which would violate privacy guarantees and regulatory separation of powers.

10.2 Tiered Algorithmic Hygiene and Uncertainty Calibration

The second pillar of the proposed framework concerns developing tiered algorithmic responsibility and uncertainty calibration standards. Informed by modern industrial hygiene practices, algorithmic control and oversight are calibrated according to the level of process criticality: low, medium, and high. High-criticality applications, such as computer vision systems for real-time injury risk estimation, have higher algorithmic oversight and reliability expectations than low-criticality applications, such as post hoc statistical coding of survey items. However, a universal requirement is introduced by the second pillar is the explicit uncertainty calibration that employs conformal prediction algorithms. The framework suggests that before any automated analytical system, including machine learning and computer vision applications, is approved for production use in occupational safety settings, it must demonstrate measurable and reliable performance characteristics on prospective data. In particular, the framework suggests that automated systems should be capable of providing uncertainty estimates that go beyond point estimates (e.g., probability of incident) and explicitly state safety thresholds, thus enabling informed human judgment in safety management.

10.3 Outcome-Linked Evaluation Protocols

The third pillar concerns the development of outcome-linked evaluation protocols that define the requirements for prospective validation of predictive performance. Predictive analytics has a well-deserved reputation of producing overly optimistic performance estimates due to the use of convenient but unrepresentative datasets (e.g., retrospective data). According to the third pillar, the requirement for positive predictive validity (i.e., a demonstrated ability of a predictive model to actually affect the incidence rate of incidents or injuries in real life) is critical for the adoption of any predictive analytics tools in occupational safety settings. The framework recommends that before predictive models, including computer vision systems, are adopted for use in occupational safety settings, they must go through prospective validation studies that assess their impact on safety outcomes, such as incidence rates, near miss reporting rates, hazard management efficacy, and worker safety behaviors. In other words, an analytical tool cannot be used to optimize incidence rate calculations without demonstrating that it can actually affect the frequency of incidents.

10.4 A Phased Transition Toward Regulatory Compliance

The fourth and final pillar concerns the development of a phased transition toward regulatory compliance. The framework suggests developing voluntary standards for the adoption and use of machine learning applications in workplaces during the early stages of maturity. Subsequently, the transition from voluntary standards to regulatory compliance can be initiated as analytical methods and tools mature and sufficient outcome evidence is accumulated. This approach to regulatory transitions was successfully implemented in the domain of industrial hygiene by first developing voluntary exposure assessment guidelines and later transitioning to regulatory approvals for specific exposure control technologies. By proposing a similar approach, the integrated framework promotes the establishment of a culture of algorithmic safety that would enable the safe adoption of emerging technologies while protecting workers from unsubstantiated interventions. In particular, the framework suggests that voluntary (but later standardized) guidelines and technical standards must be developed, which would then create the legal basis for the regulation of automated systems in occupational safety settings. Thus, as the body of evidence accumulates, regulatory agencies would gain jurisdiction over algorithmic auditing, bias testing, data governance, and other aspects of automated decision-making.

To summarize the operational mechanisms and strategic objectives of this nationwide model, the four core pillars of the proposed framework are structured across their target limitations, technical mechanisms, and intended protective outcomes in **Table 1**.

Table 1: Key Pillars of the Proposed Integrated Nationwide Framework

Framework Pillar	Limitation Addressed	Core Mechanism	Intended Outcome	Safety
1. Federated Data Interoperability Layer	Dataset fragmentation and jurisdictional silos (OSHA, BLS, MSHA, State Comp)	Standardized data exchange formats, privacy-preserving APIs, federated learning, secure multiparty computation	Models trained across jurisdictions without centralizing	or

			exposing raw data
2. Tiered Algorithmic Hygiene and Uncertainty Calibration	Uncalibrated point-estimate predictions; uniform oversight regardless of application risk	Risk-tiered oversight (low/medium/high mandatory prediction for uncertainty bounds)	Explicit confidence and safety thresholds supporting human judgment, especially for high-criticality uses
3. Outcome-Linked Evaluation Protocols	The retrospective outcomes gap: high accuracy on historical splits without demonstrated real-world effect	Mandatory validation studies measuring incidence rates, near-miss reporting, and hazard management efficacy	Empirical evidence that a tool reduces incidents before broader deployment
4. Phased Transition Toward Regulatory Compliance	Current governance is voluntary and non-binding.	Voluntary standards maturing into binding regulation as evidence accumulates, modeled on industrial hygiene's own historical transition	Legal and regulatory basis for algorithmic auditing, bias testing, and data governance

11 CONCLUSION

The machine learning literature in the context of US occupational safety is characterized by a high level of sophistication in terms of computational complexity. The practice starts at the largest statistical agencies performing production-scale transformer-based language models and extends to scientific experiments achieving excellent scores in terms of classification accuracy on various narrative corpora. Nevertheless, the field is still overwhelmingly retrospective. The primary direction for future research should be focused on bridging the gap between the accuracy of machine learning predictions and tangible increases in occupational safety. The area requiring immediate attention is the application of positive results from individual studies to nationwide practice. To achieve this goal, the researchers should consider designing studies with multiple datasets at once, engaging in field trials with predictive models, developing valid uncertainty estimation frameworks, and defining regulatory response surfaces. This approach will allow occupational safety analysts to truly shift from retrospective descriptions to the actual improvement of workplace safety.

REFERENCES

[1] American Society of Safety Professionals. (2024). *Position statement on the use of artificial intelligence in occupational safety and health*. ASSP.

[2] Baidwan, N. K., Gerberich, S. G., Kim, H., Ryan, A. D., Church, T. R., & Capistrant, B. (2018). A longitudinal study of work-related injuries: Comparisons of health and work-related consequences between injured and uninjured aging United States adults. *Injury Epidemiology*, 5(35), 1–9. <https://doi.org/10.1186/s40621-018-0166-7>

[3] Danish, A. A. K., Chatterjee, S., Raj, K. V., Kaunda, R., Miller, H., & Majdara, A. (2025). Accidents analysis of mining industry through semantic text representation and dimensionality reduction: An integrated clustering framework. *Engineering Applications of Artificial Intelligence*, 161, 112164. <https://doi.org/10.1016/j.engappai.2025.112164>

[4] Dworsky, M. (2025). Racial and Ethnic Disparities in Occupational Health. *JAMA Health Forum*, 6(9). <https://jamanetwork.com/journals/jama-health-forum/fullarticle/2839235>

[5] Escorpizo, R., Theotokatos, G., & Tucker, C. A. (2024). A Scoping Review on the Use of Machine Learning in Return-to-Work Studies: Strengths and Weaknesses. *Journal of Occupational Rehabilitation*, 34, 71–86. <https://doi.org/10.1007/s10926-023-10127-1>

[6] Guerin, R. J., Reichard, A. A., Derk, S., Hendricks, K. J., Menger-Ogle, L. M., & Okun, A. H. (2020). Nonfatal Occupational Injuries to Younger Workers—United States, 2012-2018. *MMWR Morbidity and Mortality Weekly Report*, 69(35), 1204–1209. <https://doi.org/10.15585/mmwr.mm6935a3>

[7] Hoch, C. P., Caughman, A., Griffith, A., Rex, J. C., Scott, D. J., & Gross, C. E. (2022). A Detailed Analysis of Workplace Foot and Ankle Injuries. *Foot & Ankle Orthopaedics*, 7(1). <https://doi.org/10.1177/2473011421S00241>

- [8] Jetha, A., Bakhtari, H., Irvin, E., Biswas, A., Smith, M. J., Mustard, C., Arrandale, V. H., Dennerlein, J. T., & Smith, P. M. (2025). Do occupational health and safety tools that utilize artificial intelligence have a measurable impact on worker injury or illness? Findings from a systematic review. *Systematic Reviews*, 14(146), 1–10. <https://doi.org/10.1186/s13643-025-02869-1>
- [9] Khairuddin, M. Z. F., Hasikin, K., Abd Razak, N. A., Lai, K. W., Osman, M. Z., Aslan, M. F., Sabanci, K., Azizan, M. M., Satapathy, S. C., & Wu, X. (2022). Predicting occupational injury causal factors using text-based analytics: A systematic review. *Frontiers in Public Health*, 10, 984099. <https://doi.org/10.3389/fpubh.2022.984099>
- [10] Khairuddin, M. Z. F., Hasikin, K., Razak, N. A. A., Mohshim, S. A., & Ibrahim, S. S. (2023). Harnessing the Multimodal Data Integration and Deep Learning for Occupational Injury Severity Prediction. *IEEE Access*, 11, 85284–85302. <https://doi.org/10.1109/ACCESS.2023.3304328>
- [11] Khairuddin, M. Z. F., Sankaranarayanan, S., Hasikin, K., Abd Razak, N. A., & Omar, R. (2024). Contextualizing injury severity from occupational accident reports using an optimized deep learning prediction model. *PeerJ. Computer Science*, 10, e1985. <https://doi.org/10.7717/peerj-cs.1985>
- [12] Khalifa, W. M. S. (2025). Work Place Safety: Machine Learning Techniques for Assessing Workplace Incident Severity. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2025.3543136>
- [13] Kim, D., & Xiong, S. (2025). Enhancing Worker Safety: Real-Time Automated Detection of Personal Protective Equipment to Prevent Falls from Heights at Construction Sites Using Improved YOLOv8 and Edge Devices. *Journal of Construction Engineering and Management*, 151(1), 04024187. <https://doi.org/10.1061/JCEMD4.COENG-14985>
- [14] Kumar, A., Senapati, A., Upadhyay, R., Chatterjee, S., Bhattacharjee, A., & Samanta, B. (2025). An uncertainty-aware decision support system: Integrating text narratives and conformal prediction for trustworthy accident code classification. *Process Safety and Environmental Protection*, 204. <https://www.sciencedirect.com/science/article/abs/pii/S0957582025014016>
- [15] Meyers, A. R., Al-Tarawneh, I. S., Wurzelbacher, S. J., Bushnell, P. T., Lampl, M. P., Bell, J. L., Bertke, S. J., Robins, D. C., Tseng, C.-Y., Wei, C., Raudabaugh, J. A., & Schnorr, T. M. (2018). Applying Machine Learning to Workers' Compensation Data to Identify Industry-Specific Ergonomic and Safety Prevention Priorities: Ohio, 2001 to 2011. *Journal of Occupational and Environmental Medicine*, 60(1), 55–73. <https://doi.org/10.1097/JOM.0000000000001162>
- [16] Nath, N. D., Behzadan, A. H., & Paal, S. G. (2020). Deep learning for site safety: Real-time detection of personal protective equipment. *Automation in Construction*, 112, 103085. <https://doi.org/10.1016/j.autcon.2020.103085>
- [17] National Institute for Occupational Safety and Health. (2022). *NIOSH Strategic Plan: FYs 2019-2026 (Version 6)*. Centers for Disease Control and Prevention. <https://www.cdc.gov/niosh/about/strategicplan/>
- [18] National Institute for Occupational Safety and Health. (2024). *Center for Workers' Compensation Studies*. Centers for Disease Control and Prevention. <https://www.cdc.gov/niosh/centers/workers-comp.html>
- [19] National Safety Council. (2022). *Using Computer Vision as a Risk Mitigation Tool*. Work to Zero Initiative. <https://www.nsc.org/getmedia/2499ccbe-1bfe-4992-9cbe-111db89a9a0f/computer-vision-camera-analytics-tech-wp.pdf>
- [20] Occupational Safety and Health Administration. (2026). *2024 Annual Report of Severe Injuries and Illnesses Reported from Employers Covered by Federal OSHA*. U.S. Department of Labor. https://www.osha.gov/sites/default/files/2024_Annual_Report_of_Fed_OSHA_SIRs.pdf
- [21] Oladele, M., Jeong, M. M., & Jung, Y. E. (2025). *Classification of Fatal and Non-fatal Construction Incidents in the Southeastern U.S. Using Machine Learning*. 6, 856–865. <https://easychair.org/publications/paper/h45l/download>
- [22] Peterson, C., Xu, L., & Barnett, S. B. L. (2021). Average lost work productivity due to non-fatal injuries by type in the USA. *Injury Prevention*, 27(2), 111–117. <https://doi.org/10.1136/injuryprev-2019-043607>
- [23] Sadowski, J. P., & Howard, J. (2026). Practical Strategies to Manage AI Hazards in the Workplace. In *NIOSH Science Bulletin*. Centers for Disease Control and Prevention. <https://www.cdc.gov/niosh/bulletin/2026/practical-strategies-to-manage-ai-hazards-in-the-workplace.html>
- [24] U.S. Bureau of Labor Statistics. (2026a). *Automated Coding of Injury and Illness Data*. U.S. Department of Labor. <https://www.bls.gov/iif/automated-coding.htm>
- [25] U.S. Bureau of Labor Statistics. (2026b). *Employer-Reported Workplace Injuries and Illnesses, 2023-2024* (USDL-26-0101). U.S. Department of Labor. <https://www.bls.gov/news.release/pdf/osh.pdf>

- [26] U.S. Bureau of Labor Statistics. (2026c). *National Census of Fatal Occupational Injuries in 2024* (USDL-26-0230). U.S. Department of Labor. <https://www.bls.gov/news.release/pdf/cfoi.pdf>
- [27] U.S. Department of Labor, E. and T. A. (2026). *Training and Employment Notice No. 07-25: The U.S. Department of Labor's Artificial Intelligence Literacy Framework*. <https://www.dol.gov/agencies/eta/advisories/ten-07-25>
- [28] Vivian, G. A., Bauder, R. A., & Khoshgoftaar, T. M. (2025). A comprehensive survey on machine learning for workplace injury analysis: Risk prediction, return to work strategies, and demographic insights. *Journal of Big Data*, 12(167). <https://doi.org/10.1186/s40537-025-01229-z>
- [29] Wang, J., Wu, S., Qu, Z., Shen, J., Memon, S. A., Skitmore, M., & Ma, L. (2025). Classifying OSHA construction accident reports: Leveraging ensemble learning and lightweight large language models (l-LLMs). *Engineering, Construction and Architectural Management*. <https://doi.org/10.1108/ECAM-08-2025-1273>