

---

# Predictive Analysis of Causal Factors Influencing Occupational Accidents in Construction Workplaces Using a Machine Learning Unified Data Model

---

[Hajar Ait Lamkademe](#)\*, Ahmed Naddami, [Karim Choukri](#)

Posted Date: 20 September 2024

doi: 10.20944/preprints202409.1611.v1

Keywords: machine learning; predictive analytics; data mining; safety; occupational accidents; construction; workplace; artificial intelligence; big data; incidents



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

# Predictive Analysis of Causal Factors Influencing Occupational Accidents in Construction Workplaces Using a Machine Learning Unified Data Model

Hajar Ait Lamkademe <sup>1,\*</sup>, Ahmed Naddami <sup>1</sup> and Karim Choukri <sup>2</sup>

<sup>1</sup> Laboratory of Advanced Research in Industrial and Logistics Engineering (LARILE), Higher National School of Electricity and Mechanics (ENSEM) | Hassan II University of Casablanca

<sup>2</sup> EEIS Laboratory, ENSET Mohammadi, Hassan II University of Casablanca

\* Correspondence: hajar.aitlamkademe-etu@etu.univh2c.ma; Tel.: +212-6-18-43-17-56

**Abstract:** Occupational incidents in construction workplaces present a persistent challenge, with severe consequences for both workers and project outcomes. This study aims to examine the specific causal factors contributing to these incidents, focusing on the interaction between worker-related, environmental, and procedural factors. The objective is to identify key contributors to workplace accidents and develop a unified predictive model to mitigate future risks. To achieve this, we employed a machine learning approach on a comprehensive accidentology dataset collected from contractors across multiple construction sites. The dataset includes variables such as worker experience, environmental conditions, adherence to safety protocols, and more. By analyzing direct, indirect, and root causes, the methodology uncovers hidden patterns and interdependencies that traditional analysis might overlook. The study's findings indicate that worker experience and environmental factors are the most significant contributors to incident occurrence, with a clear interaction effect between these variables. The results not only confirm previous research but also offer enhanced predictive capabilities for future safety measures. This research demonstrates the value of machine learning in generating data-driven insights, ultimately aiding in the development of targeted interventions to improve safety standards in the construction industry.

**Keywords:** machine learning; predictive analytics; data mining; safety; occupational accidents; construction; workplace; artificial intelligence; big data; incidents

## 1. Introduction

The construction industry is inherently hazardous, with a high incidence of occupational accidents and fatalities. Despite advances in safety protocols and regulations, construction workplaces continue to experience significant numbers of incidents that not only result in human suffering but also incur substantial economic costs. Understanding the causal factors behind these incidents is essential for developing effective prevention strategies and improving overall workplace safety.

Traditional approaches to investigating construction accidents often involve retrospective analyses of incident reports and expert surveys. While these methods have provided significant insights, they are limited by issues such as underreporting, subjectivity, and the inability to handle large volumes of data. Therefore, Machine Learning is used to uncover complex patterns and relationships within large datasets that are not easily discernible through conventional methods (APC, et al. 2023).

Building upon recent research ((Chen and al. 2020), (Khan, et al. 2023)), data preprocessing and optimal machine learning techniques have been proposed to handle the complexity and diversity of construction accident data, enabling the derivation of meaningful variables and correlations (Lee, et al. 2020). A comprehensive review of machine learning applications in occupational accident analysis identified key research domains and highlighted the need for further exploration of ML algorithms

in this field (Sarkar and & Maiti, Machine learning in occupational accident analysis: A review using science mapping approach with citation network analysis. 2020).

This paper presents a comprehensive data exploration of accidentology historic and the application of predictive analytics in construction workplace safety to uncover the impact of causalities on the occurrence of all types of workplace incidents, focusing on a machine learning approach.

In this present work, we advocate for a proactive approach to dynamic adaptation. The benefits can be summarized as follows: **1)** avoiding oscillatory safety practices/Plans of Action **2)** managing allocation of exhaustible resources, and **3)** proactivity in front of seasonal behavior. Focusing on this project, we propose to enhance dynamic adaptation by adopting a computational model on the historic HSE Data of every Business Unit's project to predict future performance and using real-time statistics to produce reliable predictions on future events. Thus, we leverage the business decision making from the area of static analysis to operational dynamic predictive analysis.

For better pre-analysis visualization, the literature is divided into four operationally defined predictive dataset families in order to facilitate understanding about safety prediction: **(1)** Safety Observation Reports (SOR) dataset that take into account the climate assessment of hazards and characteristics of the work; **(2)** Accidentology dataset that contain the historic dataset of the event; **(3)** Key Performance Indicators (KPI) dataset that take into account the leading and lagging safety indicators and the event tracking reflecting the volume of safety management activities; and **(4)** Training dataset as part of the safety metrics in consideration.

Within the existing body of investigation paper, a single model is suggested where all dataset families are taken into consideration together as well as allowing opportunities for synergy and cross-validation are to be exploited. As perspectives of the present study, this model's application can be expanded to produce more precise and reliable safety predictions that take into consideration the interconnections of the work-related attributes, human resources, and management approaches which impact safety.

By synthesizing and exploiting recent advancements in both predictive analytics and construction safety management, we aim also to address the importance of direct, indirect and root cause correlation analysis with the occurrence of accidents per type and criticality. Specifically, we investigate the development and implementation of predictive models capable of identifying safety risks in real-time, thereby enabling proactive interventions to prevent accidents and injuries.

## **2. Materials and Methods**

### *2.1. Related Works*

The construction industry is a high-risk sector, with a significant number of work-related accidents. To address this, recent research has explored the use of machine learning (ML) and predictive analytics to enhance workplace safety. (Cavalcanti, Lessa and & Vasconcelos 2023) conducted a systematic review of ML applications in construction accident prevention, highlighting the need for further studies in this area. (Fargnoli and Lombardi 2020) emphasized the potential of Building Information Modelling (BIM) to improve occupational safety in construction activities, suggesting practical applications such as safety training and risk analysis.

(Gao, et al. 2019) developed a model using ML and the Big Five personality taxonomy to predict construction workers' safety behavior, identifying workers prone to unsafe behaviors. (Baker, Hallowell and Tixier 2020) further advanced this work by using ML to predict independent construction safety outcomes, achieving significant improvements in injury severity prediction. These studies collectively underscore the potential of ML and predictive analytics in enhancing construction workplace safety.

One study developed a predictive model using machine learning to identify the potential risk of fatality accidents at construction sites, utilizing a dataset from the Ministry of Employment and Labor of the Republic of Korea (Jongko Choi et al., 2020). The study found that the random forest method had the highest predictive success rate, with influential factors including the month of the accident and employment size. Another research effort focused on predicting the consequences of

construction accidents in China, analyzing 16 critical factors with eight different algorithms (Zhu, et al. 2021). The study highlighted the importance of the 'Type of accident' and 'Accident reporting and handling' as critical factors, with Naive Bayes and Logistic regression achieving the best F1-Score on the raw dataset.

Predictive modeling has also been applied to the mining industry, which shares similar safety concerns with construction. Machine learning models, including decision trees and artificial neural networks, were used to predict outcomes of mining accidents and days away from work, with narrative data providing additional insights compared to structured data (Yedla, Kakhki and Jannesari 2020).

A framework for predicting safety performance before the implementation of construction projects was proposed in another study (Abbasianjahromi and & Aghakarimi 2021). It utilized a decision tree algorithm coupled with the k-Nearest Neighbors algorithm, identifying key criteria for safety performance prediction such as safety employees, training, rule adherence, and management commitment.

A comprehensive literature review on ML applications in construction safety literature revealed trends and gaps in the field (Koc and & Gurgun 2021). It found that severity evaluation of construction accidents was the most widely investigated sub-topic, with linear regression and logistic regression commonly used as benchmark models. The performance of machine learning techniques in predicting injury severity in agribusiness industries was tested, with models achieving high accuracy rates (Kakhki, Freeman and Mosher 2019). This study emphasized the importance of quantitative analysis of empirical injury data in safety science.

The integration of machine learning into construction safety management has shown promising results, with various studies demonstrating the effectiveness of different algorithms in predicting accident outcomes. These advancements hold the potential to significantly improve safety measures and reduce the incidence of accidents in the construction industry as we will argue in this paper. The following (see Table 1) are some of the leading predictive models in the construction safety domain mentioned in our state of art and their relevant information.

**Table 1.** Inventory of Reviewed Safety-Related Studies in Predictive Approaches.

<i>Study</i>	<i>Main findings</i>	<i>Authors, Year</i>	<i>DOI</i>
Machine Learning Predictive Model Based on National Data for Fatal Accidents of Construction Workers	Machine learning can effectively predict fatal accidents at construction sites, with month, employment size, age, weekday, and service length being the most influential factors.	Jongko Choi, Bonsung Gu, Sangyoon Chin, Jong-seok Lee (2020)	10.1016/j.autcon.2019.102974
Application of Machine Learning Techniques for Predicting the Consequences of Construction	Naive Bayes and Logistics regression are the best machine learning algorithms for predicting the severity of construction accidents, with accident type, reporting, and handling	Rongchen Zhu, Xiaofeng Hu, Jiaqi Hou, Xin Li (2021)	10.1016/j.psep.2020.08.006

<i>Study</i>	<i>Main findings</i>	<i>Authors, Year</i>	<i>DOI</i>
Accidents in China	being the most critical factors.		
Predictive Modeling for Occupational Safety Outcomes and Days Away from Work Analysis in Mining Operations	Machine learning techniques, such as decision trees and random forests, can improve mining safety by predicting accident outcomes and days away from work.	Anurag Yedla, Fatemeh Davoudi Kakhki, A. Jannesari (2020)	10.3390/ijerph17197054
Customized AutoML: An Automated Machine Learning System for Predicting Severity of Construction Accidents	Customized AutoML is an automated machine learning system that accurately predicts construction accident severity for professionals with limited data science knowledge, offering higher scalability, accuracy, and result-oriented insight.	V. Toğan, F. Mostofi, Y. Ayözen, Onur Behzat Tokdemir (2022)	10.3390/buildings12111933
Safety Performance Prediction and Modification Strategies for Construction Projects Via Machine Learning Techniques	The decision tree algorithm effectively predicts safety performance in construction projects, with safety employees, training, rule adherence, and management commitment being key criteria.	H. Abbasianjahromi, Mehdi Aghakarimi (2021)	10.1108/ecam-04-2021-0303
Component-Based Machine Learning for Performance Prediction in Building Design	This paper presents a component-based machine learning approach for predicting building performance, enabling high prediction quality with errors as low	P. Geyer, Sundaravelpandian Singaravel (2018)	10.1016/J.APENERGY.2018.07.011

<i>Study</i>	<i>Main findings</i>	<i>Authors, Year</i>	<i>DOI</i>
	as 3.7% for cooling and 3.9% for heating.		
Machine Learning Applications in Construction Safety Literature	Machine learning methods, particularly support vector machine and decision tree, are widely used in construction safety literature to predict accident outcomes and identify potential safety risks.	K. Koc, A. Gurgun (2021)	10.14455/isec.2021.8(1).csa-05
Evaluating Machine Learning Performance in Predicting Injury Severity in Agribusiness Industries	Machine learning techniques can accurately predict injury severity in agribusiness industries using workers' compensation claims, with a 92-98% accuracy rate.	Fatemeh Davoudi Kakhki, S. Freeman, G. Mosher (2019)	10.1016/j.autcon.2019.102974

## 2.2. Process Overview

The purpose of the following section is to point out the various techniques used to make safety predictions in our use case. When applicable, this paper tracks the progression of research done. In the review of each method of safety prediction in our use case, the following were discussed:

- Attributes or units of the analysis, which we consider to be the information used as the input data to the predictive model.
- Summary of approaches implemented in past research, including framing of the problem, sources of data, analytical techniques, and strengths and weaknesses as they apply to scientific validity and reliability.

Also in our study, we have presented and organized safety prediction research into four major families, based upon the information that has been provided by the organization. Our work was in a pilot phase aimed at evaluating the techniques available. The potential of data-mining techniques not only derived from the possibility for processing large quantities of data but also from the following:

- Their capacity to deal with large-dimension problems, which is necessary when endeavoring to identify relevant variables among many potential factors.
- Their flexibility in reproducing the data-generation structure, irrespective of complexity, thanks to a non-linear structure that is adaptable to the data (non-parametric philosophy).
- Their great predictive and, in some cases, interpretative, potential.

Many manuscripts in the study were, therefore, an attempt to organize the widely dispersed literature on these topics. Then, we suggest a unified model of safety prediction that leverages the unique features of the four families of prediction while also exploiting overlap among the methodologies as a potential source of cross-validation.

While there are significant differences among the predictive families, the goals are similar. Thus, we postulate that there is an opportunity for synergy and joint prediction by simultaneously applying multiple techniques. Before suggesting a unified model for safety prediction, we examined the

theoretical connections among the families of prediction upon the global project process flowchart illustrated in Figure 1.

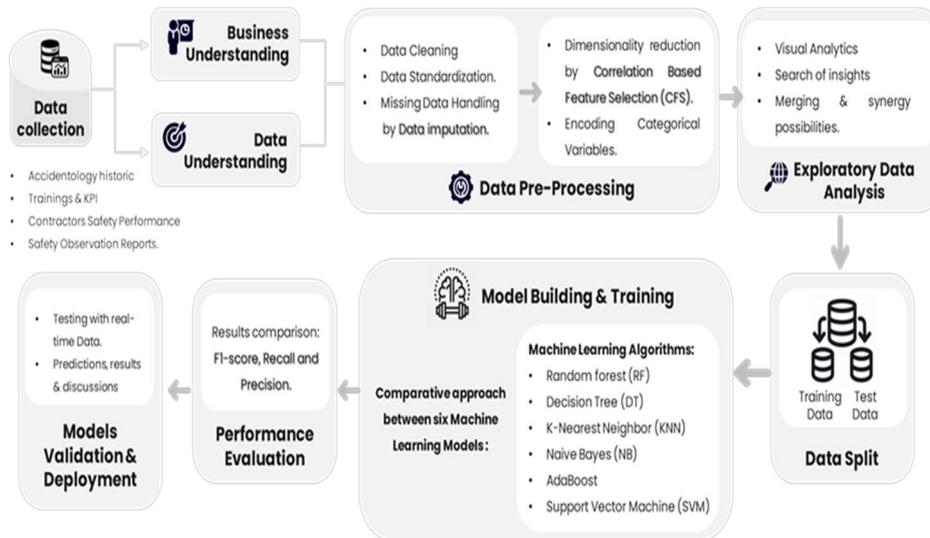


Figure 1. Project process flowchart.

Some families of safety prediction measure fundamentally different dimensions of the safety system. When the methods are completely independent, we postulate that the combination of methods offers synergy and improved predictive accuracy. When the methods measure fundamentally similar aspects of the safety system, we postulate that this is an opportunity for cross-validation.

### 2.3. Data in Use

As we have delved into the examination of safety prediction methods relevant to our use case, we aimed to explore diverse safety prediction techniques to be utilized and enhanced in our study and, when pertinent, to chronicle the research evolution for it to trace a scientific improvement to the business case.

Unlike previous studies that predominantly focused on a singular dataset, a distinctive challenge encountered in our study lies in managing and integrating multiple categories of datasets. This multifaceted data landscape necessitates a more intricate approach to data preprocessing, feature engineering, and model development to ensure comprehensive and accurate multiple predictive modeling (see Figure 2).

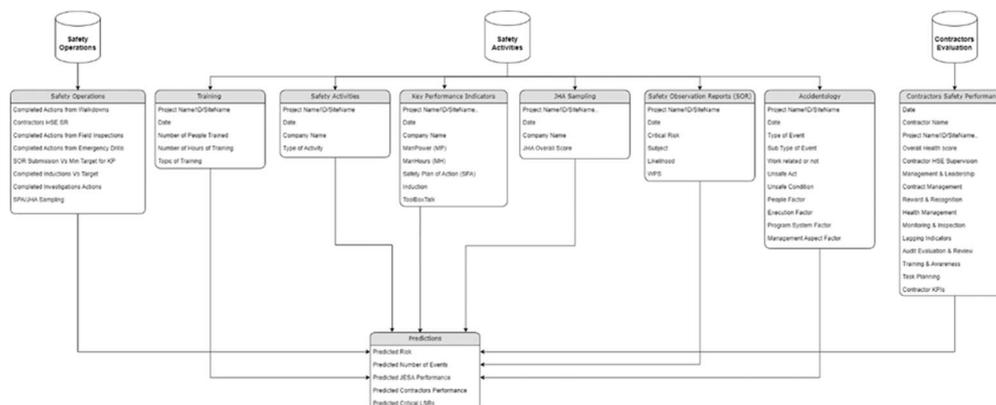


Figure 2. Datasets and Variables in Use.

The study population for this use case was derived from a comprehensive historical accidentology dataset encompassing over 103 construction sites. This dataset includes a diverse

workforce exceeding 132,500 employees distributed across all locations and employed by more than 3,000 contractors. The construction sites engage in 15 critical activities daily, providing a realistic simulation of the inherently chaotic and hazardous conditions typical of construction environments. This robust dataset allows for a more nuanced and comprehensive analysis of accident-related causes.

Each recorded incident is meticulously documented with detailed descriptions of the direct, indirect, and root causes, which are further categorized into subcategories (see Table 2).

**Table 2.** Causal Factors Classification.

Injury Categories	Direct Causes Categories	Indirect Causes Categories	Root Causes Categories
- First Aid Case - Medical Treatment Case - Restricted Work Case - Lost Time Injury - Asset Damage	<b>Unsafe Act (UA)</b> - Individual behavior/ attitude - Tools or Equipment Use - Procedures implementation	<b>People Factor (PF)</b> - Physical Capabilities - Mental Capabilities - Physiological	<b>Management Aspect (MA)</b> - Resource Management - Leadership - Contractors & Subcontractor Mgt.
	<b>Unsafe Condition (UC)</b> - Workplace Hazards - Process Hazards - Tools & Equipment Condition - Protective Defenses - Weather conditions	<b>Execution Factor (EF)</b> - Engineering / Design - Project level execution - Communication - Skill & Knowledge - Tools & Equipment Provision	<b>Program System Aspect (PSA)</b> - Work Standards / Procedures - Risk Evaluation - Task Planning - Training - Inspection and Audit program

This extensive categorization facilitates a thorough assessment and evaluation of the correlations and potential causal links leading to the occurrence of injuries on-site. The richness and granularity of the data enable the identification of patterns and trends that are critical for understanding the underlying factors contributing to workplace accidents.

Consequently, this analysis not only sheds light on the immediate and obvious causes of accidents but also delves into the more complex and interrelated factors that can precipitate such incidents, providing valuable insights for improving safety protocols and preventive measures in the construction industry.

#### 2.4. Synergy Possibilities

- Key Performance Indicators (KPI) And Training Datasets: A Synergy of Long-Term Predictive Methods:

The two families of safety prediction, safety leading indicators and training dataset, are long-term in nature and are typically measured over weeks and months. Therefore, they are not useful in making situational predictions; rather, they are used to forecast injury rates over months or even years. Although training and safety leading indicators measure different aspects of the safety system, they may not be completely independent.

In contrast, safety leading indicators measure the quality of safety management activities (Hinze, J, Hallowell, M. et Baud, K. 2013) and training measures general perception of safety which may include perceptions of safety management. For instance, perception on management's commitment

to safety, perception on supervisor's role, or perception on adequacy and efficacy of training (i.e., safety climate dimensions) may be influenced by quantity and quality of training programs, number of audits, and type of incentive programs (i.e., leading indicators). Therefore, we postulate that using training and safety leading indicators in concert may offer both synergistic and cross-validation opportunities.

- Safety Observation Reports (SOR) And Preliminary Event Notifications (PEN): A Synergy of Situational Predictive Methods:

Safety Observation Reports (SOR) and Preliminary Event Notifications (PEN)/Accidentology are considered here as situational methods because they attempt to make safety predictions for single events based upon information available for a specific work situation. For the situational methods, we postulate that the two methods can be used synergistically to make predictions that are far more robust than a single method applied in isolation.

For instance, the predictions with Preliminary Event Notifications produce very strong predictions about the type of injury and its related direct, indirect and root causes. However, these attributes of the work do not predict the severity of the injury. Fortunately, the safety observation reports, makes a prediction of severity and has shown skill in predicting the likelihood that an injury will occur by differentiating success from failure (Alexander, D., Hallowell, M. et Gambatese, J 2017). So, both methods attempt to make situational predictions but are based upon fundamentally different aspects of the safety system.

### *2.5. Synergy and Cross-Validation among Situational and Long-Term Methods*

To review all possible interactions among the families of safety prediction, there are some interesting potential combinations of methods that extend across single timeframes:

First, there is logical synergy between safety leading indicators and safety risk assessment by Safety Observation Reports (SOR), even though one method is situational and the other is long-term. Colloquially, leading indicators measure the effort over relatively long timeframes and safety risk measures the danger of specific work. Theoretically, safety outcomes are ultimately defined by the balance between the effort and the danger of the work, where more dangerous work demands more organizational effort for safety management. We postulate that these two families can be used synergistically to better predict safety performance, but we do not believe that these two families can be used for cross-validation because they measure fundamentally different aspects of safety.

Second, there is a logical opportunity for cross-validation among safety leading indicators (KPI) and Preliminary Event Notification. Ultimately, an organization must choose the safety activities that will be performed and how often. This decision logically impacts safety performance as more activities have been correlated to improved future safety performance. In a more nuanced and indirect way, increased safety efforts (e.g., increased training) should manifest in reduced prevalence of precursors (e.g., workers not knowing the safe work procedure or not performing strong pre-job planning).

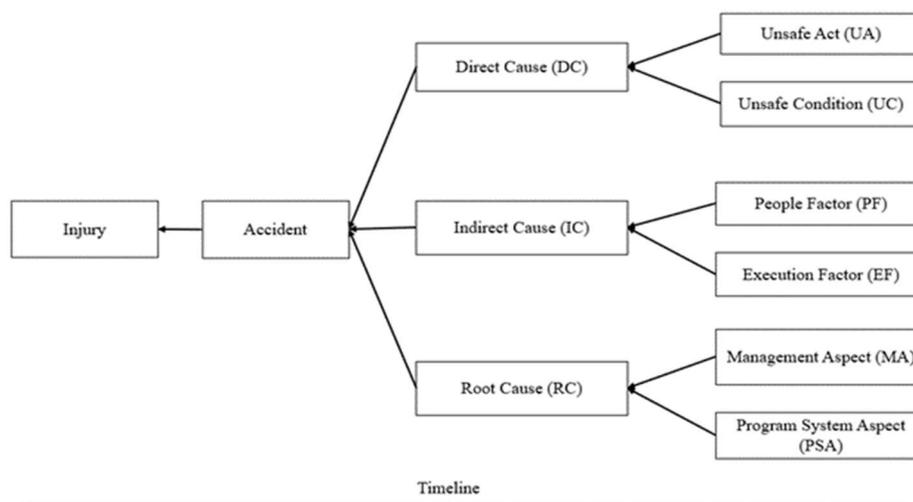
Although not all precursors are connected to all safety leading indicators, there are some indirect relationships. Therefore, as an organization increases or decreases safety efforts, they should expect to see corresponding changes in the prevalence of specific precursors. There is a logical hierarchy in these relationships where safety leading indicators are the primary drivers and precursor scores mediate the relationship to injury rates.

### *2.6. Research Hypothesis*

The suggested model, illustrated in Figure 3, based originally on the Heinrich's domino model of accident causation (Heinrich 1941) aims to explore and analyze and defend the assumptions aligned with this study. This model will be specifically applied to the construction industry.

The hypothesis for this study are whether the various subcategories originally linked to the direct, indirect and root causes have an underlying impact on the occurrence of occupational

accidents per type (hypothesis 1) and whether this impact can be predictively materialized in terms of accuracy, for predicting the different types of accidents (hypothesis 2).



**Figure 3.** Projection of Heinrich's Domino Model on the Hypothesis of the Analysis.

### 2.6.1. Risk Mitigation Statistics

The accidentology data (see Table 3, Figure 4) reveals a nuanced landscape of occupational hazards across construction sites, underscoring critical areas for targeted risk mitigation. First Aid Cases (FAC) are the most frequent, with 117,430 events, accounting for 40.9% of all reported incidents. This high prevalence indicates that while these injuries are generally minor, there is a substantial volume of accidents that require immediate attention, suggesting potential gaps in everyday safety practices and minor hazard management (Shrestha 2020).

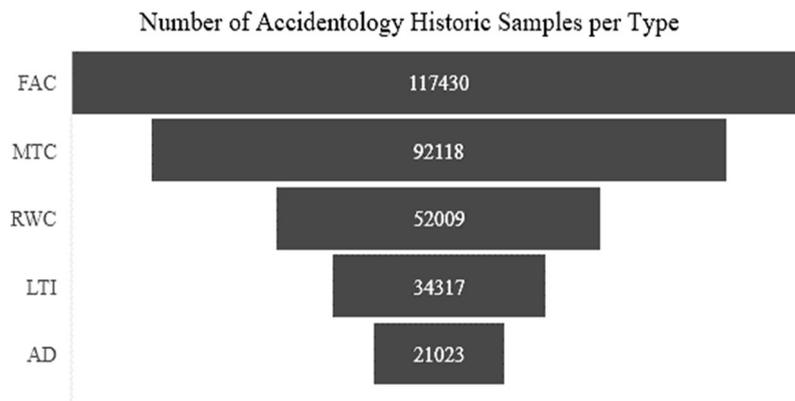
Medical Treatment Cases (MTC), with 92,118 events (32.1%), point to more serious incidents necessitating professional medical care. The frequency of MTCs signals significant underlying risks that warrant more robust safety interventions to prevent these injuries. Restricted Work Cases (RWC), totaling 52,009 events (18.1%), highlight injuries severe enough to limit workers' duties temporarily, emphasizing the need for preventive measures that can address the conditions leading to such incapacitating incidents.

Lost Time Injuries (LTI), with 34,317 events (12.0%), represent injuries that result in significant work absences, impacting both worker well-being and project timelines. This category's considerable proportion underscores the critical importance of enhancing safety protocols to reduce incidents with such profound effects on productivity. Asset Damages (AD), while the least frequent at 21,023 events (7.3%), reveal a troubling reality of hazards present on construction sites.

**Table 3.** Data Classification of Causation Per Injury Type per Man Hours.

<i>Injury Type</i>	<i>% of Data related to UA</i>	<i>% of Data related to UC</i>	<i>% of Data related to PF</i>	<i>% of Data related to EF</i>	<i>% of Data related to MA</i>	<i>% of Data related to PSA</i>
FAC	0.3521	0.2735	0.1234	0.0732	0.0987	0.0791
MTC	0.3056	0.2845	0.1467	0.1123	0.0874	0.0635
RWC	0.2245	0.2987	0.1342	0.1748	0.0981	0.0697
LTI	0.1678	0.2512	0.1543	0.2187	0.1426	0.0654

<i>Injury Type</i>	<i>% of Data related to UA</i>	<i>% of Data related to UC</i>	<i>% of Data related to PF</i>	<i>% of Data related to EF</i>	<i>% of Data related to MA</i>	<i>% of Data related to PSA</i>
AD	0.1034	0.2113	0.1321	0.2314	0.1987	0.1231

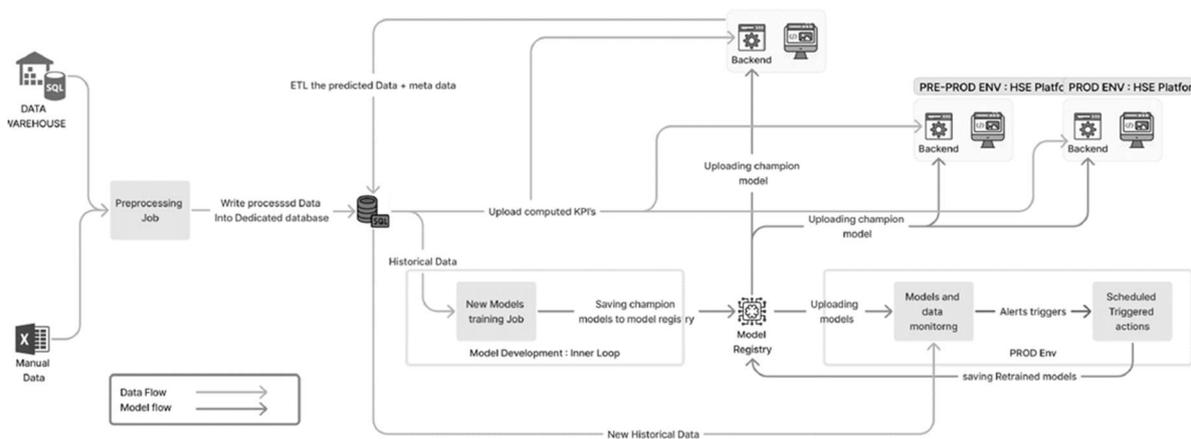


**Figure 4.** Overview of data quantity per type of injury.

Overall, this data highlights the need for a multi-faceted predictive approach to safety management, focusing on reducing the frequency and severity of all injury types.

### 2.6.2. Theoretical Approach

Our study capitalizes on a synthesis of methodologies employed in prior research, encompassing the conceptualization of the problem, data source selection, analytical techniques utilized, and the respective strengths and limitations of each approach in terms of scientific rigor and robustness following the framework annotated in Figure 5 and Figure 6:



**Figure 5.** Suggested architecture for the model deployment.

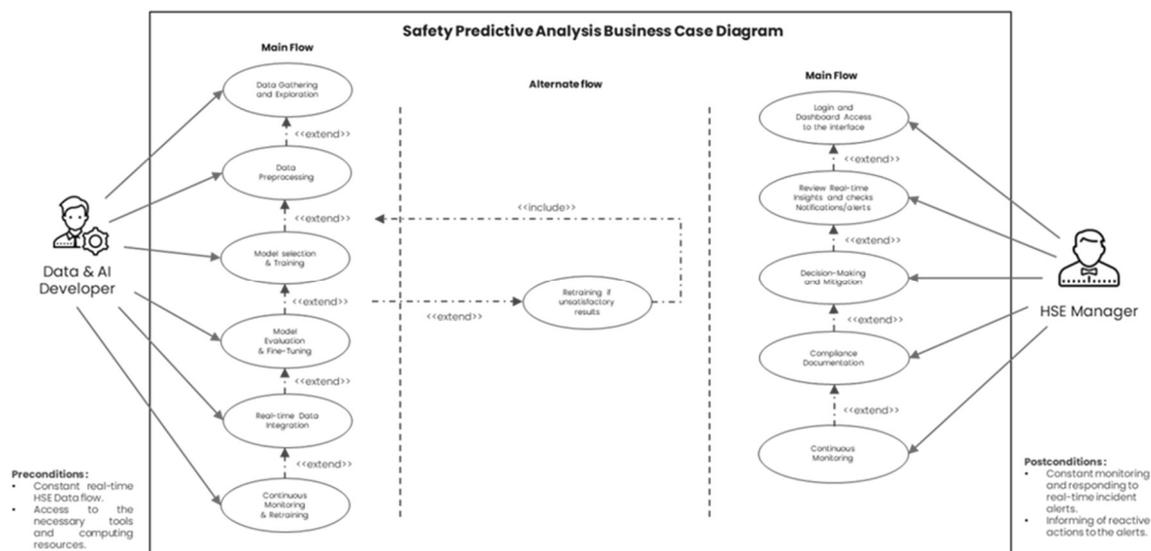


Figure 6. Safety Predictive Analysis Business Case Diagram.

In addition to presenting and organizing safety prediction research, we have also broken it down into three main families based on the data provided by the organization. Our work was at a pilot stage to evaluate the available techniques. The usefulness of data-mining approaches stems not only from the ability to process large amounts of data, but also from:

- Their ability to solve large-scale issues is essential when trying to determine important variables from a wide range of variables.
- Their ability to replicate the data generation process, no matter how complex, is due to their non-linear data structure (non-procedural approach).
- Their predictive and sometimes interpretative capabilities.

Many of the papers in the study were an attempt to consolidate the large body of literature on these topics, and then propose a single model for safety prediction that takes advantage of the unique characteristics of all three families of prediction, while also taking advantage of the overlap between the methodologies as an opportunity for cross-validation.

Significant differences exist between the predictive families, but the objectives are similar. Therefore, we propose that there is a potential for synergies and joint prediction by using multiple techniques at the same time.

Some safety prediction families measure very different aspects of a safety system. We argue that the combination provides synergy and predictive accuracy if the methods are independent. If the methods measure very similar aspects, we indicate that this is a case of cross-validation.

The data will be rigorously analyzed using various classification algorithms within a machine learning framework to achieve two main objectives: (1), to explore the correlation between the occurrence of incidents and different causation categories, and (2), to assess how direct, indirect, and root causes influence the prediction of injury types.

By applying classification algorithms such as decision trees and random forests, we aim to identify patterns and relationships within the data that reveal how different causation categories contribute to incidents. Additionally, we will evaluate the predictive impact of each causation type on injury outcomes to determine which factors are most influential. This comprehensive analysis will provide valuable insights for enhancing safety protocols and reducing injury rates by targeting the most significant causative factors.

### 3. Results

#### 3.1. Unified Model Approach

- **Safety Observation Reports (SOR) & Accidentology Historical:** Safety Observation Reports (SORs) and accidentology histories serve as situational methods in this context, aiming to forecast

safety outcomes for individual events based on specific environmental data. We hypothesize that combining these situational methods can yield predictions that are more reliable than those obtained using a single method in isolation. For instance, Random Forest and Decision Tree algorithms excel in predicting the type of injury and identifying its direct, indirect, and root causes. However, they do not inherently predict injury severity. In contrast, Safety Observation Reports demonstrate proficiency in forecasting injury severity and have proven effective in differentiating between successful and failed safety outcomes. Thus, while both methods focus on situational predictions, they are grounded in distinct aspects of the safety system.

- **Safety Key Performance Indicators Datasets:** Both safety prediction families, namely safety leading indicators, and the training dataset are time-dependent, as well as safety activities and operations, typically measured over weeks or months. Consequently, they are not suited for situational predictions but rather for forecasting injury rates over extended periods, spanning months to years. While training and safety leading indicators assess distinct facets of the safety system, they may not be entirely independent. Safety leading indicators gauge the efficacy of safety management, as documented by Hinze J, Hallowell M, & Baud K. (2013), whereas training evaluates overall safety perceptions. This encompasses perceptions regarding management's safety commitment, the role of supervisors, and the adequacy and effectiveness of training, encapsulated as "safety climate dimensions". These perceptions may be influenced by various factors, such as the quality and quantity of training programs, audit frequency, and incentive structures. Therefore, we postulate that integrating training data with safety-leading indicators could yield synergistic effects and opportunities for cross-validation.
- **Contractors Safety Performance Datasets:** This section aims to predict safety performance in construction sites using contractors' safety performance data. By analyzing the comprehensive dataset, including factors such as accident types, severity, and frequency, as well as contractor characteristics and historical safety records, examination of past safety performance, including incident rates, corrective actions taken, and adherence to safety regulations, to gauge the overall safety culture and performance trajectory. We employ various machine learning algorithms to identify patterns and predict potential safety incidents related to the accidentology historic dataset.

In short, these families of safety prediction can be carefully designed to cross-validate one another and to make more accurate and complete predictions of future performance.

To propose a higher-level systems model of the relationships among the families, we offer Figure 7. Here, we attempt to visually link the various methods of safety prediction and show what can be predicted from each method, how the methods can be used synergistically and opportunities for cross-validation. When two constructs measure similar attributes of safety, we consider these to be theorized relationships and opportunities for cross-validation, which are denoted by dashed lines in:

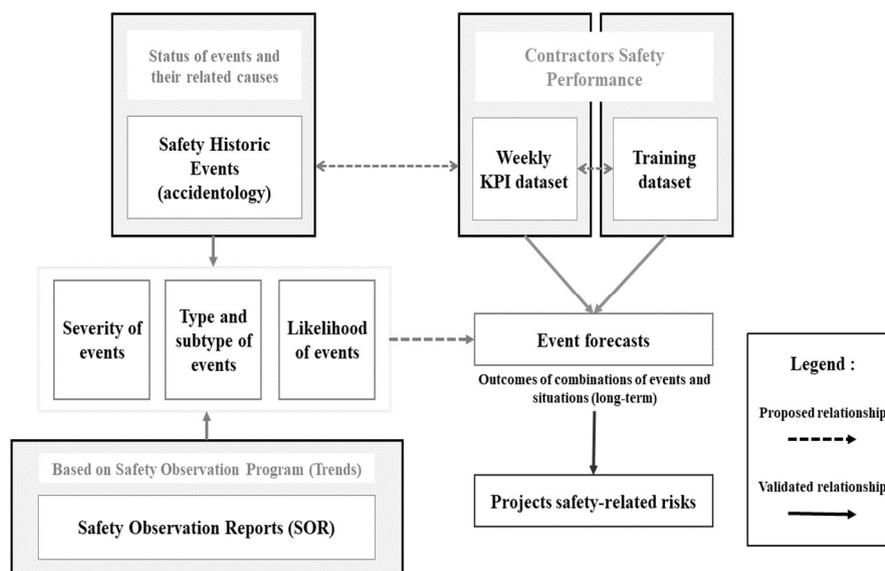


Figure 7. Suggested Model of the Relationships Among the Datasets Families.

### 3.2. Correlation of Causal Factors

A correlation matrix for the studied data groups relevant to multiple workplace sites is presented in Table 4. As illustrated in the table, all the studied variables exhibit a significant relationship with accident occurrence, underscoring the intricate interplay among the factors contributing to workplace incidents. The table also reveals a remarkable correlation between the variables themselves, highlighting their interconnected nature.

In addition to the correlation analysis, the mean and standard deviation for each variable are also calculated and presented, providing a comprehensive overview of their distribution. The statistical significance of these findings is indicated by a p-value of less than 0.03, further reinforcing the robustness of the observed relationships.

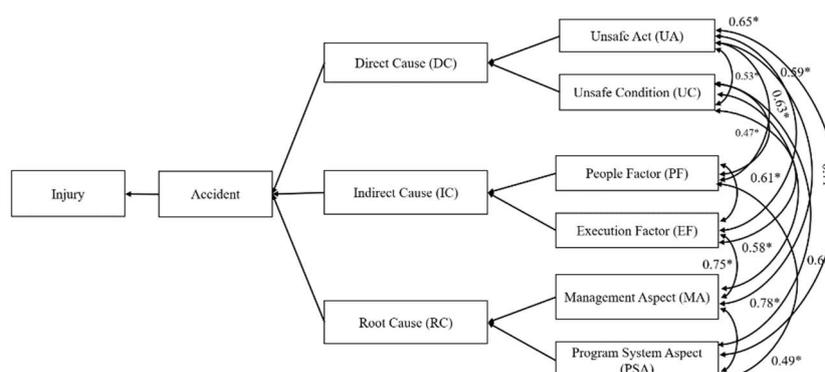
**Table 4.** Correlation Matrix of The Studied Parameters for the Studied Group.

<i>Injury Type</i>	1	2	3	4	5	6	<i>SD</i>	<i>Mean</i>
<b>UA</b>	-----	0.65*	0.71*	0.47*	0.53*	0.49*	0.13	0.56
<b>UC</b>	0.65*	-----	0.78*	0.60*	0.58*	0.61*	0.15	0.70
<b>PF</b>	0.71*	0.78*	-----	0.63*	0.59*	0.62*	0.13	0.72
<b>EF</b>	0.47*	0.60*	0.63*	-----	0.71*	0.68*	0.14	0.68
<b>MA</b>	0.53*	0.58*	0.59*	0.71*	-----	0.75*	0.14	0.69
<b>PSA</b>	0.49*	0.61*	0.62*	0.68*	0.75*	-----	0.14	0.69

(1) Unsafe Act, (2) Unsafe Condition, (3) People Factor, (4) Execution Factor, (5) Management Aspect, (6) Program System Aspect \* $p < 0.01$ . *SD*, standard deviation.

The correlation matrix presented in Table 4 reveals several key insights into the relationships between various workplace safety factors. Unsafe Acts (UA) exhibit strong positive correlations with People Factors (PF) (0.71) and Unsafe Conditions (UC) (0.65), indicating that individual behaviors and poor environmental conditions are closely linked to the occurrence of unsafe actions. Unsafe Conditions are also strongly associated with People Factors (0.78) and Execution Factors (EF) (0.60), suggesting that both individual factors and execution issues contribute significantly to unsafe conditions.

Execution Factors show a strong relationship with Management Aspects (MA) (0.71) and Program System Aspects (PSA) (0.68), highlighting the importance of effective management and robust systems in mitigating execution problems. Lastly, Management Aspects and Program System Aspects are highly correlated (0.75), underscoring the critical role of integrated management and system improvements in enhancing overall workplace safety. The correlations, with a significance level of  $p < 0.01$  (see Figure 8), indicate that addressing these interrelated factors could substantially improve safety outcomes.



**Figure 8.** Theoretical Model of Correlation Between Causal Factors of The Study.

### 3.3. Impact on Accident Occurrence

The correlation matrix in Table 5 reveals significant positive relationships between accident occurrence and all the causation factors studied. Unsafe Acts, Unsafe Conditions, and People Factors are particularly strongly associated with accidents, suggesting that both individual behaviors and environmental conditions play crucial roles in incident rates.

Execution Factors and Management Aspects also show meaningful correlations, highlighting the importance of effective execution processes and management practices in influencing accident occurrences. Additionally, Program System Aspects are closely related to accidents, underscoring the role of system deficiencies.

Overall, these findings highlight the complex interplay between various factors and emphasize the need for comprehensive improvements across all areas to enhance workplace safety and reduce accident rates.

**Table 5.** Correlation Matrix of The Causations & Accident Occurrence.

	0	1	2	3	4	5	6	SD	Mean
AO	-----	0.75*	0.70*	0.77*	0.65*	0.68*	0.74*	0.12	0.72
UA	0.75*	-----	0.65*	0.71*	0.47*	0.53*	0.49*	0.13	0.56
UC	0.70*	0.65*	-----	0.78*	0.60*	0.58*	0.61*	0.15	0.70
PF	0.77*	0.71*	0.78*	-----	0.63*	0.59*	0.62*	0.13	0.72
EF	0.65*	0.47*	0.60*	0.63*	-----	0.71*	0.68*	0.14	0.68
MA	0.68*	0.53*	0.58*	0.59*	0.71*	-----	0.75*	0.14	0.69
PSA	0.74*	0.49*	0.61*	0.62*	0.68*	0.75*	-----	0.14	0.69

(0) Accident Occurrence, (1) Unsafe Act, (2) Unsafe Condition, (3) People Factor, (4) Execution Factor, (5) Management Aspect, (6) Program System Aspect \*p < 0.01. SD, standard deviation.

### 3.4. Predictive Analysis of Accident Occurrence

Given the insights from the correlation matrixes, a predictive analysis of accident occurrence will be conducted using various machine learning algorithms. The strong correlations between accident occurrence and the identified causation factors provide a robust foundation for this analysis (Sarkar and & Maiti, Machine learning in occupational accident analysis: A review using science mapping approach with citation network analysis. 2020). By employing multiple machine learning algorithms, we aim to train and test models that can accurately predict the likelihood of accidents based on these factors (Choi, Gu and Chin 2020).

In this case study, several machine learning algorithms were utilized to leverage the correlations identified in the data. Random Forest was used for its ability to manage complex interactions between multiple factors (Zhang, et al. 2022) and reduce overfitting (see Figure 9).

```

Classification report:
precision    recall  f1-score   support

0           1.00     1.00     1.00         1
1           0.72     0.89     0.80        53
3           0.44     0.47     0.45        15
4           0.86     0.46     0.60        13
5           1.00     1.00     1.00        13
6           1.00     1.00     1.00        56
7           1.00     1.00     1.00        26
8           0.00     0.00     0.00         8

accuracy    0.84        185
macro avg   0.75     0.73     0.73        185
weighted avg 0.82     0.84     0.83        185

Summary of Accuracy Score:
Random Forest Model: 0.8432

```

**Figure 9.** Classification report of Random Forest Algorithm.

Decision Tree was also incorporated (Eetvelde, et al. 2021) to offer clear, interpretable insights into how different factors influence accident predictions (see Figure 10).

```

Classification report: -
precision    recall  f1-score   support

   0         1.00      1.00      1.00         1
   1         0.69      0.81      0.75        53
   3         0.33      0.47      0.39        15
   4         0.67      0.31      0.42        13
   5         1.00      1.00      1.00        13
   6         1.00      1.00      1.00        56
   7         1.00      1.00      1.00        26
   8         0.00      0.00      0.00         8

 accuracy          0.81      185
 macro avg         0.71      0.70      0.69      185
 weighted avg      0.79      0.81      0.79      185

Accuracy Score: 0.8108

```

**Figure 10.** Classification report of Decision Tree Algorithm.

AdaBoost was applied (see Figure 11) to enhance prediction accuracy by focusing on correcting misclassifications and boosting the performance of weaker models (Augustine and & Shukla 2022).

```

Classification report:
precision    recall  f1-score   support

   0         1.00      1.00      1.00         1
   1         0.60      1.00      0.75        53
   3         0.00      0.00      0.00        15
   4         0.00      0.00      0.00        13
   5         1.00      1.00      1.00        13
   6         0.68      1.00      0.81        56
   7         0.00      0.00      0.00        26
   8         0.00      0.00      0.00         8

 accuracy          0.66      185
 macro avg         0.41      0.50      0.44      185
 weighted avg      0.45      0.66      0.54      185

Summary of Accuracy Score:
AdaBoost Model: 0.6649

```

**Figure 11.** Classification report of AdaBoost Algorithm.

#### 4. Discussion

The ultimate objective of this study was to model the factors influencing workplace accidents, and the results reveal significant insights into how various causation factors are interrelated. Our findings indicate a substantial correlation between accident occurrence and a range of direct, indirect, and root causes within the case group. Notably, accident occurrence was inversely related to several variables, suggesting that an increase in these factors correlates with a decrease in accident rates, or vice versa, depending on the nature of the variable (Zarei, et al. 2021).

Our analysis highlights several critical pressure factors that impact workplace safety. Specifically, family-to-work conflict and occupational responsibilities were identified as significant sources of stress that adversely affect safety outcomes. The lack of effective management in balancing personal and professional demands exacerbates this issue, leading to increased accident risks. Furthermore, inadequate feedback and rewards from the work environment contribute to this pressure, affecting employees' overall safety performance.

Another key finding is the role of perceived control over one's behavior. Employees who feel that their behavior is heavily regulated by external controls rather than self-determined are under greater stress, which in turn increases the likelihood of workplace accidents. This perception of diminished personal agency contributes to a heightened sense of pressure, further elevating accident risks.

The study underscores the importance of general health on occupational safety. Individuals who are physically and psychologically healthier are better equipped to handle job-related stress, which can influence their ability to manage safety risks effectively. Improved health conditions are associated with greater resilience to stress, reducing the likelihood of accidents and mitigating their potential consequences.

Overall, these findings suggest that a multifaceted approach is necessary to address workplace safety. Effective management strategies, support systems for balancing work and personal life, and health-promoting interventions are crucial to reducing accident rates. By addressing these pressure factors and enhancing individual well-being, organizations can create a safer work environment and potentially lower the incidence of occupational accidents.

The training and testing results supported both hypotheses across all data groups, demonstrating a positive impact on occupational accident occurrences. These findings align with the research by (TD, et al. 2020) who also investigated factors influencing occupational accidents, including work-family conflict. Their study highlighted that work-family conflict, among other organizational parameters, showed a strong association with accident occurrences and severity.

As this is consistent with the present study, it confirms that several causations significantly affect accident rates. They represent a misalignment between personal and professional responsibilities, increasing the potential risk levels and, consequently, the likelihood of accidents. The results suggest that managing and mitigating these causations and related stressors can effectively reduce accident rates and enhance employee health. Thus, organizations should focus on addressing these conflicts to improve workplace safety.

## 5. Conclusions

The current analysis identified several safety climate variables—underlying in direct, indirect and root causes as significant risk factors for occupational accidents. To improve workplace safety, it is essential for management to involve workers in safety decisions and to enhance safety programs through targeted proactive action plans based on the predictive insights proven possible and accurate in this study results.

The study found that increased safety climate issues correlate with higher accident rates, particularly among workers in sectors construction. This underscores the need for targeted preventive and proactive measures in these high-risk areas.

Evaluating the predictive results during construction work could help identify individuals prone to accidents, allowing for better safety. The study's strength lies in the comparable sampling of several machine learning algorithms used in data that has been collected across multiple sites and circumstances, though future research should address the gap in comparing injured versus non-injured workers for more detailed insights into accident causation.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The data used for this research are not publicly available due to privacy and legal restrictions related to companies' ethics.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Abbasianjahromi, H., and M. & Aghakarimi. 2021. "Safety performance prediction and modification strategies for construction projects via machine learning techniques." *Engineering, Construction and Architectural Management*.
2. Alexander, D., Hallowell, M., and Gambatese, J. 2017. "Precursors of construction fatalities. II: predictive modeling and empirical validation." *Journal of construction engineering and management* 143(7).
3. APC, Chan, Guan J, Choi TNY, Yang Y, Wu G, and Lam E. 2023. "Improving Safety Performance of Construction Workers through Learning from Incidents." *Int J Environ Res Public Health* 5 (4570): 4-20.
4. Augustine, T., and S. & Shukla. 2022. "Road accident prediction using machine learning approaches." *2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*. 808–811.
5. Baker, Henrietta, Matthew R. Hallowell, and Antoine J.-P. Tixier. 2020. "AI-based prediction of independent construction safety outcomes from universal attributes." *Automation in Construction* 118.
6. Baradan, S., and Usmen, M. A. 2006. "Comparative injury and fatality risk analysis of building trades." *J. Constr. Eng. Manage.* 533-539.
7. Cavalcanti, M., L. Lessa, and B. & Vasconcelos. 2023. "Construction accident prevention: A systematic review of machine learning approaches.. ." *Work*.
8. Chen, J. R., and Yang, Y. T. 2004. "A predictive risk index for safety performance in process industries." *J. Loss Prev. Process Ind.* 17(3): 233-242.

9. Chen, W., and et al. 2020. "Artificial Intelligence Marvelous Approach for Occupational Health and Safety Applications in an Industrial Ventilation Field: A Short-systematic Review." *Electronics* 9 (10.3390/electronics9040591).
10. Choi, J., B. Gu, and S. et al. Chin. 2020. "Machine Learning Predictive Model Based on National Data for Fatal Accidents of Construction Workers." *Automation in Construction* (102974): 110.
11. Chua, D. K. H., and Goh, Y. M. 2005. "A Poisson model of construction incident occurrence." *J. Constr. Eng. Manage.* 715-722.
12. Cooper, M. D., and Phillips, R. A. 2004. "Exploratory analysis of the safety climate and safety behavior relationship." *J. Saf. Res.* 35(5): 497-512.
13. Eetvelde, H., L. Mendonça, C., Seil, R. Ley, and T. & Tischer. 2021. "Machine learning methods in sport injury prediction and prevention: a systematic review. ." *Journal of Experimental Orthopaedics*, 8 (10.1186).
14. Fang, D. P., Chen, Y., and Louisa, W. . 2006. "Safety climate in construction industry: A case study in Hong Kong." *J. Constr. Eng. Manage.* 573-584.
15. Fagnoli, Mario, and Mara Lombardi. 2020. "Building Information Modelling (BIM) to Enhance Occupational Safety in Construction Activities: Research Trends Emerging from One Decade of Studies." *Buildings* 10(6):98.
16. Gao, Yifan, Vicente Gonzalez, Kenneth Tak Wing Yiu, and Guillermo Cabrera-Guerrero. 2019. "The Use of Machine Learning and Big Five Personality Taxonomy to Predict Construction Workers' Safety Behaviour." *Computer Science*.
17. Gillen, M., Baltz, D., Gassel, M., Kirch, L., and Vaccaro, D. 2002. "Perceived safety climate, job demands, and coworker support among union and nonunion injured construction workers." *J. Saf. Res.* 33(1): 33-51.
18. Glendon, A. L., and Litherland, D. K. 2001. "Safety climate factors, group differences and safety behavior in road construction." *J. Saf. Sci.* 39(3): 157-188.
19. Hallowell, M. R., and Gambatese, J. A. 2009. "Activity-based safety and health risk quantification for formwork construction." *J. Constr. Eng. Manage.* 990-998.
20. Heinrich, H. W. 1941. *Industrial Accident Prevention: A Scientific Approach*. McGraw-Hill.
21. Hinze, J, Hallowell, M., and Baud, K. 2013. "Construction-safety best practices and relationships to safety performance." *J. Constr. Eng. Man.* (04013006): 1943-7862.
22. Johnson, S. E. 2007. "The predictive validity of safety climate." *J. Saf. Res.* 511-521.
23. Kakhki, Fatemeh Davoudi, Steven A. Freeman, and Gretchen A. Mosher. 2019. "Evaluating machine learning performance in predicting injury severity in agribusiness industries." *Safety Science* 117: 257-262.
24. Khan, Rafi Ullah, Jingbo Yin, Faluk Shair Mustafa, and Wenming Shi. 2023. "Factor assessment of hazardous cargo ship berthing accidents using an ordered logit regression model." *Ocean Engineering* 284 (115211).
25. Kim, Y., and S. & Chi. 2021. "Hazardous material releases in construction: Analysis with the decision tree approach. ." *Journal of Construction Engineering and Management* 2 (04020150): 147.
26. Koc, K., and A. & Gurgun. 2021. "MACHINE LEARNING APPLICATIONS IN CONSTRUCTION SAFETY LITERATURE." *Proceedings of International Structural Engineering and Construction*.
27. Lee, J., Y. Yoon, T. Oh, S. Park, and S. & Ryu. 2020. "A Study on Data Pre-Processing and Accident Prediction Modelling for Occupational Accident Analysis in the Construction Industry." *Journal of Safety Research* 73 (10.1016): 285-297.
28. Lee, S., and Halpin, D. W. 2003. "Predictive tool for estimating accident risk." *J. Constr. Eng. Manage.* 4(431): 431-436.
29. Mahamulkar, S, V H Lad, and K A Patel. 5-7 September 2022. "Development of a Framework for Selection of a Tunnel Lining Formwork System." *Proceedings 38th Annual ARCOM Conference*. Glasgow, UK: Association of Researchers in Construction Management. 359-368.
30. Rozenfeld, O., Sacks, R., Rosenfeld, Y., and Baum, H. 2010. "Construction Job Safety Analysis." *J. Saf. Sci.* 48(4): 491-498.
31. Sarkar, S., and J. & Maiti. 2020. "Machine learning in occupational accident analysis: A review using science mapping approach with citation network analysis. ." *Safety Science* (104900): 131.
32. Sarkar, S., and J. & Maiti. 2020. "Machine learning in occupational accident analysis: A review using science mapping approach with citation network analysis." *Safety Science* (104900): 131.
33. Shrestha, S. 2020. "Occupational Hazards in Building Construction." *SCITECH Nepal* (10.3126).
34. Shuang, Q., and Z. & Zhang. 2023. "Determining Critical Cause Combination of Fatality Accidents on Construction Sites with Machine Learning Techniques." *Buildings*.
35. Tam, C. M., and Fung, I. W. H. 1998. *Effectiveness of safety management strategies on safety performance in Hong Kong*. 16(1) vols. J. Construction Management Economy.
36. TD, Smith, Mullins-Jaime C, Dyal MA, and DeJoy DM. 2020. "Stress, burnout and diminished safety behaviors: An argument for Total Worker Health® approaches in the fire service." *J Safety Res.* (10.1016/j.jsr.2020.09.010): 75:189-195.

37. Yedla, Anurag, Fatemeh Davoudi Kakhki, and Ali Jannesari. 2020. "Predictive Modeling for Occupational Safety Outcomes and Days Away from Work Analysis in Mining Operations." *Int. J. Environ. Res. Public Health* 17(19).
38. Zarei, E., A. Karimi, E. Habibi, Barkhordari, and Reniers, G. A. 2021. "Dynamic occupational accidents modeling using dynamic hybrid Bayesian confirmatory factor analysis: An in-depth psychometrics study." *Safety Science* (105146): 131.
39. Zhang, Shuguang, Afaq Khattak, Caroline Mongina Matara, Arshad Hussain, and Asim Farooq. 2022. "Hybrid feature selection-based machine learning Classification system for the prediction of injury severity in single and multiple-vehicle accidents." *PLoS One* (10.1371).
40. Zhu, K., K. Du, and Y. & Tang. 2020. "Integrating machine learning with human factors for analyzing construction safety risk." *Automation in Construction* (103366): 119.
41. Zhu, R., X. Hu, J. Hou, and X. & Li. 2021. "Application of machine learning techniques for predicting the consequences of construction accidents in China. ." *Process Safety and Environmental Protection*.
42. Zohar, D. 1998. "Safety climate in industrial organizations: Theoretical and Applied Implications." *J. Appl. Psychol.* 78-85.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.