Unravelling Safety Concerns in Construction: A Comprehensive Data Analysis

Soroush Piri, MS, and Kamalesh Panthi, PhD
Morgan State University
Baltimore, Maryland

Construction site hazards and safety concerns have been a longstanding issue within the industry. These concerns extend beyond immediate physical risks to workers and public safety. Ensuring safety in construction is pivotal, as it not only safeguards lives but also has a significant societal impact. Previous studies have explored various methods and strategies to enhance construction site safety. However, there remains a gap in understanding the intricate interplay between contributing factors, personal attributes, and accident occurrence. This research aims to fill this gap by employing advanced data analysis techniques, including text mining, time series, comparative, correlation analyses. In this regard, we present a novel approach to address construction site safety. We begin by manually collecting data from the National Institute for Occupational Safety and Health (NIOSH) accident reports and subsequently employ an automatic algorithm for data cleaning and extraction. The data we collect includes information on Gender, Age, Contributing factors, Cause of Death, Recommendations by experts, State, and Summary from these reports. Through these advanced data analysis techniques, we aim to identify the most effective strategies for reducing the occurrence of accidents in the construction industry, thereby improving overall safety performance.

Keywords: Construction site safety, Comparative analysis, Correlation analysis, Text mining, Time-series analysis

Introduction

Safety within the construction industry is of paramount importance due to the inherent risks, including falls, electrical hazards, and machinery-related accidents (Malekitabar et al., 2016). These risks not only jeopardize the well-being of workers but also pose a substantial threat to public safety. Recent statistics reveal the urgent need for innovative safety solutions, as the construction sector accounted for 46.2 percent of all fatal incidents related to falls, slips, and trips in 2021 (U.S. Burea of Labor Statistics, 2021). Globally, approximately 2.3 million individuals succumb to work-related accidents or diseases each year, underscoring the magnitude of the issue (International Labour Organization, 2011). Recent data also indicates a concerning trend, with a rising frequency of accidents and work-related illnesses (Liang et al., 2020). From a historical perspective, the industry had minimal safety regulations, resulting in high injury and fatality rates (Abudayyeh et al., 2006; Teizer et al., 2010). The 1970 Occupational Safety and Health Act (OSHA) in the United States marked a turning point by

introducing comprehensive safety standards and enforcement (Blanc & Escobar Pereira, 2020). Its influence extended globally, as other countries adopted similar regulations.

As part of the ongoing efforts to enhance construction safety, recent research by Chi and Han sheds light on the intricacies of construction accidents (Chi & Han, 2013). Their study involved empirical and statistical analyses of 9,358 accidents in the U.S. construction industry from 2002 to 2011. The research aimed to investigate relationships between accidents and injury elements, such as injury type, part of the body injured, and injury severity. However, despite these efforts, preventing work-related injuries, especially falls, remains a significant challenge among U.S. workers. In this regard, results from recent studies, such as the one conducted by (Socias-Morales et al., 2018) indicate that work-related falls continue to be a problem, especially among certain worker groups.

While numerous research studies focus on quantitative data available on the OSHA website, there is a limited amount of research utilizing the vast qualitative data available, such as those present in fatality investigation reports compiled by the National Institute for Occupational Safety and Health (NIOSH) through the Fatality Assessment and Control Evaluation (FACE) program. Since the inception of NIOSH-FACE in 1982, investigators have conducted traumatic occupational fatality investigations throughout the United States and provided technical assistance to 15 state health or labor departments with cooperative agreements with NIOSH to conduct traumatic fatality surveillance, targeted investigations, and prevention activities at the state level (Higgins, 2001). Currently, there are 263 NIOSH FACE construction accident reports and 619 State FACE construction accident reports (NIOSH ,n.d.).

Data mining technology can help us examine correlations between labor safety inspections and legal regulations. This approach holds promise for constructing a predictive model to identify potential safety hazards in construction. The effectiveness of this model increases as more construction accident data accumulates over many years. In recent times, there has been a growing interest in utilizing data to enhance our understanding of safety incidents, particularly with the increasing availability of digitally recorded incident reports. Baker et al employed Natural Language Processing (NLP), Convolutional Neural Networks (CNN), and Hierarchical Attention Networks (HAN), along with the established Term Frequency-Inverse Document Frequency representation (TF-IDF) + Support Vector Machine (SVM) approach. This methodology aimed to identify valid injury precursors and predict accident outcomes (Baker et al., 2020b). The same researchers further advanced their work, achieving more accurate predictions of injury severity outcomes(Baker et al., 2020a). (Tixier et al., 2017) proposed and used a Data Mining technique to extract valuable insights from accident reports without relying on subjective expert opinion. In a previous study, (Tixier et al., 2016) applied natural language processing (NLP) to quickly and automatically scan unstructured injury reports, replacing the conventional manual content analysis. Notably, researchers like Xu et al. (2021) have taken a proactive approach to addressing this issue. Their research leverages text-mining technology to automatically extract critical safety risk factors from accident reports. By streamlining the analysis process and considering report length, their study contributes to enhancing safety in the construction industry. Key outcomes include the development of a text-mining framework, a domain-specific lexicon, and an innovative method for assessing term importance. Huseh et al. (2013) used data mining technology to explore labor safety strategy in the construction industry. Our research aims to identify effective strategies to reduce accidents in the construction industry by employing advanced data analysis techniques such as text mining and time series analysis, while focusing on both safety-related and personal factors. This effort contributes to the overall goal of improving safety performance in the construction industry.

Methodology

Data Collection

In this comprehensive methodology for analyzing occupational accidents using NIOSH reports, the data collection process begins with manually extracting historical accident reports from the National Institute for Occupational Safety and Health (NIOSH) database. This manual extraction involves the meticulous selection of report numbers through hyperlinks one by one, ensuring a detailed and accurate compilation of incident data. This hands-on approach is imperative to guarantee the precision and reliability of the dataset, as it allows for a nuanced examination of each individual report, considering unique circumstances and contextual factors that may influence the incident analysis. The manual extraction process not only involves the retrieval of numerical data but also extends to capturing qualitative information, such as narrative descriptions, root causes, and recommended preventive measures. This comprehensive data collection strategy aims to provide a holistic understanding of each incident, facilitating a nuanced analysis that goes beyond numerical frequencies.Furthermore, this meticulous manual extraction process allows researchers to account for any potential variations in reporting formats, terminology, or data presentation across different incident reports. By actively engaging in the extraction process, researchers gain firsthand insights into the intricacies of each incident, enabling them to identify nuances that may be overlooked in automated extraction methods.

While the manual extraction may be time-intensive, it serves as a crucial step in ensuring the integrity of the dataset and the robustness of subsequent analyses. The detailed scrutiny of each incident report becomes an integral part of the research methodology, contributing to the depth and accuracy of the findings. The methodology utilizes RStudio (*RStudio: Integrated Development Environment for R*, n.d.) and the 'rvest' (Wickham, 2023a) to automatically extract valuable information from the NIOSH database. R (R Core Team, 2021) is a programming language and environment for statistical computing and graphics, offering various statistical and graphical tools, including linear and nonlinear modeling, statistical tests, time-series analysis, classification, and clustering. RStudio provides a user-friendly interface and a comprehensive set of tools, making it easier for data scientists, statisticians, and analysts to work with R. Utilizing RStudio, we develop a custom algorithm that leverages the 'rvest' package, specifically designed for web scraping and parsing HTML content. This package enables the extraction of valuable information from websites, a crucial step in collecting historical accident reports from the NIOSH database. The combined use of RStudio and 'rvest' automates the data collection process and efficiently compiles the reports. By developing a custom algorithm that follows hyperlink report numbers, we can efficiently compile historical accident reports, laying the groundwork for our analysis. This approach ensures the creation of a comprehensive and up-to-date dataset, contributing to the improvement of workplace safety and the prevention of accidents.

Data Cleaning

Following data collection, a crucial step in this methodology involves rigorous data cleaning and pre-processing, which encompasses the systematic elimination of errors, handling missing values, removing duplicates, and resolving inconsistencies within the collected data. This meticulous process is essential for ensuring the utmost data quality and reliability, making it suitable for in-depth

analysis. To achieve these data processing tasks, we utilized the 'dplyr' (Wickham et al., 2023) and 'stringr' (Wickham, 2023b) packages within the R programming environment.

Text Mining Algorithm

To further enrich the analysis, a robust text mining algorithm is developed and implemented, which is adept at extracting crucial information from NIOSH accident reports. This algorithm is tailored to extract specific details such as contributing factors, expert recommendations, state information, and comprehensive accident summaries, age, gender, and the date of incident.

Variable Categorization

In the methodology, a critical step involves the definition of key variables for the analysis, including, but not limited to, gender, age, contributing factors, recommendations by experts, state, summary from accident reports. These variables are carefully selected to encompass a comprehensive range of factors affecting workplace safety. Categorizing variables helped us to organize our data into meaningful categories, making it easier to work with and interpret. It also helped to simplify complex datasets by grouping similar variables together. This reduced the number of variables we need to consider, making the analysis more manageable. We categorized the values of contributing factors to 10 categories, which were "Human Error", "Lack of Training", "Unsafe Work Practices", "Equipment Failures", "Environmental Conditions", "Communication Issues", "Procedural Failures", "Lack of Safety Culture", "Fatigue", "Inadequate Safety Equipment". Then we categorized the values of recommendations to 7 categories as well, which are "Hazard Identification and Control", "Safety Management and Training", "Contractor and Employer Responsibilities", "Equipment and Technology Enhancement", "Conduct a site assessment", "Emergency Response Planning".

Time Series Analysis

Time Series Analyses enabled a comprehensive understanding of how the frequency of various contributing factors has evolved over time. This involves the creation of time series plots, which visually depict the temporal changes in contributing factors related to workplace safety. The time series analyses results were visually represented through time series plots, enabled easy interpretation of patterns and trends. To conduct these time series analyses, we leveraged the 'lubridate' (Grolemund & Wickham, 2011) package in the R programming environment.

Association Network Diagram

This analysis aimed to harness the power of association network diagrams to unravel the intricate web of relationships between contributing factors and recommended safety measures in the context of workplace safety and incident prevention. The resulting association network diagram serves as a dynamic visual representation, offering numerous advantages in understanding and leveraging the insights derived from the data. We used 'igraph' (Csárdi et al., 2023) package, and the GEM (Graph Embedding Methods) (Goyal & Ferrara, 2018) layout optimization algorithm, which is suitable for large and complex networks. It was helpful because we had a graph with many nodes and edges.

Keywords Frequencies Analysis

This analysis calculates the frequencies of each contributing factor to determine which factors are most common. This analysis also helped to identify the frequencies of recommendations. We utilized the 'ggplot2' (Wickham, 2016) package, which provided insightful plot for research. By determining which contributing factors are most common, construction industry stakeholders can identify the most pressing safety concerns. This allows them to allocate resources, training, and preventive measures more effectively to address the factors that pose the highest risks.

Results and Discussions

In our comprehensive analysis of incident reports sourced from the National Institute for Occupational Safety and Health (NIOSH), we conducted a keyword frequency analysis as shown in Figure 1 to discern the prevailing contributing factors of safety incidents. Our findings reveal a hierarchy of contributing factors ranked by frequency as follows: Unsafe Work Practices, Environmental Conditions, Procedural Failures, Human Error, Communication Issues, Lack of Training, Equipment Failures, Inadequate Safety Equipment, Fatigue, and Lack of Safety Culture. Remarkably, Unsafe Work Practices and Environmental Conditions emerged as the most frequently cited factors, suggesting their prominent roles in safety incidents. This analysis not only underscores the imperative need to address these primary factors but also hints at potential interrelationships and patterns among them.
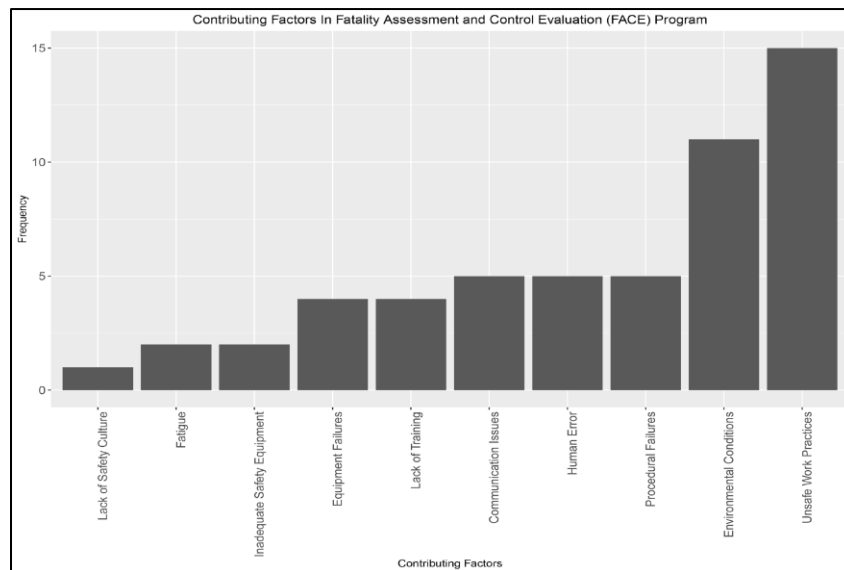

Figure 1. Frequency of contributing factors

Figure 2 shows the frequency of recommendations made in the FACE reports. Our analysis revealed common recommendations for incident prevention, emphasizing *Safety management and training*, *Hazard identification and control*, *Equipment and Technology Enhancement, Contractor and employer responsibilities, Conduct site assessments, and Emergency response planning*. The focus on Hazard Identification and Control underscores the importance of proactive risk assessment and

mitigation. The inclusion of *Equipment and Technology Enhancement* in the recommendations reflects the evolving landscape of safety practices. Advancements in technology can greatly enhance safety measures, from improved protective equipment to real-time monitoring and data analysis, demonstrating the need for organizations to stay current with safety technologies. *Contractor and Employer Responsibilities* indicate a shared commitment to safety. This recommendation highlights that safety is not the sole responsibility of one party but requires a collaborative effort. Clear definitions of responsibilities ensure accountability across the board. *Regular Site Assessments* are crucial for ongoing safety improvement. They underscore the importance of continually evaluating the workplace for potential hazards and making necessary adjustments. Finally, *Emergency Response Planning* highlights the necessity of being prepared for unforeseen incidents. It's a reminder that while prevention is crucial, having a well-thought-out response plan can mitigate the impact of incidents that do occur.
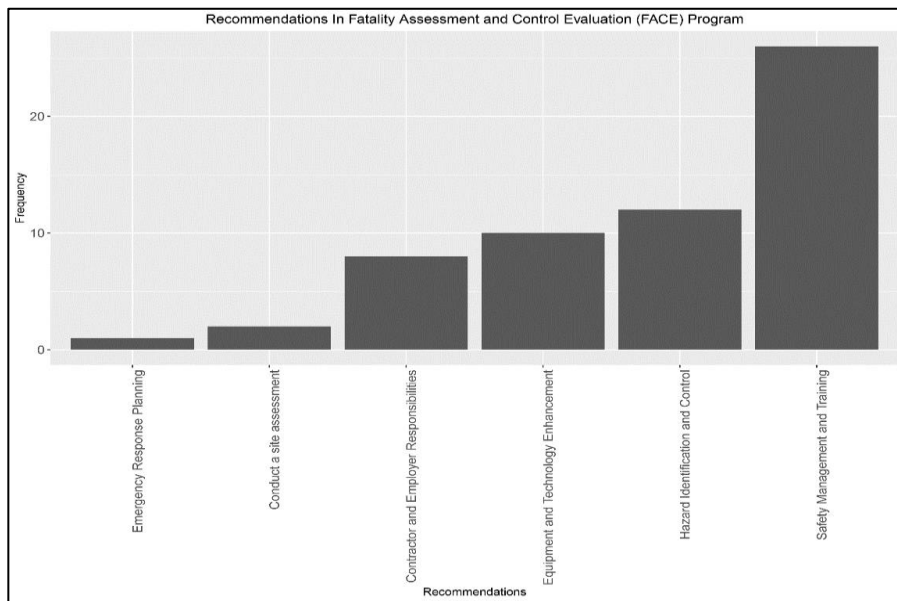

Figure 2. Frequency of Recommendations

The application of Time Series Analysis as shown in Figure 3 to the incident reports dataset spanning from 2006 to 2014 has yielded valuable insights into the temporal evolution of contributing factors to workplace incidents. Notably, our analysis revealed a marked increase in the frequency of "Unsafe Work Practices", "Communication Issues", and" Human Error" as contributing factors. This suggests that over the specified time frame, there has been a noticeable uptick in incidents related to unsafe work practices. It may indicate a need for enhanced training, stricter enforcement of safety protocols, or other measures to address and mitigate this particular issue. Similarly, the analysis highlights a marked increase in the frequency of "Communication Issues" as contributing factors. This trend suggests a growing concern in workplace communication, reflecting potential challenges in information flow, clarity, or process effectiveness. Addressing it may involve communication training, protocol enhancements, or fostering a culture that encourages open and effective dialogue among team members. In the same manner, the analysis highlights a marked increase in the frequency of 'Human Error' as contributing factors. This trend suggests potential challenges in human performance and decision-making within the workplace. The elevated occurrence of 'Human Error'

may point to areas where additional training, process improvements, or error-prevention measures are needed. Addressing this trend could involve implementing targeted training programs, enhancing processes to reduce the likelihood of human error, or adopting technologies that provide additional safeguards. In contrast, a declining trend was observed for "Environmental Conditions", and "Procedural Failures", which suggests several positive developments. These include improved environmental safety practices and enhanced procedural adherence within the workplace. Such positive trends may signify the successful implementation of safety measures, increased employee awareness, and a commitment to continuous improvement in environmental and procedural aspects. Organizations may benefit from further reinforcing these positive practices through ongoing training, regular safety audits, and fostering a safety-conscious culture among employees.
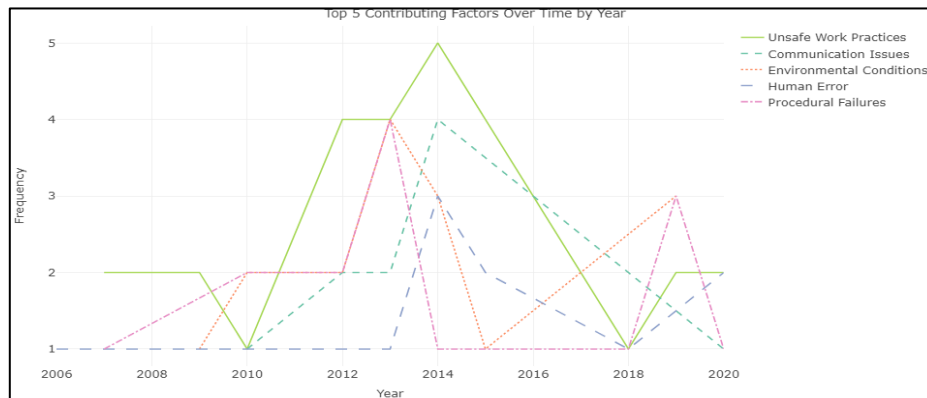

Figure 3. Top 5 contributing factors-time series analysis

The association network diagram as shown in Figure 4, derived from the associations of contributing factors and recommended safety measures, unveils valuable insights for enhancing workplace safety and incident prevention. Notably, "Environmental Conditions" demonstrates a strong association with safety management and training, equipment and technology enhancement, and the necessity to conduct site assessments. Conversely, "Equipment Failures" is closely linked to hazard identification and control, emphasizing the importance of recognizing and rectifying equipment-related risks and reinforcing contractor and employer responsibilities. "Fatigue" emerges as a critical factor linked to safety management and training, hazard identification and control, and the need to develop, implement, and train on emergency response plans, highlighting the need for strategies to address worker exhaustion. "Human Error" is intricately tied to various recommendations, reinforcing the multifaceted approach required to tackle this contributing factor effectively. "Inadequate Safety Equipment" emphasizes the need for safety management and training, equipment and technology enhancement, site assessments, and contractor and employer responsibilities. "Lack of Safety Culture" is clearly associated with a wide array of safety recommendations, indicating the necessity for holistic cultural shifts within organizations. "Lack of Training" echoes the significance of safety management and training, hazard identification and control, equipment and technology enhancement, site assessments, and contractor and employer responsibilities, emphasizing the need for comprehensive training programs. Lastly, "Procedural Failures" stresses the importance of procedural adherence, particularly in the context of contractor and employer responsibilities, equipment and technology enhancement, and hazard identification and control. This network diagram provides a structured framework for organizations to tailor their safety strategies. The network diagram underscores the dynamic and evolving nature of safety practices within the construction industry, highlighting the need for continuous improvement and adaptability to address specific safety concerns effectively.
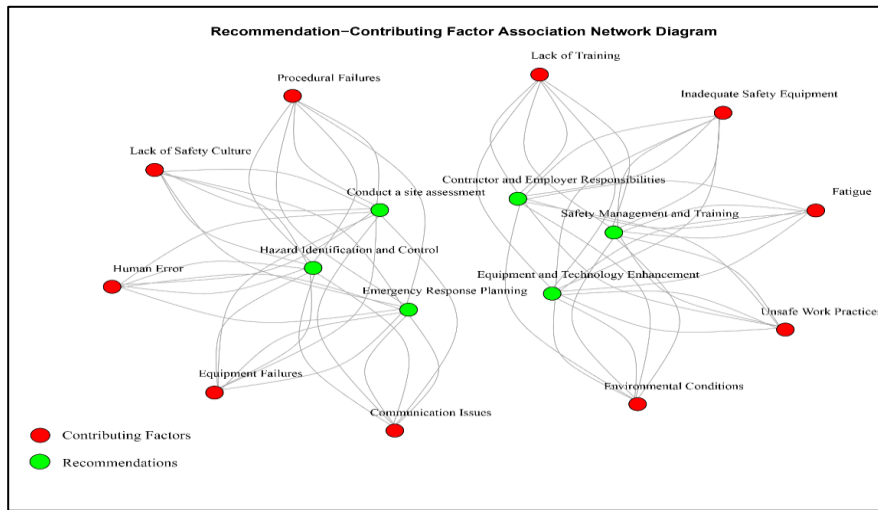
918

Figure 4. Recommendation-contributing factor association network diagram

Conclusion and Recommendations

In conclusion, our analysis of NIOSH incident reports has highlighted key contributing factors and common safety recommendations. "Unsafe Work Practices" and "Environmental Conditions" emerged as central issues, emphasizing their critical roles in incidents. Temporal analysis revealed evolving patterns, notably the rise of "Communication Issues", "Unsafe Work Practices"," Human Error" and a decrease in "Environmental Conditions", and "Procedural Failures". The association network diagram demonstrated interrelationships between contributing factors and recommendations. These findings not only underscore the need to address these primary factors but also hint at potential interrelationships and patterns among them. To translate these findings into practical action, we recommend the following: Enhanced Training Programs-organizations should invest in more comprehensive training programs, particularly focusing on addressing "Unsafe Work Practices." These programs can target specific areas where incidents are prevalent. Embracing modern technology and equipment to enhance safety measures is also a viable solution. This includes real-time monitoring, data analysis, and advanced protective gear to reduce the impact of incidents. We suggest further research into the potential connections between the rise in safety recommendations and the increased reporting of incidents related to "Unsafe Work Practices" and "Environmental Conditions."

References

Abudayyeh, O., Fredericks, T. K., Butt, S. E., & Shaar, A. (2006). An investigation of management's commitment to construction safety. *International Journal of Project Management*, *24*(2), 167–174.

Baker, H., Hallowell, M. R., & Tixier, A. J.-P. (2020a). AI-based prediction of independent construction safety outcomes from universal attributes. *Automation in Construction*, *118*, 103146.

Baker, H., Hallowell, M. R., & Tixier, A. J.-P. (2020b). Automatically learning construction injury precursors from text. *Automation in Construction*, *118*, 103145.

Blanc, F., & Escobar Pereira, M. M. (2020). Risks, Circumstances and Regulation – Historical development, diversity of structures and practices in Occupational Safety and Health inspections. *Safety Science*, *130*, 104850.

Chi, S., & Han, S. (2013). Analyses of systems theory for construction accident prevention with specific reference to OSHA accident reports. *International Journal of Project Management*, *31*(7), 1027–1041.

Csárdi, Nepusz, T., Traag, V., Szabolcs Horvát, Fabio Zanini, Daniel Noom, & Kirill Müller. (2023). *igraph: Network Analysis and Visualization in R*.

Goyal, P., & Ferrara, E. (2018). Graph embedding techniques, applications, and performance: A survey. *Knowledge-Based Systems*, *151*, 78–94.

Grolemund, G., & Wickham, H. (2011). *Dates and Times Made Easy with lubridate*. *40*, 1--25.

Higgins, D. N. (2001). The Fatality Assessment and Control Evaluation program's role in the prevention of occupational fatalities. *Injury Prevention*, *7*(90001), 27i–233.

Hsueh, S.-L., Huang, C.-F., & Tseng, C.-Y. (2013). *Using Data Mining Technology to Explore Labor Safety Strategy- A Lesson from The Construction Industry*.

International Labour Organization. (2011, July 13). *World Statistic*. http://www.ilo.org/moscow/areas-of-work/occupational-safety-and-health/WCMS_249278/lang--en/index.htm

Liang, H., Zhang, S., & Su, Y. (2020). The structure and emerging trends of construction safety management research: A bibliometric review. *International Journal of Occupational Safety and Ergonomics*, *26*(3), 469–488.

Malekitabar, H., Ardeshir, A., Sebt, M. H., & Stouffs, R. (2016). Construction safety risk drivers: A BIM approach. *Safety Science*, *82*, 445–455.

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. https://www.R-project.org/

NIOSH. NIOSH Fatality Assessment and Control Evaluation Program Brochure (Accessed November 10, 2023 from: https://www.cdc.gov/niosh/face/inhouse.html

RStudio Team. (2020). *RStudio: Integrated Development Environment for R* [Computer software]. RStudio, PBC. http://www.rstudio.com/

Socias-Morales, C. M., Chaumont Menéndez, C. K., & Marsh, S. M. (2018). Fatal work-related falls in the United States, 2003-2014. *American Journal of Industrial Medicine*, *61*(3), 204–215.

Teizer, J., Allread, B. S., Fullerton, C. E., & Hinze, J. (2010). Autonomous pro-active real-time construction worker and equipment operator proximity safety alert system. *Automation in Construction*, *19*(5), 630–640.

Tixier, A. J.-P., Hallowell, M. R., Rajagopalan, B., & Bowman, D. (2016). Application of machine learning to construction injury prediction. *Automation in Construction*, *69*, 102–114.

Tixier, A. J.-P., Hallowell, M. R., Rajagopalan, B., & Bowman, D. (2017). Construction Safety Clash Detection: Identifying Safety Incompatibilies among Fundamental Attributes using Data Mining. *Automation in Construction*, *74*, 39–54.

U.S. Burea of Labor Statistics. (2021). *Construction deaths due to falls, slips, and trips increased 5.9 percent in 2021: The Economics Daily: U.S. Bureau of Labor Statistics*.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. https://ggplot2.tidyverse.orgWickham, H. (2023a). *Easily Harvest (Scrape) Web Pages*. https://rvest.tidyverse.org

Wickham, H. (2023b). *stringr: Simple, Consistent Wrappers for Common String Operations*. https://stringr.tidyverse.org

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *dplyr: A Grammar of Data Manipulation*. https://dplyr.tidyverse.org

Xu, N., Ma, L., Liu, Q., Wang, L., & Deng, Y. (2021). An improved text mining approach to extract safety risk factors from construction accident reports. *Safety Science*, *138*, 105216.