JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Integrative Analysis of Safety Risks in Construction Projects: Merging Cascading Effects and Machine Learning Methodologies"

S.P.SANKAR [*] U VINITH[1]
[*] - Corresponding author: Associate Professor, Department of Mechanical Engineering,
[1]- ME Industrial Safety Engineering, KSR College of Engineering (Autonomous), Tiruchengode, Tamil Nadu, India

Abstract : Addressing safety incidents remains a persistent challenge in construction, requiring constant attention from project management teams. Despite prioritizing accident prevention, the dynamic nature of projects and varying management expertise contribute to ongoing safety mishaps, disrupting schedules and inflating costs, often with tragic consequences. To manage this, it's crucial to identify and reduce risk factors throughout the project lifecycle, considering their complex interplay. A nuanced understanding is required, navigating uncertainty to anticipate and address potential risks at every stage. Integrating insights from academia and experience, leveraging advanced methodologies and technologies, is becoming increasingly important for effective safety management. Traditionally, safety risk analysis has been limited, but a novel approach combining cascading effects and machine learning offers a more thorough and objective understanding. This approach recognizes the ripple effects of incidents and employs machine learning to extract insights from vast data sets, enabling a more accurate depiction of safety landscapes and informing better risk mitigation strategies. Overall, this combined approach represents a significant advancement, promising safer environments and more successful project outcomes through enhanced risk identification, assessment, and mitigation.

Keywords: Risk factors Heterogeneous data Machine learning Cascading effect

1.Introduction

Safety incidents pose a persistent challenge within the construction industry, demanding unwavering attention from project management teams. These teams consistently prioritize accident prevention as a core responsibility (Nadhim et al., 2016). However, the dynamic landscape of construction projects and varying levels of management expertise contribute to the ongoing occurrence of safety mishaps. These incidents not only disrupt project schedules and inflate costs but also tragically result in casualties, highlighting the urgent need for bolstered risk mitigation strategies (Zhang et al., 2013; Xia et al., 2018).

The unpredictable nature of risk factors complicates the effective management of safety incidents. Addressing this challenge requires the identification and reduction of risk factors, as well as the mitigation of their adverse effects, to diminish the likelihood and severity of accidents. Wang and Yuan (2017) emphasize that risk factors are pervasive throughout the entire lifecycle of a project, intertwining in complex ways. Moreover, accidents often stem from the interplay of multiple interconnected risk factors, each with its own distinct underlying causes (Marcelino-Sadaba et al., 2014; Fang and Wu, 2013).

Given this intricate web of influences, managing project safety demands a nuanced understanding of the multifaceted relationships among various risk factors. Project managers must navigate through uncertainty, proactively identifying and addressing potential sources of risk at every stage of project development. This comprehensive approach not only involves minimizing the occurrence of safety incidents but also entails mitigating their potential impact when they do occur.

As such, the construction industry is increasingly focusing on interdisciplinary approaches that integrate insights from both academia and practical experience. By leveraging advanced methodologies and technologies, project management teams can enhance their ability to anticipate, prevent, and effectively respond to safety incidents, thereby fostering safer working environments and more successful project outcomes.

The comprehensive analysis of safety risk factors in construction projects represents a critical endeavor aimed at minimizing accidents and ensuring the well-being of workers. Traditionally, such analyses have been limited by subjectivity and the inability to fully harness the wealth of data available. However, a novel approach combining the concepts of cascading effects and machine learning has emerged, promising a more thorough and objective understanding of safety risks in construction projects.

The combined cascading effect and machine learning approach integrates two distinct methodologies to provide a holistic analysis of safety risk factors.

Firstly, the cascading effect principle recognizes that safety incidents in construction projects often have ripple effects, where the consequences of one risk factor can propagate and exacerbate the impact of others. By incorporating this concept into the analysis, the interconnected nature of risk factors is acknowledged, allowing for a more nuanced assessment of their collective influence on project safety.

Secondly, machine learning techniques are employed to sift through vast amounts of heterogeneous data and extract meaningful insights regarding safety risk factors. These advanced algorithms are capable of identifying patterns, trends, and correlations within the data that may not be immediately apparent to human analysts. By leveraging machine learning, the analysis can more efficiently uncover hidden risk factors and prioritize them based on their potential impact on project safety.

The combined approach offers several advantages over traditional methods. By considering the cascading effects of risk factors, the analysis accounts for the complex interactions between different elements of a construction project, thereby providing a more accurate depiction of the overall safety landscape. Additionally, the use of machine learning enhances the objectivity and efficiency of the analysis, enabling project managers to make more informed decisions regarding risk mitigation strategies.

Overall, the combined cascading effect and machine learning approach represents a significant advancement in the field of construction project safety. By integrating these complementary methodologies, project teams can better identify, assess, and mitigate safety risks, ultimately fostering safer work environments and more successful project outcomes.

2.Background

2.1Understanding the Imperative for Integrative Safety Risk Analysis in Construction Projects

Understanding the Imperative for Integrative Safety Risk Analysis in Construction Projects" encapsulates the urgent need for a more holistic approach to safety risk management within the construction industry. It emphasizes the importance of gaining insight into why adopting an integrative safety risk analysis approach is essential, suggesting that traditional methods may not adequately address the intricate challenges of safety risk management in construction projects.

The phrase "For Integrative Safety Risk Analysis" specifies the nature of the analysis being discussed – one that integrates various factors or methodologies. This indicates a departure from isolated approaches towards a more comprehensive understanding of safety risks. By incorporating both qualitative and quantitative assessments, this integrative approach aims to offer a nuanced and thorough perspective on safety risks in construction projects.

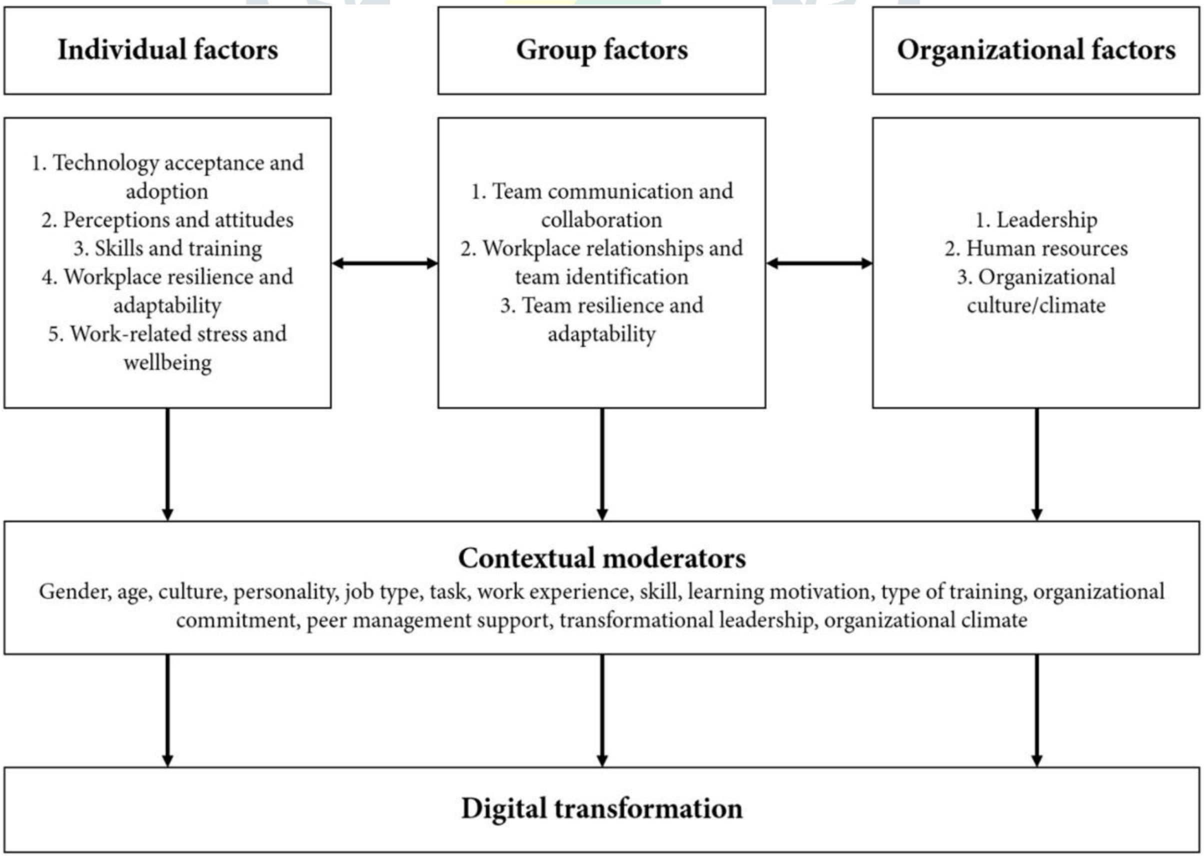
Moreover, by including "In Construction Projects," the title focuses specifically on the context of construction. It acknowledges that safety risk analysis in construction projects presents unique challenges and dynamics compared to other industries. Factors such as the ever-changing work environment, involvement of multiple stakeholders, and presence of various hazards necessitate tailored approaches to safety risk management. Therefore, the title underscores the relevance and applicability of the proposed integrative approach within the construction sector.

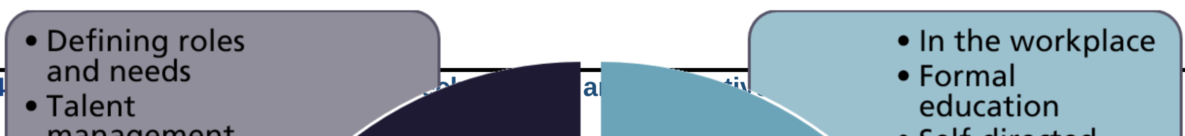
Overall, the title effectively communicates the importance of acknowledging and addressing the complexities of safety risk management in construction projects through an integrative approach. It lays the groundwork for further exploration into the rationale behind this imperative and the potential benefits it offers for improving safety outcomes in construction. By adopting such an approach, researchers and practitioners can gain a deeper understanding of safety risks in construction projects and develop strategies to effectively mitigate them, ultimately fostering safer work environments and enhancing project outcomes.

2.2 Exploring the Integration of Machine Learning in Risk Assessment.

Machine learning (ML) stands out as a promising tool for analyzing uncertain risk factors, benefiting from its capabilities in handling diverse data types. Current research in this area can be broadly categorized into two main aspects. Firstly, ML techniques are applied to unstructured text data, including safety accident reports and meeting records, to uncover potential risk factors through data mining. For instance, Brown (2016) utilized the Latent Dirichlet Allocation (LDA) algorithm to analyze accident text and identify factors contributing to severe accident impacts. Similarly, Rajasekaran et al. (2008) combined various ML methods such as Naive Bayes (NB) and Support Vector Machine (SVM) to classify injury events effectively based on large datasets from labor bureau records. Advanced ML models like Random Forests (RF) and Stochastic Gradient Tree Boosting (SGTB) are also employed to estimate the likelihood of risk factors occurrence and predict risks using extensive accident text data. Moreover, Natural Language Processing (NLP) techniques are utilized to encode and extract risk factors from unstructured text data.

On the other hand, ML techniques are also applied to numerical data, including accident characteristics and time-series data. Researchers, such as Zou et al. (2016), have utilized association rule mining to identify relationships between fall accidents and contributing factors, shedding light on the influence of different variables. To assess the relative importance of risk factors, Kwon et al. (2015) employed Naive Bayesian classifiers and decision trees. Some studies primarily focus on risk prediction; for example, Rajasekaran et al. (2008) utilized SVM to predict surge events caused by ocean storms, accident probabilities, and potential impact severity. Additionally, artificial neural networks (ANN) and convolutional neural networks (CNN) have been utilized to evaluate risk factors.





Framework for Analyzing Safety Risk Factors in Projects

2.3 Understanding the Spread and Ripple Effects of Risk.

The phenomenon of cascading effects, often observed in the realm of risk transfer, represents a critical yet sometimes overlooked aspect of project management. It manifests when seemingly minor oversights or errors escalate into a chain of events with profound and sometimes catastrophic consequences. These consequences may include significant casualties, financial losses, or disruptions to project timelines.

Cascading effects highlight the interconnectedness of various elements within a project system. Actions taken in one part of the project can propagate throughout the entire system, triggering a cascade of unforeseen events. This amplification of risk underscores the need for vigilant risk management practices to mitigate potential consequences.

Project failures frequently trace back to fundamental issues such as flawed structural designs or a culture of negligence within organizations. These underlying weaknesses serve as fertile ground for the initiation and propagation of cascading effects. For instance, a project's structural deficiencies may create vulnerabilities that, when left unaddressed, can escalate into larger issues affecting multiple aspects of the project.

Moreover, the presence of an interconnected network within the project environment exacerbates the propagation of faults. Information, decisions, and actions flow through this network, facilitating the transmission of risks and their associated impacts. As a result, a single fault or oversight can have far-reaching implications, affecting not only the immediate project tasks but also broader organizational objectives.

Recognizing the cascade effect necessitates understanding the mechanisms through which risks propagate within a project. By identifying potential points of failure and implementing proactive measures to address them, project managers can minimize the likelihood and severity of cascading effects. This may involve enhancing structural resilience, fostering a culture of accountability, and implementing robust risk management strategies tailored to the project's unique circumstances.

In essence, the cascade effect serves as a stark reminder of the interconnected nature of project risks and the importance of proactive risk management practices in safeguarding project success. Only through a comprehensive understanding of these dynamics and diligent risk mitigation efforts can projects effectively navigate the complexities of risk transfer and minimize the impact of cascading effects.

The construction project schedule operates as a fundamental activity network comprising activities and nodes. Liu et al. (2013) conducted an in-depth analysis into how safety accidents reverberate throughout project timelines and costs. Beyond merely examining the direct consequences of accidents on schedule delays, they also scrutinized the resulting resource consumption, thereby offering valuable insights into the broader implications of risk factors on project dynamics.

While casualties directly result from risk factors, their ramifications extend beyond mere disruptions to project schedules. Risk factors can profoundly impact various facets of project management, including cost, progress, and quality (Flyvbjerg, 2013). Consequently, schedule delays frequently manifest as downstream consequences of these risk factors.

When an accident disrupts a task within the project, it not only stalls that particular activity but also holds the potential to impede subsequent tasks in the project timeline. This delay propagation through the activity network can trigger a cascade effect, where the repercussions of the initial incident spread across interconnected activities. Understanding and acknowledging this cascade effect marks a significant breakthrough in project management research, shedding light on the intricate mechanisms of risk transmission within construction projects.

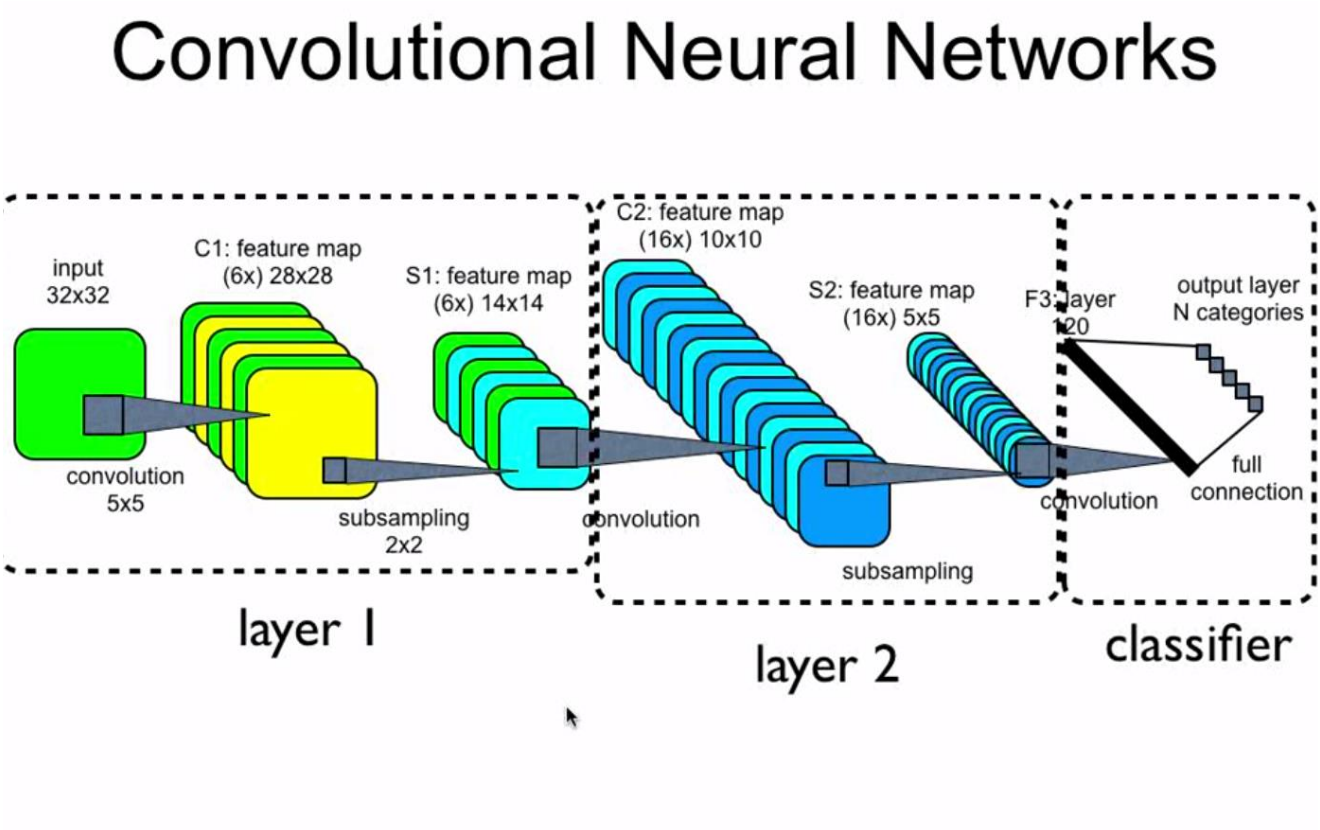
By recognizing the interconnected nature of project activities and the potential for cascading effects, project managers can proactively mitigate risks and implement strategies to minimize the impact of accidents on project schedules and costs. This enhanced understanding of risk propagation dynamics contributes to more effective risk management practices, ultimately leading to improved project outcomes and enhanced project resilience.

3. Methodology

This study employs a comprehensive analysis to identify and evaluate project safety risk factors, integrating existing knowledge of project safety management. The framework consists of three main components: 1) Identifying project safety risk factors from various unstructured texts generated during project management; 2) Defining risk factors based on key "features" and "outcomes"; and 3) Establishing an Activity-on-Node (AON) network to assess cascading effects and measure the impact of risk factors on projects. Figure 1 illustrates our proposed Integrated Analysis Framework (IAF) alongside a comparison with previous studies. Detailed discussions on each component follow in subsequent sections.

Given the nature of project safety management, the IAF not only aids in identifying risks during the construction stage but also in predicting risks throughout subsequent construction phases. Safety incident reports from the construction stage contain valuable hidden risk information. Extracting safety risk factors from these reports enables the formulation of proactive measures against similar hidden dangers in future project phases.

Additionally, risk information may also be present in meeting minutes, construction logs, and other documents, which the IAF can help uncover. As the framework only requires one-time establishment, it holds universal applicability across construction projects, enabling managers to analyze risk factors with the assistance of safety-related documents throughout the entire construction process. Numerous unstructured data are generated in projects



3.1. Identification and Prioritization of Risk Factors.

3.1.1 Categorization of Textual Documents.

Project teams frequently adopt diverse strategies to mitigate personnel safety risks, tailoring these measures to address the specific causes of accidents. This study extends its focus to analyze safety incidents attributed to various causes, aiming to pinpoint safety risk factors associated with each cause. The initial step involves categorizing textual documents according to the identified accident causes and assigning specific document types as Ti. This categorization facilitates the extraction of safety risk factors corresponding to each accident cause, enabling a targeted approach to risk management.

To accomplish this task, the study leverages the CNN model, renowned for its adeptness in handling intricate structures and diverse components. Additionally, the CNN model demonstrates effectiveness in processing extensive heterogeneous data with cross-features, making it well-suited for the multifaceted nature of textual documents (Li et al., 2019). Given the inherent variability and complexity of textual data, rigorous data cleaning procedures are imperative before analysis can commence. Thus, CNN is introduced as the primary tool for classifying a wide range of materials and meeting records.

By employing the CNN model for document classification, the study aims to streamline the process of identifying safety risk factors associated with different accident causes. This approach not only enhances the efficiency of risk assessment but also enables project teams to tailor control measures more effectively to address specific safety concerns. Ultimately, by understanding and mitigating safety risks at their root causes, project teams can foster a safer working environment and minimize the occurrence of accidents throughout the project lifecycle.

The Convolutional Neural Network (CNN) is structured with multiple layers, each playing a distinct role in processing data. These layers include the input layer, convolutional computing layer, excitation layer, pooling layer, connection layer, and output layer, as visualized in Figure 2.

To ensure the effectiveness of the CNN model, a robust training set comprising 150 representative safety accident cases was created. These cases were carefully selected from safety accident reports issued by the Occupational Safety & Health Administration (OSHA), encompassing a wide range of incidents in different types of buildings. The training set, denoted as R*, was utilized to train the CNN model, leveraging TensorFlow 1.14. Additionally, a separate test set, labeled as R, was used to evaluate the model's performance.

Simultaneously, a text classification framework was developed, as depicted in Figure 2. This framework integrates various components, including vector-matrix transformation, convolution, activation function, maximum pooling layer, and a multi-layer perceptron classifier. The operation of the framework is delineated into three primary processes

Word Vector Transformation: This process involves converting textual data into numerical representations, such as word embeddings or vectors, to facilitate processing by the CNN model.

Extraction of Text Features: The framework extracts salient features from the transformed word vectors, capturing relevant information to characterize the input text.

Automatic Classification: Leveraging the extracted text features, the framework employs a multi-layer perceptron classifier to automatically classify the input text into predefined categories or labels.

By segmenting the operation into these three distinct processes, the text classification framework efficiently analyzes textual data and categorizes it based on predefined criteria. This approach enhances the model's ability to classify safety-related documents accurately, thereby contributing to improved risk assessment and management practices within the context of project safety.

In the context of utilizing a Convolutional Neural Network (CNN) for text processing, the initial challenge lies in preparing the textual data for input into the model. Since the CNN model cannot directly process vocabulary, a transformation into a vector-matrix format is necessary, enabling computers to recognize and process the data efficiently. Taroun (2014) highlighted the significance of the word vector space, where words sharing similar meanings exhibit close associations, despite potential differences in properties.

To bridge this gap between textual data and CNN processing, the word2vec technique is employed to convert input words into vectors. This approach facilitates the representation of words in a continuous vector space, capturing semantic similarities among them. However, it's essential to acknowledge that context information within text documents pertaining to project safety incidents might be incomplete, and text features could be ambiguous.

To address this challenge, the skip-gram model is adopted for text processing. This model is particularly adept at handling sparse and incomplete textual data by predicting surrounding words given a target word. By leveraging the skip-gram model, the CNN framework can effectively capture the contextual nuances and semantic relationships embedded within the textual data, despite potential information gaps.

In summary, the process involves transforming textual data into a format suitable for CNN processing, leveraging word2vec for word embedding and the skip-gram model for contextual understanding. This approach ensures that the CNN model can effectively interpret and analyze textual data related to project safety incidents, even in scenarios where context information may be incomplete or ambiguous

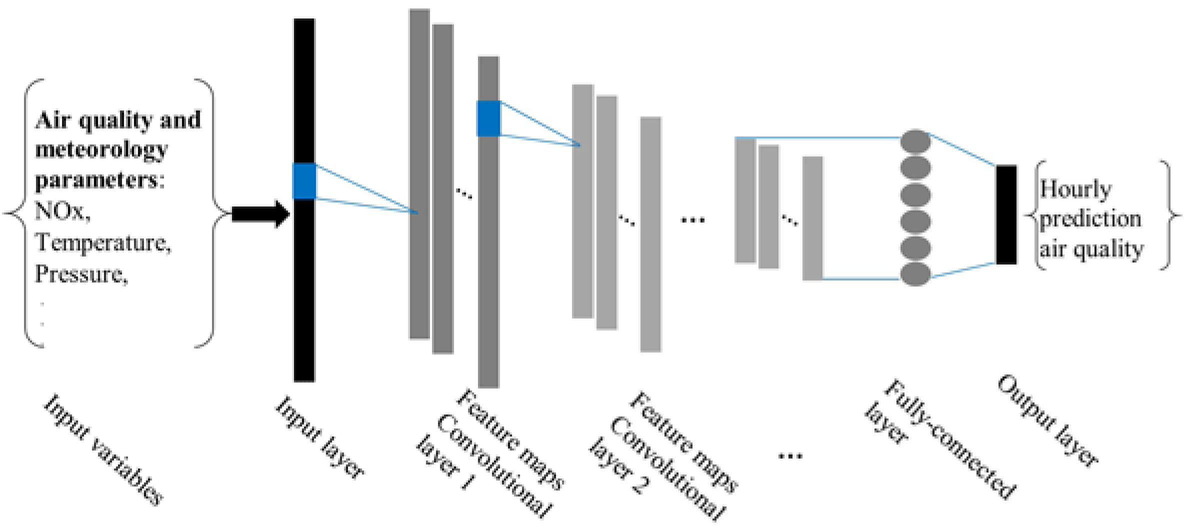


Fig3.CNN model

Word Embedding: This stage focuses on converting vocabulary information into a numerical word vector space. Utilizing the word2vec tool, vocabulary information is mapped to a semantic space within the training set, resulting in the generation of a word vector model. For instance, considering the word "collapse," its related words in the semantic space include "platform" with a similarity score of [0.97393211233], "injure" with a score of [0.969927314301], "kill" with a score of [0.96290013827], and "concrete" with a score of [0.95327622081]. This process enables the representation of words in a continuous vector space, capturing semantic relationships and contextual associations among them.

(0.01324947 0.01233419 − 0.01283749 − 0.02473493 0.03432512 0.01864652 0.03228544 0.01927345 − 0.0327201 0.09328726 − 0.01774350 − 0.02065491 − 0.2290923 0.0532193 − 0.04624433 0.11873217 0.1320874 ...

Text Classification Using the CNN Model: The CNN model emerges as a preferred choice for text classification tasks due to its efficiency and effectiveness. Renowned for its adaptability in handling diverse document types, the CNN model offers a robust solution for multi-type document classification. This study employs pip to establish a Python virtual environment, named "virtual," and installs TensorFlow 1.14 within this environment. Pool layer configurations and classification features are customized based on the requirements of the CNN model and the characteristics of the text data. The results of the final testing phase are depicted in Figure 3.

Evaluation of Results: In assessing the effectiveness of the model, Precision, Recall, and F1 scores are employed. These metrics provide valuable insights into the model's performance in accurately classifying text labels. The test results, as illustrated in Figure 3, reveal the accuracy of the text label "collapse" to be 0.990099.

This high accuracy underscores the effectiveness of the CNN model in achieving accurate and reliable text classification results**.**

3.1.2 Refining Safety Risk Factor Extraction for Enhanced Hazard Identification.

After classifying text documents, we employ the Apriorism algorithm to extract potential safety risk factors, as outlined by Anand and Dinakaran in 2017. Utilizing the dataset R, the Apriorism algorithm iteratively searches through R to identify frequent item sets. Upon reaching reasonable levels of support and confidence, connection and pruning strategies are applied to construct a candidate set, denoted as Candidate. For instance, consider the safety accident type "Electrocution" (T1=Electrocution), which may involve risk factors such as "circuit aging," "lack of electrical certification," and "charge overload." These factors collectively constitute an itemset, where itemset = {item1, item2, ..., item}. The calculation formula for this process can be expressed as follows:

The Apriorism algorithm functions by iteratively scanning through the dataset to identify frequent item sets, i.e., combinations of items that appear together with high frequency. Initially, it identifies individual items that surpass a predefined support threshold, indicating their frequent occurrence. Then, it gradually extends these item sets by adding more items, ensuring that each generated itemset still meets the support threshold. This iterative process continues until no further extensions are possible, resulting in a set of frequent item sets.

Once the frequent item sets are determined, the algorithm employs association rules to evaluate the relationship between different items within these sets. These association rules, typically measured by metrics such as confidence and lift, help identify significant associations between items. For instance, a high confidence value indicates a strong association between items, implying that the occurrence of one item is highly likely to be accompanied by the occurrence of another.

Moreover, the Apriorism algorithm employs pruning strategies to optimize the search process by eliminating subsets of infrequent item sets that are guaranteed to be infrequent themselves. This pruning step reduces the computational overhead associated with generating and evaluating candidate item sets, thereby improving the efficiency of the mining process.

In summary, the Apriorism algorithm is a fundamental technique for mining frequent item sets and extracting association rules from large datasets, enabling the identification of potential safety risk factors from text documents efficiently.

$$\mathrm{Support}(x, y, z) = P(xyz)\frac{\mathrm{num}(xyz)}{2\mathrm{num}(\mathrm{Allsamples})}$$

$$\mathrm{Confidence}(x \Leftarrow yz) = P(x|yz)\ \frac{P(xyz)}{2aP(yz)}$$

At this stage, Support serves as a metric indicating the likelihood of occurrence for specific risk factor associations. It quantifies the frequency of these associations within the dataset R by measuring the proportion of times they appear. For instance, if a particular association between two risk factors occurs frequently across various documents in R, it would receive a higher support value, suggesting a stronger association.

Confidence, on the other hand, offers insights into the conditional probability of one risk factor occurring given the presence of another. It reflects the likelihood of the occurrence of a secondary factor following the occurrence of a primary factor. In essence, confidence evaluates the strength of association between two risk factors by calculating the conditional probability of their co-occurrence.

To illustrate, consider a scenario where "circuit aging" is a primary risk factor, and "electrocution" is a secondary risk factor. A high confidence value between these factors implies that when "circuit aging" is identified in a safety assessment, there's a high likelihood of "electrocution" being present as well. This indicates a strong correlation between the two factors and underscores the importance of addressing both in safety protocols.

Confidence provides valuable insights for risk assessment and mitigation strategies. By understanding the likelihood of one risk factor occurring in conjunction with another, safety measures can be tailored more effectively. For instance, if the confidence between "no electrical certificate" and "electrocution" is found to be high, safety protocols can prioritize ensuring proper certification to mitigate the risk of electrocution incidents.

In summary, while Support quantifies the frequency of risk factor associations, Confidence delves deeper into the conditional probability of these associations, offering nuanced insights into their interrelationships. Together, these metrics empower safety professionals to identify, prioritize, and address potential risk factors more comprehensively, thereby enhancing overall safety protocols and minimizing hazards.

A strong correlation is established when Support values are high, indicating frequent co-occurrences of specific risk factors. In essence, higher Support values suggest that certain risk factor associations occur consistently across the dataset, signaling their relevance and importance. Conversely, low Support values indicate infrequent occurrences of these associations, suggesting that they are less significant and may not warrant immediate attention in safety assessments.

Confidence, serving as another pivotal parameter, complements Support by assessing the reliability of mining results. It measures the conditional probability of one risk factor occurring given the presence of another, thus gauging the strength of association between them. When Confidence values are elevated, it signifies a robust correlation between associated risk factors. In practical terms, this implies that the likelihood of one risk factor occurring alongside another is high, reinforcing the need to address both factors in safety protocols.

Moreover, an increase in Confidence values corresponds to a heightened incidence of related factors, indicating a more stable correlation between them. This suggests that as Confidence levels rise, the reliability of identified risk factor associations strengthens, providing clearer insights into potential hazards and their interrelationships.

Effectively, the interplay between Support and Confidence enables safety professionals to discern meaningful correlations from spurious associations within the dataset. While high Support values highlight frequent and significant risk factor associations deserving attention, elevated Confidence values affirm the reliability and stability of these associations, guiding prioritization efforts in safety planning and mitigation strategies.

Therefore, by considering both Support and Confidence metrics, safety practitioners can make informed decisions regarding risk assessment and mitigation, focusing resources on addressing the most pertinent and reliable risk factor associations, thereby enhancing overall safety outcomes.

Drawing on the principles of the Apriorism algorithm, we have devised a comprehensive algorithmic approach tailored for mining safety risk factors embedded within text documents. Our algorithm unfolds in a step-by-step fashion, strategically sifting through the dataset to unveil pertinent risk factors.

Initially, in the inaugural phase of the algorithm, each item representing a potential risk factor (e.g., "circuit aging," "no electrical certificate," "charge overload," etc.) undergoes a tallying process to accumulate their respective counts. Subsequently, items failing to meet the predefined Support threshold, such as "no electrical certificate," are promptly pruned from further consideration. The surviving items, satisfying the minimum Support criterion (e.g., "circuit aging," "charge overload," etc.), compose what we term as frequent 1_itemsets, denoted as s1.

Moving forward, the algorithm embarks on a successive search within the dataset, wherein frequent 1_itemsets are interconnected to form candidate frequent 2_itemsets. Again, items failing to meet the Support threshold are systematically pruned from consideration, yielding the emergence of frequent 2_itemsets, exemplified by occurrences of "circuit aging," among others.

This iterative process persists until no further frequent (k + 1)_item sets can be derived, signifying the culmination of the algorithm's search. The resultant ensemble of frequent itemset, exemplified by occurrences of "circuit aging," constitutes the algorithm's output, encapsulating identified safety risk factors.

This meticulous algorithmic process is meticulously delineated in Table 1, providing a comprehensive roadmap for navigating the intricate terrain of safety risk factor extraction from textual data. By adhering to this systematic methodology, safety professionals can efficiently distill actionable insights from voluminous textual sources, empowering them to bolster safety protocols and preemptively address potential hazards.

The holistic extraction process of safety risk factors encompasses two pivotal stages: data preprocessing and element extraction. Particularly within the domain of construction engineering, safety accident data is predominantly articulated in human language, often characterized by convoluted logic and structural complexities. To enable effective computerized analysis, this data necessitates meticulous preprocessing.

The preprocessing phase serves as a crucial preparatory step aimed at enhancing the readability and interpretability of safety accident data for computational systems. It involves a series of methodical procedures tailored to streamline the data and render it amenable to automated analysis.

The primary objective of preprocessing is to rectify the inherent intricacies present within the textual data, thus facilitating coherent interpretation by computational algorithms. This involves tasks such as noise reduction, where irrelevant or redundant information is eliminated to enhance signal clarity. Additionally, textual data may undergo standardization processes to ensure uniformity and consistency in representation, simplifying subsequent analysis tasks.

Furthermore, preprocessing encompasses the vital task of linguistic parsing, wherein the syntactic and semantic structures of textual data are deciphered and disentangled. This involves segmenting the text into discernible units, such as sentences or paragraphs, and identifying the underlying grammatical and semantic elements. By decomposing the text into its constituent parts, computational systems can more effectively discern patterns and extract meaningful insights.

Moreover, data preprocessing often involves the application of techniques such as tokenization, stemming, and lemmatization, aimed at reducing lexical variations and standardizing textual representations. This fosters a more cohesive and homogeneous dataset, conducive to robust analysis and inference.

In essence, data preprocessing serves as the foundational bedrock upon which subsequent analytical tasks rely. By meticulously preparing and refining safety accident data, computational systems can navigate through the inherent complexities of human language and extract actionable insights pertaining to safety risk factors in construction engineering.

Table 2

Importance ranking algorithm of risk factors.

Input: Original data set Si , k* ∈ Si ,i = 1, 2, ..., 11

Output: List of the importance of risk factors

Initial: According to the 11 safety accident causes, the data sets are constructed respectively, and then the extracted safety risk factors are put into the data set. For example, the possible safety factors of "electric shock" accidents may include "circuit aging", "electric charge overload" and "employment without a certificate", etc.

1) Determine the optimal parameters

Building a random forest model, mtyr = 1, 2, ..., n, then when errOOB → Optimal solution and stability, then obtain optimal mtry and ntree

2) Calculating MDA and MDC

The optimal parameters are used to optimize the random forest algorithm model, and the results of MDA and MDC are calculated

XMDA : The value of MDA

XMDC : The value of MDC

3) Comprehensive calculation of importance

when $XMDA \rightarrow xMDA = \frac{xMDA - xmin}{xmax - xmin} \times 100$, $XMDC \rightarrow xMDC = \frac{xMDC - xmin}{xmax - xmin} \times 100$, then

Si X = XMDA + XMDC

Sequencing, end

3.1.3 Prioritization of Risk Factors.

Before delving into the intricate web of risk propagation, it's crucial to establish a foundational understanding of safety risk factors' significance. Enter Random Forest (RF), a powerful tool adept at tackling nonlinear modeling challenges sans the need for data normalization, as demonstrated by Zhou et al. in 2019. RF not only navigates through varied data dimensions with ease but also effectively gauges the importance of factor features.

RF operates by employing a random and repeated self-sampling technique to extract N groups of datasets from the original data. Each dataset comprises two-thirds of the total original data. Subsequently, utilizing these N datasets, RF selects an optimal number of feature nodes M (where M is less than mtry) to construct decision trees (ntree). The growth of decision trees primarily hinges on mtry and ntree parameters, with the optimal solution attained through errOOB, an unbiased estimation derived from the remaining one-third dataset.

Central to the RF algorithm's efficacy is the acquisition of optimal parameter solutions for mtry and ntree. To comprehensively evaluate the importance of safety factors, RF leverages Mean Decrease Accuracy (MDA) and Mean Decrease Gini (MDC). MDA quantifies the contribution of individual factors to error rate reduction.

Dynamic Framework of Risk Propagation.

3.2Dynamic Framework of Risk Propagation

In the realm of construction processes, the interconnectedness among various tasks amplifies the repercussions of any failure within the system. When one process falters, it inevitably ripples through the project's activity network, affecting the quality (Qi) of downstream tasks (Lee et al., 2009). Despite the limitations of using time deviation as a sole indicator of task completion quality, the paramount influence of risk factors on project scheduling necessitates its continued relevance.

This study introduces the concept of cascading effects, offering a more objective and rigorous approach to assessing risk factors' impact. Instead of merely considering local impacts within a safety incident, it evaluates the broader repercussions across the entire project. The formula for Qi, though simplistic, encapsulates the essence of this evaluation.

Project planning serves as the project's pulse, delineated through the activity-on-the-node (AON) network notation. In this directed graph, tasks flow sequentially, with each downstream task dependent on the completion of its predecessors. Consequently, any failure cascades downstream, either halting progress or necessitating delays.

Quantifying the actual impact of risk factors hinges upon assessing risk levels, categorized into three grades of importance: "extremely important," "relatively important," and "important." The actual time spent on a task is adjusted accordingly, reflecting the importance of safety risk factors.

Tasks transition between states of "affected" and "unaffected" based on a predefined threshold. A failure in one task triggers a cascading effect, affecting downstream tasks based on their susceptibility to the upstream failure. This cascading effect is determined by two key mechanisms: the infecting power of the upstream task and the spreading power of the downstream task.

The infecting power of a task is influenced by its topological and temporal features within the AON network. Likewise, the spreading power considers both topological and temporal characteristics, defining the task's ability to influence downstream tasks. Together, these mechanisms dictate the degree of cascading failure within the project, with the intensity varying based on the importance of the implicated risk factors."

Serial	Project information		Project schedule		AON network	
	Project function	Types of safety accidents	Expected duration (days)	Actual duration (days)	Node count	Edge count, *E*
P1	Teaching building	A5	105	122	197	816
P2	Library	A3, A8	306	397	258	1182
P3	Residential building	A4, A5	285	311	304	828

Table 4 summary of project information

Hence, we introduce the notion of floating time between consecutive tasks, allowing for a transition from the "unaffected" to the "affected" state. This floating time, denoted as IPj, signifies the count of upstream tasks directly impacting task j within the topological structure. The greater the number of upstream tasks, the higher the likelihood of affecting connected tasks. Meanwhile, Item, derived from the duration of task j, reflects its susceptibility to upstream task failures, with longer durations mitigating the risk of infection.

The new threshold θ(new) for task j is a combination of its spreading power (SP') and infecting power (IP'). If this new threshold exceeds a certain proportion of the initial threshold (θ(initial) × α), task j transitions from the "unaffected" to the "affected" state, denoted by nj.

The degree of importance, Importance, is calculated based on the new threshold value θ(new), emphasizing the significance of the task's role in propagating failures.

To ensure meaningful comparisons, the values of IPi and IPj are normalized to fit within the range [0, 1], facilitating clear interpretation and analysis.

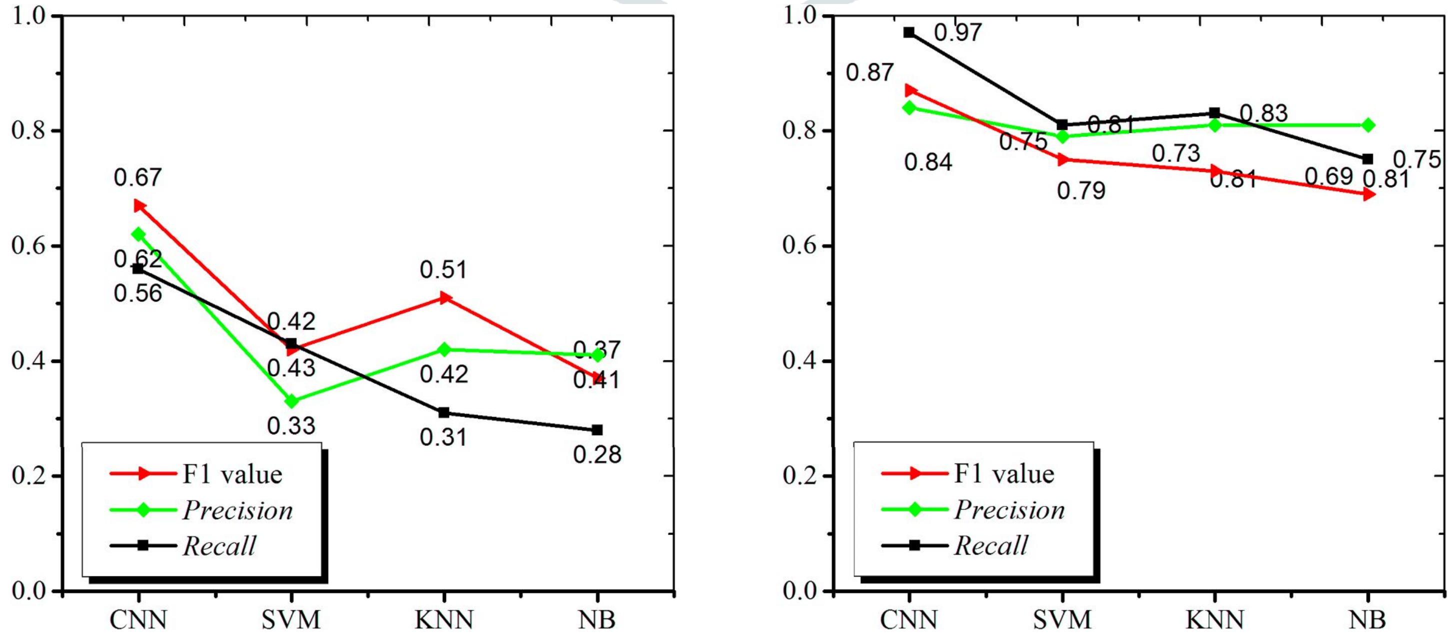
For clarity, the key concepts of task i's "spreading power" and task j's "infecting power" are summarized in Table 3.

3.2.1Ripple Effect Dynamics

External safety risk factors interfere with the completion quality of each node within the Activity-on-Node (AON) network, influencing both planned and actual completion times. Consequently, the actual completion time of a node is subject to these risk factors, resulting in variations in node completion quality. This variability can induce adjustments in downstream node thresholds, triggering the cascade effect within the AON network. Thus, the completion quality of each task, denoted as Qi, corresponds to the initial fault threshold of each node in the network, expressed as

θi(initial) = Qi

Regardless of its position within the AON network, except for the endpoint, each node represents a primary task susceptible to external safety risk factors, termed the initial task (i'). Task i' is particularly sensitive to threshold adjustments...



(a) *Electrocution* (A3) **(b)** *Exposure to Extreme Temperature* (A5)

Fig 5.Text Classification Impact of 'Electrocution' and 'Extreme Temperature Exposure' Tags

Project	Types of safety accidents	Safety risk factors	Num	Result	IoR
P1	A5	High temperature, Work overtime, Lack of heatstroke prevention measures	2	Injured (0.3)	0.6
P2	A3	Circuit aging, No work license, Charge overload	1	Disability (0.5)	0.5
	A8	Broken rope, No helmet, Illegal operation	1	Injured (0.3)	0.3
P3	A4	No gloves, No cover, Illegal operation, No warning	2	Injured (0.3)	0.6
	A5	High temperature, Work overtime, Work hard	1	Injured (0.3)	0.3

Table 5 Identifying Safety Risk Factors in Four Incidents"

Fig 5.Analyzing Text Classification Impact: Electrocuting Hazards vs. Exposure to Extreme Temperatures

The study, conducted in Wuxi City, Jiangsu Province, draws upon relevant data provided by the project management team. With a relatively low accident rate across three projects, the impact of each risk factor can be more precisely assessed. Data sourced include accident reports and project schedules, recorded electronically and preprocessed to meet evaluation requirements. To streamline the complexity of the Activity-on-Node (AON) construction, schedules were adjusted to single-project standards for method validation within a reasonable timeframe. Three project profiles, detailed in Appendix 1, offer a comprehensive overview.

The dataset encompasses three completed construction projects, summarized in Table 4. Project 1 (P1) entails a four-story teaching building constructed during a hot summer. Project 2 (P2) involves a small to medium-sized five-story library, while Project 3 (P3) features a 15-story commercial-residential building. All three projects have been in the delivery phase for an extended period, boasting complete safety accident records. As per OSHA construction safety standards, notable accidents include Extreme Temperature Exposure (A5) on P1, Electrocution (A3) and Falling Object Strikes (A8) on P2, and Chemical Substance Exposure (A4) and Extreme Temperature Exposure (A5) on P3. Table 4 provides insight into the expected and actual project progress, along with node counts.

Data collection for safety accidents primarily comprises document-based information such as accident reports and claim documents, supplemented by materials from meeting discussions. To ensure a thorough and accurate identification of safety risk factors, the study collated data from both document and meeting sources, amassing a total of seven safety accident reports and eleven meeting minutes.

4.Exploring Safety Risk Factor Analysis

Once we acquire the textual documents from the three projects, there are three key tasks to accomplish: categorizing documents, extracting factors, and prioritizing their importance.

4.1.1Document Categorization

This study introduces a CNN model aimed at classifying text documents based on accident types. The classifiers utilized within the model have undergone rigorous training procedures. The ultimate goal is to perform text classification and label assignment for four types of safety risk accidents across three distinct projects. To achieve this, three different algorithms, namely SVM, KNN, and NB, are introduced and subsequently compared using various performance evaluation metrics.

The CNN model's primary objective is to effectively categorize text documents according to the nature of accidents they describe. This is achieved through a systematic training process that ensures the classifiers within the model are optimized for accuracy and efficiency. By leveraging the capabilities of CNNs, the model is equipped to handle the complexities and nuances present in accident-related text data.

Upon completion of the training phase, the model is put to the test by performing text classification and labeling on a dataset comprising accidents from three separate projects. This evaluation stage serves to validate the effectiveness and robustness of the CNN model in accurately identifying and categorizing various types of safety risk incidents.

Furthermore, the study extends its analysis by introducing and comparing the performance of three additional algorithms: Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Naive Bayes (NB). By subjecting these algorithms to the same evaluation metrics used for the CNN model, the study aims to provide a comprehensive comparison of their respective capabilities in handling text classification tasks related to safety risk incidents.

The results obtained from the evaluation process are based on the performance of the algorithms on the test set comprising data from the three projects. Through this comparative analysis, the study aims to elucidate the strengths and weaknesses of each algorithm, ultimately providing insights into their suitability for real-world applications in accident classification and risk assessment.

4.1.2 Extraction and Prioritization of Safety Risk Factors

Utilizing the methodology delineated earlier, we proceed by segmenting the documents, eliminating stop words, and performing part-of-speech tagging operations on the annotated text documents. Following this preprocessing stage, the Apriori algorithm is employed to extract the risk factors associated with the safety incidents across the three projects under consideration. The findings of this extraction process are synthesized in Table 5. These safety risk factors are identified from the causes listed in the attribute dictionary, which represent common combinations of single vectors within the structural framework. Additionally, Table 5 presents the number of casualties and outcomes associated with the five accidents analyzed.

The assessment of risk factors concerning project safety implications exhibits a diversified perspective, highlighting that crucial features can be discerned from various angles, including the direct repercussions of safety incidents. Particularly noteworthy is the practice wherein managers assess safety incidents attributable to risk factors by considering the number of casualties in actual project evaluations. This practice mirrors the stipulations set forth by safety management laws and regulations governing project safety management. Consequently, we acknowledge the significance of safety risk factors based on the number of casualties, defining it as the Importance of Result (IoR).

Moreover, drawing on the risk level quantification method proposed by Dikmen (2007), we introduce the variable ω to denote the severity of safety incidents, with ω taking values from the set $\{0, 0.3, 0.5, 1\}$. Specifically, when an injury occurs, ω is assigned a value of 0.3; in cases of disability, ω is set to 0.5; and when a fatality occurs, ω is assigned a value of 1. A ω value of 0 indicates an absence of consequences, rendering the associated safety risk factor negligible. Consequently, we compute the corresponding IoR for each safety risk factor, as delineated in Table 5. This systematic approach enables us to assess and prioritize safety risk factors based on their potential impact, as well as the severity of the outcomes associated with safety incidents.

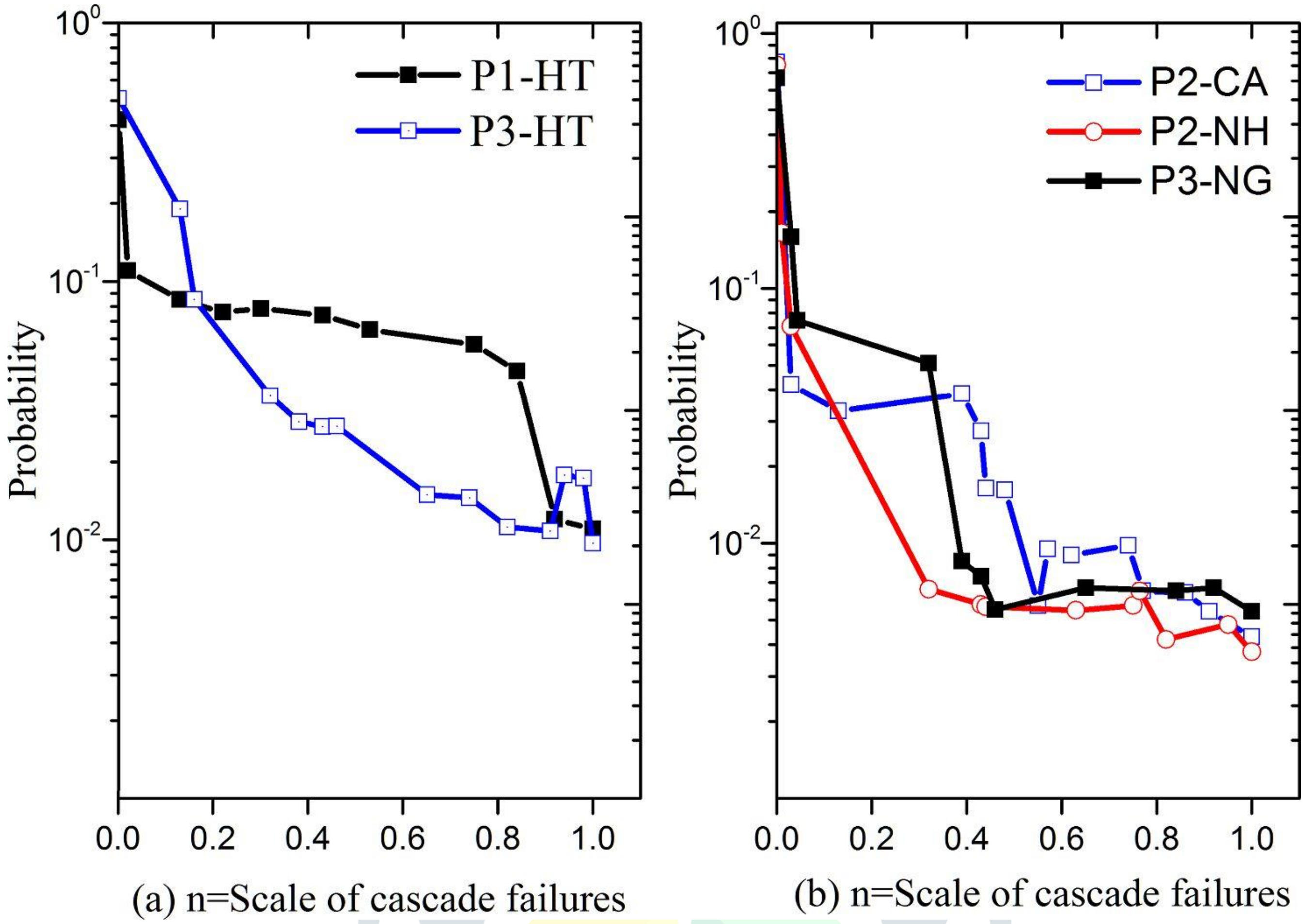


Fig 6 Probability Distribution of Cascading Failure Magnitude for Five Risk Factors. Note: The cascading fault scale is normalized, with coefficient α set at 0.5, and ω at 100%

Following the identification of 16 safety risk factors as sampling features from the original dataset, the Random Forest (RF) algorithm is applied to compute the Importance of Feature (IoF) for each safety risk factor. Given that the values of Mean Decrease Accuracy (MDA) and Mean Decrease Gini (MDC) are normalized to a range between 0 and 2, they represent the combined influence of IoR and IoF on the project's safety.

In Fig. 5, the values presented are the summation of IoR and IoF, signifying the collective impact of safety risk factors considering both aspects. To classify the importance, two criteria are established: Standard 1 designates a degree of "high" importance when either IoR > 0.5 or IoF > 1, while Standard 2 categorizes a degree of "very high" importance when the summation value exceeds 2, and a degree of "moderate" importance when the summation value is below 1. It's noteworthy that Standard 1 takes precedence over Standard 2, as surpassing a certain threshold in single importance inherently entails a qualitative change in its project impact, regardless of other factors.

As depicted in Fig. 6, five of the 16 mined safety risk factors attain the "very high" level of importance, including "high temperature", "circuit aging", "lack of helmet", "lack of gloves", and another instance of "high temperature" (classified under A5), while only "working overtime" (0.68) and "excessive physical exertion" (0.72) fall into the "moderate" category. Two factors are classified as "moderate", while the remaining are deemed "high". Additionally, it's pertinent to consider that the RF algorithm analyzes safety risk factors already mined, potentially overlooking hidden factors not captured in the actual project dataset due to low word frequency or indirect impact. Consequently, there's a possibility of incomplete feature traversal within the dataset.

4.3Exploration of Cascade Effects

To assess the real impact of safety risk factors on the project, we compute the likelihood of cascading failures occurring within the Activity-On-Node (AON) network nodes and evaluate these cascading failures across three distinct projects. Furthermore, considering the influence of safety risk factors on each project, we conduct a comprehensive analysis of the project's risk level to provide guidance for project risk management.

4.3.1Influence of AON Network Failures

The most critical safety risk factors across the three projects are abbreviated as "HT," "CA," "NH," "NG," and another instance of "HT" (belonging to A5). These five factors are considered of utmost importance, all classified as "very high." They serve as influencing terms to trigger cascading effects within the Activity-On-Node (AON) network, impacting the size of cascading failures for each node and the respective node distribution probabilities. The findings, illustrated in Fig. 6, reveal that the completion time for tasks is expected to double due to the direct interference of these risk factors with project scheduling.

Analyzing the probability distribution of cascading fault scale enables a better understanding of the cascading effects induced by various factors. For instance, the results for P1-HT and P3-HT (Fig. 7 (a)) indicate probabilities of 0.42 and 0.51, respectively, when cascading faults do not occur (n_i=0), suggesting that nearly half of the tasks are impacted by safety risk factors. Notably, the probability of "high temperature" activating cascading effects in P1 is the highest among all factors, reaching 1.1% (equivalent to 11 tasks experiencing failure). This high probability is attributed to the prolonged exposure to extreme temperatures during the summer construction period, leading to continuous impacts despite the occurrence of accident A5. Conversely, the probabilities of failure induced by CA, NH, and NG are relatively smaller. NH exhibits the lowest probability (0.37%), followed by CA with a probability of 0.43%. Nevertheless, the impact on tasks is significant for P2 due to its numerous AON nodes.

Furthermore, a combined analysis of IoR, IoF, and probability distribution allows for a deeper understanding of risk factors' impact. For instance, P1-HT has the highest importance value (2.30) and the largest probability of cascading failures, highlighting that higher importance correlates with a greater likelihood of project task failure. This underscores the varying impact of different risk factors on project outcomes. Despite the slight difference in importance between CA and NH, the higher IoR for A3 accident-related disabilities indicates a larger probability of cascading effects.

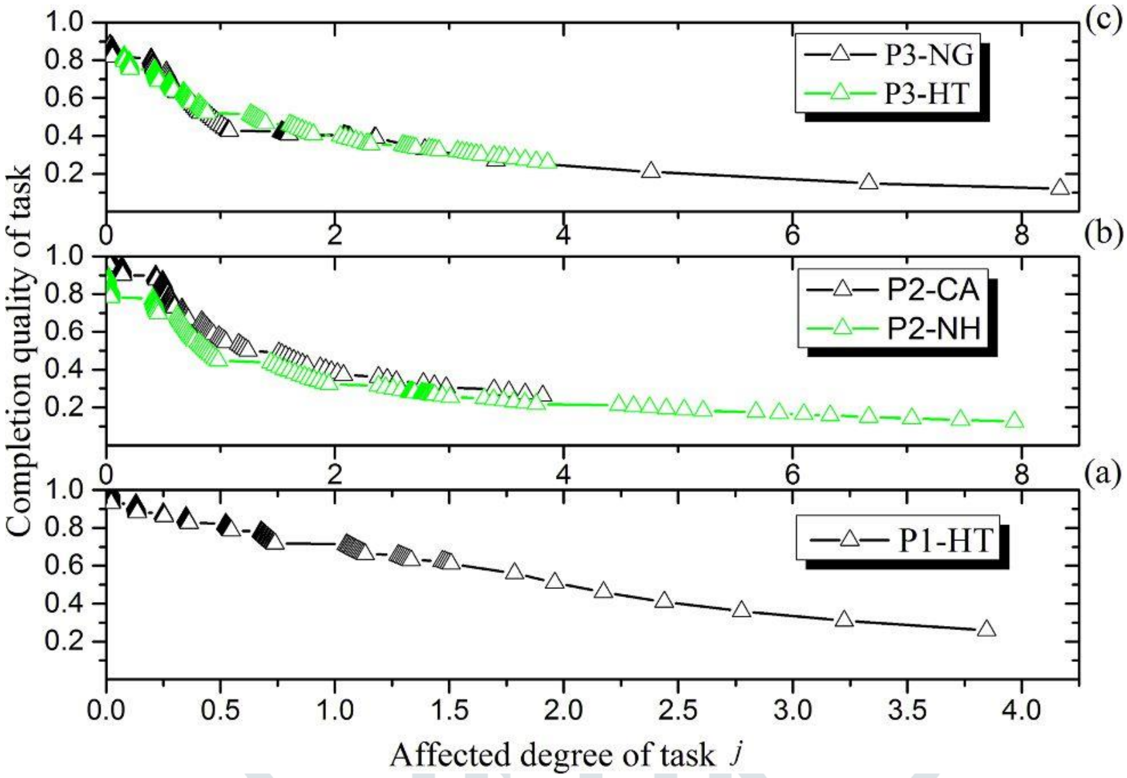


Fig 7 Effects of Five Key Risk Factors on Project Response

The disparities in project impacts resulting from risk factors are evident from the occurrences of failures.

4.3.2 Differential Analysis of Factor Effects.

The magnitude of cascading failures triggered by safety risk factors aligns with the entire project dataset. To further quantify the degree of influence exerted by each risk factor on individual tasks under the cascade effect, we calculate the initial threshold value for each node within the Activity-On-Node (AON) network to gauge the completion quality of each task. Subsequently, we determine the new threshold for each task considering the combined influence of "spreading power" and "infecting power." Given that multiple tasks are directly impacted by risk factors such as HT and NH, we set several values of j when computing the initial threshold value (θi(initial)). Consequently, additional effects emerge among the affected tasks.

The response effect of the five paramount risk factors identified across the three projects is illustrated in Figure 8. To ensure clarity, we extract data solely from nodes affected by safety risk factors and conduct case analyses specific to Projects 1, 2, and 3. Firstly, despite the risk factors resulting in a doubling of the actual completion time (ε=1) for Project 1, the completion quality of 34 tasks still exceeds 90%, with nearly half of the tasks achieving over 50% completion quality. This indicates effective risk control measures implemented by the project team, as supported by unstructured documents for Project 1. These documents detail the team's proactive measures to mitigate the adverse effects of high-temperature construction. Hence, while "high temperature" poses a significant safety risk, prudent management measures can mitigate its negative impact.

Secondly, Project 2 is affected by two risk factors, CA and NH, with CA impacting significantly fewer tasks than NH. Machine learning-derived insights indicate that "Electrocution" incidents associated with CA are infrequent and typically affect individual tasks, suggesting a limited scope of influence. In contrast, the Project 2 team conducts regular safety helmet checks, halting operations for several days if any issues are detected, thus extending the impact duration. Furthermore, the quantitative impact values for CA and NH are 3.8 and 7.9, respectively, highlighting the profound impact of not wearing safety helmets on project outcomes.

Lastly, in the context of Project 3, the variance in the effects caused by risk factors is evident once more. The number of tasks impacted by HT far exceeds those affected by NG, indicating a more significant impact of HT on Project 3. Compared to "no helmet," the impact of "no gloves" is relatively minor, as instances of wearing gloves during chemical contact are rare and may only affect specific tasks. Despite all three projects being summer constructions, the response effect on task j is lowest among them, emphasizing the contextual nuances shaping risk factor impacts.

5.Discussion of Risk Factor Impacts and Project Responses

The risk factor integrated analysis framework (IAF) presented in this study offers significant potential to enhance project management practices by enabling project managers to better predict and address safety risks. Despite the persistent emphasis on safety management, global construction accidents continue to result in fatalities each year, underscoring the urgent need for more effective risk management strategies (Farooq et al., 2018). Traditional methods of identifying project safety risk factors suffer from subjectivity issues and lack robust quantification of their impact, thereby limiting their efficacy in mitigating risks comprehensively. Thus, accurately identifying risk factors and quantifying their effects are crucial steps in addressing safety risks effectively.

Through our research and its managerial implications, several key observations can be gleaned. Firstly, leveraging heterogeneous data and employing machine learning techniques, such as CNN models and RF algorithms, facilitates the accurate identification of safety risk factors from project-related text documents. These methods demonstrate effectiveness, particularly as they can adapt and improve with updates to the dataset, enhancing their utility in practical applications. Secondly, by developing a cascade effect model, we can assess the impact of safety risk factors comprehensively. A fundamental tenet of management theory is the adoption of responsive measures, which necessitates targeted responses based on the importance of identified risk factors. Importantly, assessing the importance of safety risk factors extends beyond their impact on individual accidents; it also considers their broader implications for the entire project. Even if risk factors affect a single event within a specific task, their impact can propagate through the Activity-On-Node (AON) network once activated, underscoring the importance of considering cascade effects.

Hence, the IAF introduced in this study evaluates the significance of cascading effects resulting from risk factors leading to task failure, presenting substantial practical implications. Unstructured text data generated during project safety management contains valuable information, and leveraging machine learning to process this data can uncover potential risk factors. Integrating cascade effects into project management enables the quantification of risk factor impacts on the entire project, enhancing the relevance and practicality of project safety management efforts. The application of our research results can be categorized into two main areas. Firstly, by mining safety accident reports for risk information, project managers can identify factors leading to accidents and their corresponding consequences, providing valuable insights for future event responses. Secondly, the framework proposed herein exhibits universal applicability across various project management documents, enabling the identification of potential safety risks during analyses of meeting minutes, construction logs, and other project-related documents.

6.Conclusions

This study introduces an Integrated Analysis Framework (IAF) aimed at enhancing project safety management by supplementing existing risk management knowledge systems. It represents an innovative approach to identifying potential risk factors in heterogeneous documents through machine learning techniques. Specifically, the study utilizes accident types to filter documents, enabling the identification of associated typical risk factors. The Importance of Feature (IoF) and Importance of Result (IoR) of risk factors, incorporated into Random Forest (RF) calculations, serve as disturbance terms to activate cascade effects. Furthermore, by establishing Activity-On-Node (AON) networks corresponding to three projects, the study activates cascade effects within these networks through task "failure" induced by risk factors with importance attributes. This enables an analysis of the impact of such failures on the entire project. Through machine learning and cascade effect evaluation, the study provides insights into both the safety risk factors within each project and the degree of influence of each risk factor on the project as a whole. Case validation results demonstrate the effectiveness of the IAF in identifying and evaluating risk factors, offering direct assistance to managers in enhancing safety risk management practices.

Despite the substantive progress made in project risk management, there are certain limitations in this study. Firstly, machine learning techniques require extensive datasets for processing to achieve optimal results, encompassing both unstructured text data requiring further preprocessing and cases for training CNN models. While rich datasets can enhance the accuracy of factor mining and model classification, they also demand significant human resources. Thus, there's a need to optimize algorithms and models to reduce dependency on large datasets. Secondly, the AON network considered in this study only addresses project schedules.

7.REFERENCES

1. Choudhry, R.M., Iqbal, K. (2013). Identification of risk management system in construction industry in Pakistan. Journal of Management in Engineering, 29(1), 42–49. https://doi.org/10.1061/(asce)me.1943-5479.0000122.
2. Fang, C., Marle, F. (2012). A simulation-based risk network model for decision support in project risk management. Decision Support Systems, 52(3), 635–644. https://doi.org/10.1016/j.dss.2011.10.021.
3. Fang, D., Wu, H. (2013). Development of a safety culture interaction (SCI) model for construction projects. Safety Science, 57, 138–149. https://doi.org/10.1016/j.ssci.2013.02.003.
4. Heravi, G., Hajihosseini, Z. (2012). Risk allocation in public-private partnership infrastructure projects in developing countries: Case study of the Tehran-Chalus toll road. Journal of Infrastructure Systems, 18(3), 210–217. https://doi.org/10.1061/(asce)is.1943-555x.0000090.
5. Kaiser, M.G., El Arbi, F., Ahlemann, F. (2015). Successful project portfolio management beyond project selection techniques: Understanding the role of structural alignment.
6. Kuo, Y.-C., Lu, S.-T. (2013). Using fuzzy multiple criteria decision-making approach to enhance risk assessment for metropolitan construction projects. International Journal of Project Management, 31(4),

602–614. https://doi.org/10.1016/j.ijproman.2012.10.003.

7. Liu, J., Li, Q., Wang, Y. (2013). Risk analysis in ultra deep scientific drilling project — A fuzzy synthetic evaluation approach. International Journal of Project Management, 31(3), 449–458. https://doi.org/10.1016/j.ijproman.2012.09.015
8. Liu, J., Zhao, X., Yan, P. (2016). Risk paths in international construction projects: Case study from Chinese contractors. Journal of Construction Engineering and Management, 142(6), 05016002. https://doi.org/10.1061/(asce)co.1943-7862.0001116.
9. Marcelino-S´adaba, S., P´erez-Ezcurdia, A., Echeverría Lazcano, A.M., Villanueva, P. (2014). Project risk management methodology for small firms. International Journal of Project Management, 32(2), 327–340. https://doi.org/10.1016/j.ijproman.2013.05.009.
10. Tavakolan, M., Etemadinia, H. (2017). Fuzzy weighted interpretive structural modeling: Improved method for identification of risk interactions in construction projects. Journal of Construction Engineering and Management, 143(11), 04017084. https://doi.org/10.1061/(asce)co.1943-7862.0001395.
11. Wang, J., Yuan, H. (2017). System dynamics approach for investigating the risk effects on schedule delay in infrastructure projects. Journal of Management in Engineering, 33(1), 04016029. https://doi.org/10.1061/(asce)me.1943-5479.0000472.
12. Wang, L., Zhang, X. (2018). Bayesian analytics for estimating risk probability in PPP waste-to-energy projects. Journal of Management in Engineering, 34(6), 04018047. https://doi.org/10.1061/(asce)me.1943-5479.0000658.
13. Xia, N., Zou, P.X.W., Liu, X., Wang, X., Zhu, R. (2018). A hybrid BN-HFACS model for predicting safety performance in construction projects. Safety Science, 101, 332–343. https://doi.org/10.1016/j.ssci.2017.09.025.
14. Zhang, S., Teizer, J., Lee, J.-K., Eastman, C.M., Venugopal, M. (2013). Building information modeling (BIM) and safety: Automatic safety checking of construction models and schedules. Automation in Construction, 29, 183–195. https://doi.org/10.1016/j.autcon.2012.05.006

.