Internasional Journal of Integrative Sciences (IJIS)
Vol.5, No.2, 2026: 371-384

Predictive Risk Modeling via Natural Language Processing of Industrial Safety Reports

Kimberly Long Holt
Health and Safety Concepts-Environmental Health & Safety
Corresponding Author: Kimberly Long Holt ka3027@yahoo.com

A R T I C L E I N F O

Keywords: Natural Language Processing, Machine Learning, Random Forest, Safety Management, Injury Prediction

Received : 5 December 2025
Revised : 20 January 2026
Accepted: 21 February 2026

A B S T R A C T

Rapid management of industrial safety is producing long incident investigation reports, which are not fully reflected in the use of text. The study examines Natural Language Processing (NLP) and Machine Learning (ML) in transforming qualitative messages on safety into quantitative predictive messages. A total of 16,878 records of construction accidents were used to test different ML algorithms in terms of usefulness and Elasticity. The model of the Fields of Application was the Random Forest with an accuracy of 79.3%, 77.1% precision, 78.0% recall, and an Area Under the Receiver Operating Curve (AUROC) of 0.98. The analysis of the importance of the features revealed that accident mechanism and nature were mainly important predictors, whereas the temporal and economic factors were the least affected. These results support the effectiveness of NLP when using unstructured safety data, which provides practitioners with a resource-based and proactive risk intervention tool as opposed to a reactive mechanism

INTRODUCTION

Workplace accidents can be noted as one of the most important social, health, and economic issues in every industry. The indirect costs (human and financial costs) are not immediate incidences but have repercussions on the workers, the families, organizations and the community. According to recent epidemiological studies, non-fatal accidents cause significant productivity losses; in seven years, the insured in Greece suffered over 22,000 occupational injuries, which led to over 22,000 working days (Papazoglou et al., 2025). These are real losses in the economy and wasted potential of the workforce, huge insurance premiums, fines in the form of regulations and reputation lost to the organization.

The traditional safety management is more or less reactive-oriented, where the organisations are reactive to what has occurred, even after the incidents. Although reporting systems are taken seriously by the companies, they are still done manually or with basic trend analysis of the systematised fields. These standardized methods document the incidences and unearth the ostensible fashions. They fail to find sensitive relations and micro-indications that are embedded in written reports of near misses, hazardous situations, and actual incidents, however. Based on the accident sequence details, environmental incidents, human factors, equipment conditions and situational factors, the safety professionals document accidents in comprehensive detail in the accident account. However, the untapped information on the precursors and mechanisms of accidents is also provided in the free-text reports (Baker et al., 2020).

The opportunities that have been provided by the technologies of artificial intelligence and data sciences have presented unprecedented chances in changing the safety management practices. Natural Language Processing enables machines to automatically classify data and discover things in reports like causes and consequences (Ricketts et al., 2023). NLP is concerned with the interaction between a computer and human language, and this enables an organization to process large volumes of textual data much faster and reliably than human beings can. It has shown a lot of promise in terms of application in the aviation field (Kuhn, 2018; Nanyonga et al., 2025; Robinson, 2019), the healthcare sector (Wong et al., 2018), and even the field of chemical processing (Song & Suh, 2019).

NLP application in safety management is a revision of the paradigm with regard to reactive and proactive evaluation of risk. The NLP algorithms can cut latent linguistic patterns, recurrent patterns, as well as risk factors, that would otherwise be missed by human beings, rather than a manual review of data, taking place before an organization can infer the existence of any patterns (Liu & Yang, 2022). This progression causes the safety professionals not to rely on the numeric values of frequencies, making a complex, foreseeable modelling of injury incidences based on the story properties.

The more recent developments have shown the maturity in improving AI-based safety analytics. Ensemble-based classifier machine learning methods demonstrated good predictive ability of factor analysis and scenario-based accident prediction in the construction industry (Kim et al., 2025). At the same time, NLP use in near safety-critical industries, including aviation, remains on the rise, which contributes to the importance of transferability of these algorithms across the industries (Nanyonga et al., 2025). Altogether, this literature sets the direction for the transition to information-driven, proactive safety systems, which can help inform decision-making on large scales.

Although this has been achieved, there are still big gaps in the literature. The majority of current studies use NLP either in the classification of accidents or root causes in one industry or small datasets that are domain-specific. A limited number of studies have deployed an end-to-end NLP pipeline - i.e., raw incident narrative preprocessing - to ensemble classification of multi-category injury type prediction under the construction industry. Furthermore, the practical external validity of these types of pipelines between different organizational and project backgrounds is underresearched. To fill this gap, a methodology needs to not only be validated in large-scale on an empirical basis, but must also be replicable across different settings.

The given paper provides a data-driven approach that forecasts the types of damages, and it has a proactive prediction of risk points by NLP and machine learning. Thanks to big data of construction accidents and the latest algorithms, one can demonstrate how the organizations might transform the already available documentation on safety and turn it into a powerful predictive tool (Kim et al., 2025). The methodology can offer a generalizable, replicable methodology that can be used to conclude a broad spectrum of industries and organizational contexts with general practical implications to the safety managers, project managers, and organizational leaders whose responsibilities entail the safety of workers.

LITERATURE REVIEW

Natural Language Processing forms an important step towards eliciting significant information using textual information that is not textually structured. Recent sources show the increased use of NLP in the area of safety. In the study by Rickett et al. (2023), NLP used to safety occurrence reports was identified with the help of a scoping review that revealed automated classification and entity extraction features. These systems can automatically extract causes, consequences, and contributory causes out of narrative reports and can save a lot of burden on manual analysis. This is the capacity that provides the conceptual basis for the use of NLP in the construction of incident narratives in the current research, where rich yet unstructured textual reports are the major data source.

The application-specificity in the industry proves the versatility of NLP. Structural topic modeling, as developed by Kuhn (2018) in the context of aviation, was used to conduct latent topic discoveries and trends in incident reports, and temporal topic modeling with subject matter expert review was used by Robinson (2019). Nanyonga et al. (2025) give a detailed review and qualitative analysis of NLP, which is applied to aviation safety. The same can be said about healthcare apps: Wong et al. (2018) have reviewed the implications of NLP on medication safety. Song and Suh (2019) made use of narrative text-based anomaly detection methods on chemical process safety documents. Their cross-domain success also contributes to the fact that the transfer of the approaches to the construction safety setting is possible, as the stories of incidents have similar unstructured textual features, which strengthens the theoretical foundation of the approach chosen in this study.

Algorithms of machine learning, especially those that are based on ensembles, are more effective in classification tasks. Random Forest algorithms are those algorithms that build several decision trees and combine predictions to achieve higher accuracy and resilience (Ricketts et al., 2023). Kim et al. (2025) presented machine learning solutions to the issues of factor analysis of construction accidents and prediction scenarios. These experiments prove that accident stories possess substantial knowledge to make consistent predictions on the injury conditions upon running through relevant NLP pipelines. This performance advantage over alternative classifiers directly justifies the selection of Random Forest as the primary classification algorithm in the present study.

Text preprocessing algorithms have a significant impact on the model performance. According to Sankarasubramanian and Ganesh (2020), all the necessary measures, such as tokenization, stop-word removal, stemming/lemmatization, and normalization, are described. Term Frequency-Inverse Document Frequency (TF-IDF) feature engineering is an effective method to highlight such specific risk words and underline common words. In their work, Liu and Yang (2022) have illustrated the use of text mining to create the knowledge graph using accident reports in risk assessment, showing the possibility of extracting structured knowledge through the processing of unstructured descriptions. These preprocessing and feature engineering findings directly inform and validate the pipeline adopted in the present methodology.

Baker et al. (2020) have considered, in particular, automatic learning of building injury precursors in text and have demonstrated that narrative descriptions capture predictive patterns of injury occurrence. Although Baker et al. (2020) demonstrated the precursor learning in a construction text, their research was conducted in a restricted setting and not analyzed on the imaging of multi-category predicting injury types on a large dataset. It is anchored to the fact that, based on solving this deficiency, the unstructured safety reports can be compiled in a structured informational form through NLP pipelines, which can be sliced by the machine learning algorithms to produce plausible predictive intelligence so as to effect proactive safety control.

Research Contribution and Novelty

The current research has a number of contributions to the current literature collection. Even though the NLP has been proven useful in safety fields by previous research, the majority of studies were either limited to one industry or a smaller set of data, or their classification scopes were restricted to binary injury forecasting or root cause classification. These limitations are countered in this work where a full, end-to-end NLP pipeline, including text preprocessing, TF-IDF feature engineering, and Random Forest classification, are tested on a large-scale data set of 16,878 construction accidents that had to be handled in 16,878 incident reports in total (Alkaissy et al., 2023). Notably, the research presents the prediction clarity as multi-category injury type classification, differentiating between upper and lower limbs and head/ neck injuries, which is more operationally advantageous data to safety managers in comparison to binary solutions. Also, the methodology is to be transparent and replicable, as it can be applied in the different industries and organizational contexts other than the construction. Together, these aspects make the current research unique in relation to the current literature and make it an important methodological as well as an applied contribution to the concept of predictive safety analytics.

METHODS

This study is a quantitative research design in the form of an applied research design and predictive modelling orientation. The kind of research is non-experimental and secondary since the research does not involve primary data collected in the form of surveys and observations, but uses the large-scale accident database established previously. The report of individual incidence is the data unit, whereas the study sample will consist of all the registered construction accidents in the Alkaissy et al. (2023) dataset of 2018 to 2022 (N = 16,878). A general NLP pipeline was followed, which is in line with available solutions in the literature on industrial safety (Sankarasubramanian & Ganesh, 2020), and the transformation of unstructured textual descriptions of incidents into formatted numbers, which can be used to support machine learning classification. The phases incorporated in the analysis process are four, totaling data collection and quality screening (1), systematic text preprocessing and normalization (2), the TF-IDF feature engineering (3) and random forest ensemble classification with a strict version of cross-validation (4). The stages were all to overcome the very nature of unstructured safety stories, including, but not limited to, variable linguistic variation, use of domain-specific language, and unevenness of differing types of data points in terms of injury. The combination of these stages into an analytic pipeline can be utilized to produce functional predictive intelligence involving actionable forecasts of complex safety paperwork.

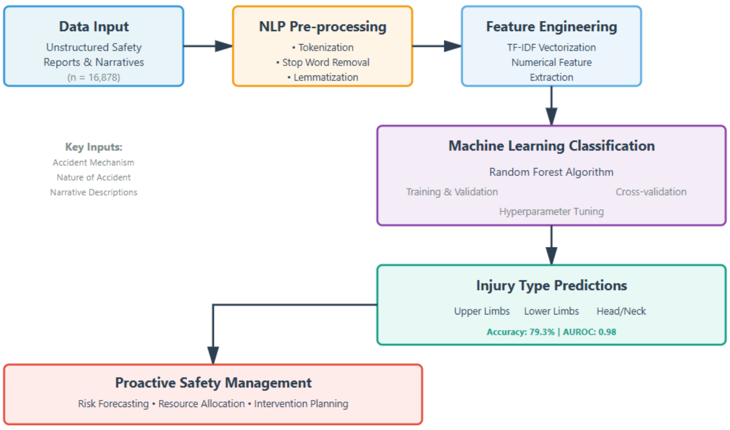

Figure 1. Conceptual Framework

Data Collection

The number of construction accidents was 16,878 (Alkaissy et al., 2023). This rich source of information was a good training and validation source because it represented the wide range of events of accidents, injury mechanisms, and outcome patterns that were seen in professional contexts of construction. Another advantage of the data choice was in the construction field, where the incidence of injury was high and detailed narrative reporting procedures were commonly employed in construction safety management. The accident record system contained both structured (date, time, location, cost category) and unstructured narrative descriptions, which explained the sequence of events, environmental influences, equipment employed, and the behavior of workers. Data were screened for quality to ensure completeness and trustworthiness, and log records were discarded where critical injury categories were absent or below 20 words required narrative content. The five-year data collection period (2018-2022) provides a temporal change that demonstrates the evolving safety strategy and the impact of seasonal changes in the accident pattern. This dataset notes some construction projects, including residential, commercial, infrastructure, and industrial facilities, which render the conclusion applicable to various construction settings. Ethical concerns were addressed by making sure that all personally identifiable information was anonymized and that the data protection laws regarding the occupational safety records were respected.

Data Preprocessing

Radical preprocessing normalised and machine-learned text. The raw incident report has a high occurrence of spelling, grammatical, abbreviation, and formatting aberrations that turn out to be counterproductive to model operations. It was also a complicated process involving a number of significant steps as described by Sankarasubramanian and Ganesh (2020). Special consideration was given to the computation of terms in the construction field, e.g., industry abbreviations (e.g., PPE, OSHA, JSA), terminologies used in equipment, and colloquial language in safety. Python 3.9 preprocessing pipeline was applied with the assistance of such libraries as Natural Language Toolkit (NLTK 3.8) and spaCy (3.5), and had a foundation in their pre-trained English language models, with the new vocabulary covering the construction safety domain.

The Unicode standard text normalization was performed to allow the text to support special characters, more frequent contractions, and remove the encoding anomalies. The cases were normalized with lower case, and this assisted in achieving the exact token matches; all the text was converted to lower case, although the acronyms were kept where they were applicable to the context. The eradication of punctuations was partial, and the ones that possess a semantic meaning (the ones that include hyphens), like fall-protection, were kept. The preprocessing sequence was optimized through iterative testing to maximize retention of safety-relevant information while minimizing noise that could degrade classification performance. The most important steps of NLP preprocessing implemented are summarized in Table 1.

Table 1. NLP Preprocessing Steps and Their Functions

Preprocessing Step	Function	Example
Tokenization	Decomposition of sentences into discrete words or tokens	"Worker fell from scaffold" → ["Worker",

		"fell", "from", "scaffold"]
Stop Word Removal	Removal of common function words that do not contribute meaning	["Worker", "fell", "from", "scaffold"] → ["Worker", "fell", "scaffold"]
Stemming/Lemmatization	Reduction of words to root form to standardize variations	["falling", "fell", "fallen"] → ["fall", "fall", "fall"]
Normalization	Conversion to lowercase and removal of special characters	"SCAFFOLD-related" → "scaffold related."

The tokenization converted continuous texts into separate linguistic units that can be analyzed separately and together. Stop word removal was used to remove function words such as the and is, which have low discriminative ability in classification tasks, and reduce computational complexity and dimensionality. Stemming and lemmatization reduced words to root morphemes, and words that had various inflections and derivations based on the same root idea were treated equally. Other methods of data quality were blank values management, elimination of duplicate records, and the resolution of text encoding problems.

Feature Engineering

Text data underwent cleaning, which was converted into numerical feature vectors. It used the Term Frequency-Inverse Document Frequency (TF-IDF) to highlight distinguishing risk words and downscale generic words (Sankarasubramanian & Ganesh, 2020). TF-IDF also attaches greater weights to words in individual documents and in rare occurrences in corpora overall, which is helpful to bring out unique vocabulary that may define particular accident situations or patterns of injury. The TF-IDF appearance came based on 1-3 n-grams to consider not only single words but also multi-word phrases characterizing a given scenario of the events that pertained to accidents (e.g., struck by a falling object, inadequate fall protection). The feature limit, which was used to perform the vectorization, was a maximum feature limit of 5,000 terms, so that the model is not too complicated but can identify the desired terms, which in this case were identified by a frequency threshold in the document (minimum frequency of 5 times in the document and 85 percent document frequency).

The importance of the features as defined by chi-squared tests was analyzed and showed the most selective terms of each kind of injury, showing a few domain-specific linguistic signs that significantly correlate with distinct forms of injuries. Other feature engineering involved the derivation of finer metadata features, which included time of day, day of week, project type, and cost type, and the expansion of finer representations of accident scenarios. To define the circadian and weekly changes in the number of accidents, temporal characteristics were treated as categorical (morning/afternoon/ evening shift; weekday/ weekend). The type of project was categorized into residential, commercial, industrial, and infrastructure construction, which had varying risk profiles. The final feature space combined 5,000 TF-IDF-weighted text features with 15 categorical metadata features, creating a comprehensive representation of each accident record that captured both narrative content and contextual factors.

Classification Algorithm

Despite trying several of the algorithms (including Support Vector Machines and Logistic Regression), the algorithm of Random Forest was chosen as it performed even better (Kim et al., 2025). Random Forest builds several decision trees and combines them to make specific and reliable predictions (Ricketts et al., 2023). In this ensemble method, the wisdom of crowds is built into a collection of weak learners that can be strong learners whose predictions are more accurate with a more potent force. The algorithm is an automatic high-dimensional feature space designed around text vectors that is easy to use with built-in feature importance measures that can be interpreted, and is resistant to outliers and noisy data. The Random Forest implementation utilized the scikit-learn 1.2 library with specific hyperparameters optimized through 5-fold stratified cross-validation. Key parameters included: number of estimators (trees) = 200, maximum tree depth = 30, minimum samples per leaf = 5, and maximum features per split = square root of total features.

These settings balanced model complexity with generalization capability, preventing overfitting while maintaining high predictive accuracy. To raise resilience to the variability of the data, the model used bootstrap sampling with replacement to generate a variety of training samples per tree. Out-of-bag (OOB) error estimation also gave a fair evaluation of the model training performance, which supplemented formal validation measures. The problem of class imbalance was tackled based on stratified sampling and adjusting the weight of classes based on the inverse frequency of classes, to ensure that the minority injury categories are well represented. The ranking of the inherent feature of the algorithm as the source of interpretation assisted in identifying masterful linguistic and contextual predictors, thus allowing high-risk accident narratives and eventualities to be identified.

The model was trained to anticipate the existence of some injury types (upper limbs, lower limbs, and head/neck injuries), based on the textual description of mechanisms/nature of accidents (Alkaissy et al., 2023). Familiar cross-validation methods were used to do model training so as to provide dependable performance estimates and to avoid overfitting. The data was divided into training and test sets, and hyperparameter optimization (model configuration) was obtained by grid search processes of the maximum model performance with regard to prediction. Specifically, the dataset was divided into 80 per cent (13,502 records) training and 20 per cent (3,376 records) final testing with stratified random sampling, and the relative representation of each type of injury in each subset was maintained. The training set was 5-fold stratified cross-validated; during the 5 validation folds, they were created, and they did not overlap to determine the stability and generalization of the model.

The grid search was applied with 48 combinations of hyperparameters with four primary values, namely the number of estimators (100, 150, 200), the maximum depth (20, 30, 40, unlimited), the minimum samples per leaf (3, 5, 10), and the maximum number of features (sqrt, log2). The F1-score was taken to be the macro-averaged F1-score on every combination, and the F1-score strikes a balance between the precision and the recall of all injury types. The highest F1-score had a configuration of 0.847 (standard deviation of 0.019), meaning that it consistently achieved a higher result among the validation folds. Model training employed early stopping based on OOB error convergence, with training terminated when error stabilized for 20 consecutive iterations. The final model was retrained on the complete training set using optimal hyperparameters before evaluation on the held-out test set, ensuring maximum utilization of available training data while maintaining rigorous validation integrity.

RESULTS

The measures of evaluation in standard machine learning evaluate model performance. The holistic evaluation system involved several complementary measures that identified various aspects of performance, giving a complete picture of predictive opportunities and possible constraints. The Random Forest algorithm performed better than other algorithms, such as XGBoost and Support Vector Machines. A detailed performance in the Random Forest classifier is presented in Table 2.

Table 2. Random Forest Classifier Performance Metrics (Alkaissy et al., 2023)

Metric	Performance	Interpretation
Accuracy	79.3%	High overall correctness in predicting injury types
Precision	77.1%	Reliability of the model when it predicts a specific injury
Recall	78.0%	Ability of the model to capture actual positive cases
F1 Score	78.5%	Balanced harmonic means of precision and recall
AUROC	0.98	Excellent ability to distinguish between classes

A 79.3% accuracy rate would approximate a 4 out of 5 cases of correct prediction of the type of injury, which is significant progress compared to two other factors, so to speak, which are baseline guessing and simple, more straightforward rules. A precision score of 77.1% is an indicator that optimistic predictions are accurate over three-quarters of the time, which is helpful for safety managers in distributing limited resources per the output of the model to decrease false alarms and prevent the wastage of resources. The ability to name all real cases of each type of injury, which is assessed by a recall score of 78.0% is important in safety considerations because omission of a possible injury would expose workers to inadequate levels of safety. An F1 score = 78.5% gives a moderate evaluation in being able to check precision and recall at the same time. The exceptionally high discriminatory ability is evidenced by the high number of 0.98 of the AUROC, which is close to perfect discrimination and far superior to the random level of 0.50.

Feature Importance Analysis

Method of permutation feature importance showed that Nature of Accident and Mechanism of Injury were the most significant attributes that gauged the day of week, and project cost was not a significant predictor of injury type (Alkaissy et al., 2023). These results find that narrative descriptions of the presence of accidents and proximate situations contain the best predictive indicators of injury consequences. This can be intuitively sound when considering biomechanical aspects of accidents and the fact that the forces and exposures caused by the accidents are the direct determinants of injuries. The insignificant role of time implies that the styles of injury are predominantly preconditioned by physical peculiarities of the accidents as opposed to cyclical variations associated with the working shifts or burnout.

DISCUSSION

The transformative opportunities of Natural Language Processing and machine learning are demonstrated by having inferences that relate to industrial safety management. The basic hypothesis of the study, that unstructured safety narratives are overloaded with rich predictive cues that can be systematically extracted and used for preventive risk management, is true. The high predictive accuracy and high discriminatory ability of the Random Forest classifier support this hypothesis. This magnitude of precision, 79.3 percent, and an AUROC of 0.98, qualify the technique as one of the most successful techniques to inform safety and the rest of the successful NLP procedures applied to predict safety in the aviation industry or the medical sector (Kuhn, 2018; Wong et al., 2018).

Further collection priorities and model improvement initiatives are made through informed knowledge. One can pay attention to narrative reporting in organizations that dwell on elaborate descriptions of the mechanisms of the accidents and immediate situations, with the knowledge that they are the most liberative in predicting the model. The stated observation aligns with the observations of Baker et al. (2020), who defined it as a significant increase in the predictive accuracy of injuries as realized by the accident antecedents, which are translated into textual representations. Potential low forecastive value areas can be minimised or even eliminated from the reporting forms by reducing administrative but not analytic capacity. This kind of strategic consolidation in the data collection protocol measures is nonetheless a massive change to the archaic method of total documentation procedures that envies most of the information that, in the real sense, is unnecessary.

Such outcomes provide a shift of the lagging indicators to the leading ones for safety managers and project managers. Other measures of safety, which are conventional measures like Total recordable Incidence rate and lost Time Injury Frequency rate, are also retrospective factors since they measure the performance of the past and should not give a guideline on how to prevent them in the future. The developed model permits entirely different approaches, and the previous descriptions of events in history are applied to inform the future determination of risks. It is this paradigm shift that Kim et al. (2025) also call scenario-based prediction that has enabled organizations to forecast the potential risks and mitigate them even before they turn into reality.

The model gives an adequate representation of latent risk factors with an AUROC of 0.98 using text narratives. This remarkable discriminating capacity is higher than the norms of other types of such research. It is a testament to the fact that narrative content contains a lot more predictive information than is usually retrieved through manual examination means. This validation defines new chances of proactive intervention. Trained models may be used by the safety managers to near-miss reports, hazard observations, and other pre-incident documentation to predict possible injury outcome prior to actual damage (Baker et al., 2020; Liu & Yang, 2022). The predictive ability allows allocating resources in advance, interventions targeted by training, and design controls based on worst-case events. A potential application that is of specific value is the potential to detect the high-risk patterns in near-miss data since such incidents are more frequent than actual injuries and encompass early warning signals of the possible catastrophic consequences.

The automation factor would discuss problematic bottlenecks in safety management where the number of incident reports is usually more than can be reviewed in its entirety under the current manual audit environment. With initial classification and scorecard on risk being the first step in machine learning, safety professionals can direct their attention to the highest-priority cases and strategic decision-making instead of data processing all the time (Ricketts et al., 2023). This allows the automation of thousands of report types, which are otherwise labour-intensive and prone to human error, and makes allocation of resources to high-risk operational mechanisms depending on models. Similarly, it was shown by Sankarasubramanian and Ganesh (2020) that automated NLP techniques save much time on analysis and do not decrease or worsen the analysis accuracy on par with traditional approaches.

Moreover, models have objective and consistent risk assessment bases that complement human judgment rather than eliminate it. The expertise in the safety domain, contextual skills, and moral judgment that safety professionals contribute to safety decisions cannot be replaced. Predictive models are not used to displace human potential, but are a means of data-driven decision support that the safety manager is able to incorporate into practice based on experience. The model of human-machine collaboration has worked in a variety of industries where NLP assists and does not replace human judgment (Robinson, 2019; Nanyonga et al., 2025).

Another critical management strength is the scalability of NLP approaches. Models developed and proven can be used in many projects, facilities, or various units of the organization with little or no extra investment. This scalability makes enterprise-wide risk intelligence possible that would otherwise be impractical using purely manual methods, and makes it possible to bring about uniform safety standards and knowledge sharing to large, geographically distributed organizations. The cost of implementing trained models is also lowered through their transferability, and deployment schedules go down to organizations with strong ambitions to increase predictive capabilities in safety.

Theoretically, the findings will add to the increasing body of literature that shows that advanced analytics can greatly assist occupational safety (Kim et al., 2025; Robinson, 2019). The outcomes of the NLP used to construct safety reports may mean that such an application can be established in other high-risk sectors like manufacturing, mining, oil and gas, and healthcare, where incident narratives are regularly gathered but not yet fully exploited in predictions (Nanyonga et al., 2025; Wong et al., 2018). Similar variations in anomaly detection competencies on chemical process safety have been demonstrated by Song and Suh (2019) and indicate the wide-ranging industry application of narrative-based prediction models.

The organizational readiness, data infrastructure, and change management are essential to be considered during the practical implementation. To facilitate the art of NLP analysis, organizations should ensure that safety reporting systems capture a reasonable amount of narrative information. Data governance systems should provide privacy to employees and allow data analysis of incident reports. Predictive models and training programs should assist safety professionals in interpreting and utilizing the outputs of predictive models correctly in decision-making. Liu and Yang (2022) underscore the need to create knowledge management mechanisms with the ability to incorporate NLP-generated insights into operational decision-making processes that guarantee predictive intelligence is turned into actual safety gains.

CONCLUSIONS

Natural Language Processing that is coupled with the Random Forest classifier comes in extremely handy in converting unstructured safety stories into risk intelligence action. This work suggests that a predictive analytics successfully implemented takes the narrative content, which is traditionally not widely involved in the practice of safety management, to infer predictive signals (Ricketts et al., 2023; Sankarasubramanian & Ganesh, 2020). The model posits accuracy (79.3%), and the absorption curve (AUROC) of the model is 0.98, which demonstrates that injuries could be predicted with the help of machine learning by structuring past data on injuries (Alkaissy et al., 2023). Such scores are out of the ordinary limits in the classification tasks in a real-life scenario and prove that it is technically viable to apply such systems in a working environment.

Consequences to the organizations are massive. This is the way under the help of which the safety managers may be able to transform the incident analysis from reactive to a proactive risk forecast so that they can implement interventions at an earlier stage and spend resources in a more efficient manner. The project managers can also do a predictive risk assessment of their projects and apply it in their project planning and implementation processes before workers face the risks. The top management can employ aggregate risk predictions in order to undertake strategic investments in safety and demonstrate measurable returns on investments in safety initiatives.

RECOMMENDATIONS

As the research findings have shown, it is possible to make some recommendations to the organizations interested in implementing the NLP-based predictive safety systems. First of all, companies should concentrate on the creation of a high-quality and comprehensive set of standards of narrative reporting. The predictive power of NLP models is premised on the fact that the richness and consistency of text data that undergo pipes are crucial determinants. The models will become more efficient and valuable as analytical tools through personnel training on the critical documentation of both models of accident mechanisms and circumstances.

Second, it is recommended that companies should invest in data infrastructure and analytics that would facilitate them to process safety narratives in a systematic manner. This includes the development of repositories of instance reports that are centralized, as well as the implementation of common information types and the development of mechanisms for model training implementation and confirmation. Third, firms should view NLP introduction as an addition to the professionalism of humans, but not a replacement for it. A combination of machine learning scalability and consistency, and the contextual understanding and ethical judgment of seasoned safety experts would compose the best safety management systems.

Fourth, companies are encouraged to work out systems of governance that would aggregate and protect ethical use of predictive models, the privacy of the people who work in them, so that they can not be utilized to penalize their workers, and should be transparent on how forecasts are utilized to guide decisions. Protecting workers should always be the purpose of any predictive analytics and not the creation of surveillance in which one should not report or be an accuser of the happenings.

FURTHER STUDY

Future research should focus on strengthening the role of Natural Language Processing (NLP) in workplace safety analysis. One key recommendation is the development of industry-based lexicons and specialized vocabularies that reflect the terminology used in specific sectors. Incorporating these domain-specific terms and taxonomies can improve model accuracy and help systems better interpret safety reports. Researchers are also encouraged to expand predictive capabilities so models can estimate injury severity, identify lost work time, and assess the possibility of recovery after incidents.Another important direction is temporal sequence analysis. Workplace accidents usually develop through a chain of events rather than occurring instantly. By analyzing the time patterns described in incident reports, researchers may identify critical intervention points where accidents could have been prevented. Understanding these dynamics can support better risk management and safety planning.

Transfer learning is also a promising approach. Models trained using data from one industry, such as construction, could potentially be adapted to other sectors like manufacturing or healthcare with limited additional training data. This would accelerate the adoption of NLP-based safety analytics across industries.Finally, longitudinal research is needed to evaluate the real-world impact of these predictive systems. Future studies should demonstrate whether their implementation can measurably reduce workplace injuries, providing stronger evidence of their effectiveness and value.

REFERENCES

Alkaissy, M., Arashpour, M., Golafshani, E. M., Hosseini, M. R., Khanmohammadi, S., Bai, Y., & Feng, H. (2023). Enhancing construction safety: Machine learning-based classification of injury types. Safety science, 162, 106102. https://doi.org/10.1016/j.ssci.2023.106102

Baker, H., Hallowell, M. R., & Tixier, A. J. P. (2020). Automatically learning construction injury precursors from text. Automation in Construction, 118, 103145. https://doi.org/10.1016/j.autcon.2020.103145

Kim, K. N., Cho, D. G., & Lee, M. J. (2025). A Machine Learning Approach for Factor Analysis and Scenario-Based Prediction of Construction Accidents. Buildings, 15(23), 4343. https://doi.org/10.3390/buildings15234343

Kuhn, K. D. (2018). Using structural topic modeling to identify latent topics and trends in aviation incident reports. Transportation Research Part C: Emerging Technologies, 87, 105-122. https://doi.org/10.1016/j.trc.2017.12.018

Liu, C., & Yang, S. (2022). Using text mining to establish a knowledge graph from accident/incident reports in risk assessment. Expert Systems with Applications, 207, 117991. https://doi.org/10.1016/j.eswa.2022.117991

Nanyonga, A., Joiner, K., Turhan, U., & Wild, G. (2025). Applications of natural language processing in aviation safety: A review and qualitative analysis. In AIAA SciTech 2025 Forum (p. 2153). https://doi.org/10.2514/6.2025-2153

Papazoglou, E., Nena, E., Kontogiorgis, C., Deligiannidou, E., Tripsianis, G., & Konstantinidis, T. (2025). Epidemiological Investigation of Occupational Accidents of Insured Salaried Employees in the Region of Thrace, Greece. OALib, 12(03), 1–20. https://doi.org/10.4236/oalib.1113080

Ricketts, J., Barry, D., Guo, W., & Pelham, J. (2023). A scoping literature review of natural language processing applications to safety occurrence reports. Safety, 9(2), 22. https://doi.org/10.3390/safety9020022

Robinson, S. D. (2019). Temporal topic modeling applied to aviation safety reports: A subject matter expert review. Safety science, 116, 275-286. https://doi.org/10.1016/j.ssci.2019.03.014

Sankarasubramanian, P., & Ganesh, E. N. (2020). Industrial Accident Report Analysis Using Natural Language Processing. International Journal of Scientific & Technology Research, 9(6), 470–476. https://www.ijstr.org/final-print/jun2020/Industrial-Accident-Report-Analysis-Using-Natural-Language-Processing.pdf

Song, B., & Suh, Y. (2019). Narrative text-based anomaly detection using accident report documents: The case of chemical process safety. Journal of loss prevention in the process industries, 57, 47-54. https://doi.org/10.1016/j.jlp.2018.08.010

Wong, A., Plasek, J. M., Montecalvo, S. P., & Zhou, L. (2018). Natural language processing and its implications for the future of medication safety: a narrative review of recent advances and challenges. Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy, 38(8), 822-841. https://doi.org/10.1002/phar.2151