International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0230
e ISSN: 2584-2854
Volume: 03
Issue: 04 April 2025
Page No: 1416 - 1420

Construction Site Incident Reporting Automation Using Natural Language Processing

Aji Antony[1], *Annie Sonia Xavier*[2]
[1]*Pg Student, Civil Engineering, KTU University, Toc H Institute of Science and Technology, Ernakulam, Kerala, India.*
[2]*Assistant Professor, Civil Engineering, Toc H Institute of Science and Technology, Ernakulam, Kerala, India.*
Email ID: *ajiantony1309@gmail.com*[1]

Abstract
In the construction sector, effective communication about incidents is important for improving safety policies and procedures. Lately, technological advancements in construction have made automating document preparation within an information system possible. However, it requires a great deal of time and effort for a human operator to prepare such documents manually, which in most cases leads to inaccuracies. This paper proposes a new design and development of construction site incident reporting systems based on the automation of event reporting using Natural Language Processing (NLP). Its purpose is to increase reporting productivity, accuracy, and the level of compliance with relevant statutes and regulations. Worksite mishaps continue to challenge both human welfare and economic resilience. Reports from global organizations indicate an upward trend in mishap frequency and critical consequences. Traditional techniques depend heavily on handwritten entries, causing postponements and obstructing immediate evaluations. This research introduces a smart reporting mechanism utilizing NLP, aimed at simplifying the notification workflow. The platform incorporates voice-to-text conversion to interpret verbal inputs from personnel, followed by NLP-driven contextual interpretation and essential data extraction. The background study reveals significant capabilities of NLP in advancing safety strategies, pointing to strengths like better reliability and shortened notification durations. The system includes information acquisition, machine learning for audio processing, and NLP-based detail identification. Predefined digital formats support fast and consistent reporting, reinforcing job site precautionary measures and preventing recurrence of hazardous events. By boosting the speed and reliability of issue communication, the model strives to refine safety procedures and assist in thorough incident analysis.

Keywords: *Automation, Construction Safety, Incident Reporting, Natural Language Processing, Speech Recognition.*

1. Introduction
Construction sites are inherently hazardous environments where ensuring the safety of workers is paramount. According to global reports such as those from the International Labour Organization (ILO), the number of construction-related accidents and near-miss incidents is on the rise. Traditional methods of incident reporting, which rely heavily on manual documentation through written or typed entries, often lead to delays, incomplete data, and human errors. These shortcomings hinder timely investigations and limit the ability to take corrective actions, ultimately increasing the risk of recurring incidents. Therefore, there is a critical need for a real-time, efficient, and accurate reporting system that can support proactive safety management. This study proposes the development of an automated incident reporting system utilizing Natural Language Processing (NLP). By incorporating speech

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0230
e ISSN: 2584-2854
Volume: 03
Issue: 04 April 2025
Page No: 1416 - 1420

recognition, the system enables workers to verbally report incidents, allowing faster and more convenient data collection on-site. NLP algorithms analyze these inputs in real time, extracting relevant details such as the type of incident, location, severity, and possible causes. The system comprises voice-to-text conversion, contextual analysis using NLP, and automated generation of structured reports. Literature supports the integration of NLP in safety management due to its potential to reduce reporting time, improve accuracy, and enhance documentation quality. The primary objective of this research is to design and evaluate an NLP-based framework that streamlines incident reporting and contributes to safer construction practices [1].

1.1 Incident Report in Construction Site

Incident reports are essential for documenting accidents, injuries, near-misses, and hazardous conditions that occur on construction sites. They play a crucial role in identifying root causes and developing preventive strategies to enhance workplace safety. Reports that are accurate, comprehensive and on time aid in the effectiveness of investigations, help fulfill legal safety requirements, and serve a purpose for the audit data. Indeed, reporting in these areas usually is done manually and is not automated, which makes it error-prone, time-consuming, and inefficient. In return, incomplete and inaccurate reporting results quite unfortunate. The implementation of modern technologies like Natural Language Processing (NLP) and speech recognition has the potential to address the issues languid reporting methods pose, automating reporting, therein providing remarkable improvements in efficiency and accuracy, therefore enhancing the quality of safety documentation.

1.2 NLP

The artificial intelligence sub-field of Natural Language Processing enables computers to interact with human language through its study of computer and human linguistic interactions. Through NLP machines acquire the ability to process human language while they understand and produce meaning that humans can comprehend. Computational linguistics teams up with machine learning approaches to analyze and process extensive amounts of unstructured text data as well as spoken materials. The real-world operationalization of NLP systems enables language translation as well as sentiment analysis and production of chatbots and speech recognition systems. Construction sites benefit from NLP utility through automated document creation while also using the technology to review incident reports together with better onsite communication systems.

2. Literature Review

Research on construction site safety indicates a rising need for effective and dependable incident reporting systems. Numerous studies have demonstrated that safety management relies heavily on the availability and accuracy of incident reporting, often citing noted gaps in reporting deadline adherence. Standard manual approaches tend to delay reporting, which then become inconsistent and inaccurate [2][5]. This makes it harder to respond to emergencies or improve safety in the future. Solutions that are automatic tend to improve consistency of data and lessen administrative procedures. One of these solutions is the use of Natural Language Processing, which is now seen as an industry standard for automating extracting and analyzing incident details. Systems of this nature can deal with actual data which include incident descriptions. Furthermore, Natural Language Processing (NLP) allows the automation of the initial step of analyzing unstructured data like incident descriptions by turning spoken or written words into organized data that can be easily understood by computers. Integrating NLP with speech recognition technology enables workers to verbally report incidents, bayoneting the speed and accuracy of reporting. Automated systems enhance the quality of incident data and boost organizational safety compliance. The Occupational Safety and Health Association, officially known as OSHA, states that the deadline for submitting a workplace fatality is capped at eight hours, while other serious injuries at twenty-four hours. These rules highlight the need for the reporting of incidents. NLP-based systems for automated reporting align to OSHA's objectives by enabling quicker, sophisticated, and ample safety documentation across construction sites.

Note: OSHA (Occupational Safety and Health

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0230
e ISSN: 2584-2854
Volume: 03
Issue: 04 April 2025
Page No: 1416 - 1420

Administration) has mandated reporting a death within eight hours and severe injuries like hospitalization, amputation, or eye removal within twenty-four hours. Considering the literature review and OSHA guidelines, we prepared a sample training dataset to train and evaluate the proposed NLP (Natural Language Processing)-based incident reporting system. The sample dataset is based on raw construction worker reports that serve as natural language text inputs for an NLP system. Each construction worker report is associated with specific incident types, injury severity levels, OSHA reportability (considering regulatory compliance thresholds), and the pertinent words or named entities for keyword extraction that will be served by NER are also noted. This specific format enables training of text classifiers, dismantle severity analyzers, and compliance classifiers [3].

Table 1 Sample Training Dataset for NLP-Based Construction Incident Reporting

Raw Text Input	Incident Type	Injury Severity	OSHA Reportable	Entites Extracted
Worker fell from the second-floor scaffold and broke his leg.	Fall From height	Serious	yes	Fall, Scaffold, second floor, leg injury
Minor electric shock occurred while fixing wires in the basement. No injury reported.	Electric shock	Minor	NO	Electric shock, basement, no injury, wires
Crane malfunctioned during lifting and dropped load near workers, no one was hurt	Equipment malfunction	Near miss	NO	Crane, malfunction, lifting, dropped load
A falling metal pipe struck the supervisor's shoulder, causing a deep wound.	Struck by object	severe	yes	Metal pipe, falling, object, deep wound, shoulder
Worker collapsed due to heatstroke while working under the sun and was taken to the hospital.	Medical emergency	Severe	yes	Heat stroke, collapse, sun exposure, hospital

3. Methodology

The research uses speech recognition alongside Natural Language Processing (NLP) to establish an automatic real-time reporting solution for construction sites incidents. The researchers started with this topic because the construction industry requires better safety reporting systems [4]. The investigation studied manual reporting techniques along with AI and NLP's possible applications for safety management through extensive literature research. A collection of data suitable for training and evaluation came from site observations along with literature research OSHA standards and questionnaire responses. The system designers developed an interface with python language which enables workers to report incidents through voice commands for practical and easy on-site operation. A speech recognition model provided by OpenAI through its Whisper application transcribed incoming voice inputs into precise written text. The examined text underwent evaluation through an NLP model (GPT-3.5 Turbo) that extracted essential information including the incident type alongside severity level and cause and place of occurrence. Alternatively, processed information served to create standardized incident reports that fulfill regulatory requirements.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0230
e ISSN: 2584-2854
Volume: 03
Issue: 04 April 2025
Page No: 1416 - 1420

Through its implementation the system offers improved incident documentation with faster response times and more accurate and consistent results that promote better safety practices in construction sites. Figure 1 shows Process of Methodology.

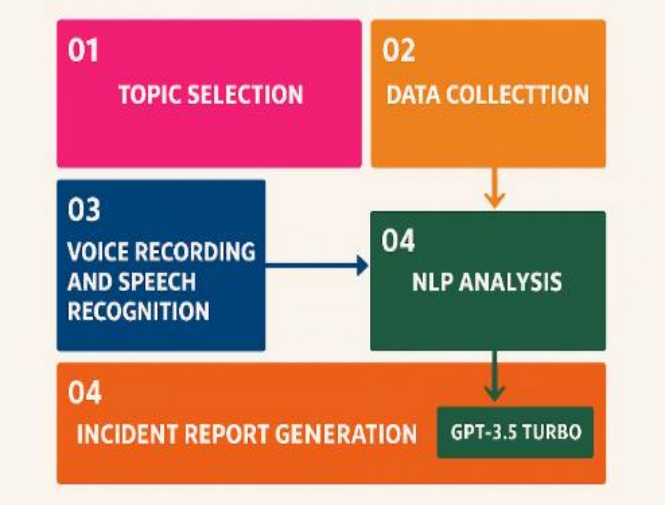

Figure 1 Process of Methodology

4. Results and Discussion

4.1 Results

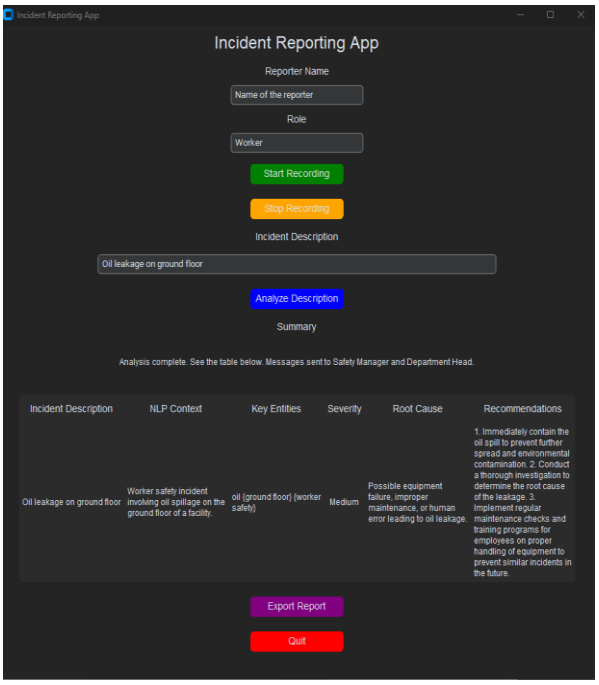

Figure 2 Incident Reporting User Interface with Sample Incident Reporting

The Incident Reporting App captures and processes safety related incidents on construction sites through manual and voice methods. The app employs NLP algorithms to decode the description after it has been entered or recorded, pinpointing its context, key entities, severity level, root cause, and other relevant recommendations. In the interface example given, the précis of “Oil leakage on ground floor” was flagged as an incident which was correctly diagnosed as medium severity and potential causes of equipment malfunction or human error. Figure 2 shows Incident Reporting User Interface with Sample Incident Reporting. The app provides feedback in real time and issues alerts to relevant personnel from the user’s point of view. Moreover, the application incident report is automatically structured in CSV (Excel) format as shown in Figure 3. Reported files contain basic reporter details, role, time, document description, NLP generated context with key components, severity, root cause, and suggested actions, and these files are automatically generated every time an incident is reported.

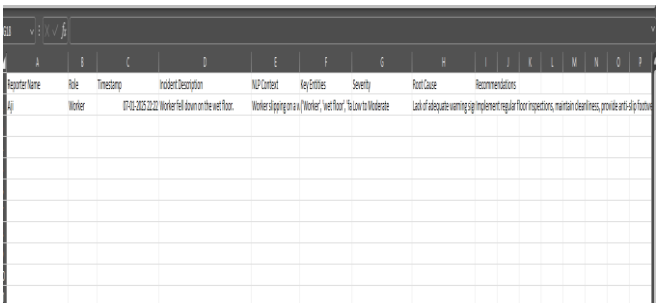

Figure 3 Sample Csv File Out Put

4.2 Discussion

In corporate environments, the addition of NLP for machine learning has been helpful when integrated into the incident reporting workflow. Automating the process of context and severity classification in the construction zone ensures that a manual documentation mistake or delay does not slow down the process. An accurate delineation of the primary causes along with supporting targeted suggestions helps site managers to proactively intervene if corrective action needs to be taken. In addition, systematic compliance with constituent reporting

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0230
e ISSN: 2584-2854
Volume: 03
Issue: 04 April 2025
Page No: 1416 - 1420

requirements is made faster by templated CSV files. This engages construction sites to switch from a reactive approach based on paperwork to reliance on proactive data driven safety management, which cultivates a prevention-centric attitude.

Conclusion

The creation of the NLP-based Incident Reporting App has showcased what is possible with the integration of AI technologies into the construction site safety management system. The system is capable of automating the incident description analysis and reporting processes which increases the speed and accuracy of the incident reporting as well as its reliability. The use of NLP technology allows for the insightful extraction of severity levels, key entities, root causes, and even provides constructive suggestions. Moreover, the Excel report generation feature enables the structured storage of incident data making it easier for advanced analytical processes and informed decisions in the future. The system fosters enhanced safety monitoring and improved response times while enabling construction sites to adopt the data-centric approach into their safety management systems.

References

[1]. Alireza Shamshiri, Kyeong Rok Ryu, June Young Park. (2024). Text mining and natural language processing in construction. Automation in Construction, 118, 105200.

[2]. Chengke Wu, Xiao Li, Yuanjun Guo, Jun Wang, & Zengle Ren. (2022). Natural language processing for smart construction: Current status and future directions. Automation in Construction, 134, 104059.

[3]. Fan Zhang, Hasan Fleyeh, Xinru Wang, & Minghui Lu. (2024). Construction site accident analysis using text mining and natural language processing techniques. Automation in Construction, 238-248.

[4]. Guntaka Akanshha, A. Vijay Kumar, G. Swapna, & Ganji Nikitha. (2023). Construction site accident analysis using text mining and natural language processing techniques. Industrial Engineering Journal, 52.

[5]. Jadhav, T., & Deshpande, R. (2023). Integrating Artificial Intelligence in Project Manager's Decision-Making Process: A Look at Built Environment Projects. PMI India Research & Academic Virtual Conference.