EPiC Series in Built Environment
Volume 7, 2026, Pages 970–978
Proceedings of Associated Schools of Construction 62nd Annual International Conference

A Deep Learning-Driven System for Real-Time Detection, Identification, and Documentation of Unsafe Behaviors in Construction: Case Study

Mohammad Dabash[1], Dyala Aljagoub[1], Edgar Small[1], Ri Na[1], and Abdulaziz Banawi[1]
[1]University of Delaware

This pilot study presents an integrated deep-learning framework that not only detects but also identifies and records noncompliance with personal protective equipment (PPE) in real time on construction sites. The framework utilizes an object detection model, achieving a mean average precision (mAP) of 93%, and employs a rolling average supported by a real-time object tracking algorithm to minimize false positives in complex site environments. Detected violations trigger a deep learning facial recognition model that identifies the individual. All relevant data, including time of occurrence, nature of violation, and identity, are then stored in a Structured Query Language (SQL) database for subsequent analysis. This system addresses a critical research gap by going beyond detection to create a comprehensive record of unsafe behaviors, thereby enabling targeted interventions and data-driven safety enhancements. Despite its promising results, limitations such as occlusions and a relatively small dataset remain, suggesting that future work should incorporate larger, more diverse datasets to further refine and validate the approach.

Keywords: Deep Learning, Construction Safety, Real-Time Object Detection, Facial Recognition

Introduction

This study demonstrates a feasible, integrated framework for automated detection and documentation of unsafe behaviors on construction sites. The construction industry has long struggled with persistently high rates of work-related injuries and fatalities. In 2023, the Occupational Health and Safety Administration (OSHA) in the U.S. reported 1,075 work-related fatalities; roughly 20% of all reported fatalities were construction-related (Occupational Safety and Health Administration, 2025). Construction was cited for having the highest fatality rate in the private sector, dating all the way back to 2011 (U.S. Department of Labor, Bureau of Labor Statistics, 2025). Current measures, including on-site risk assessments and adequate controls, are implemented but prove insufficient (Balakreshnan et al., 2020).

Building on the strengths of three deep learning models (YOLOv4 for object detection, Deep Simple Online and Realtime Tracking (SORT) for robust tracking, and a facial recognition library), this approach automatically identifies and reports PPE noncompliance incidents (Wojke et al., 2017). By combining detection, tracking, and identification in a single system, safety managers can quickly pinpoint and address hazardous behaviors. This supports faster interventions, improved compliance, and a more secure work environment. The automated logging of violations, including the individual's identity, nature of the noncompliance, and timestamp, enables targeted safety measures implementation and data-driven decisions, ultimately reducing the risk of serious accidents and bolstering overall construction site safety.

Literature Review

Deep Learning in Construction Safety

When it comes to deep learning applications in construction safety, studies have varied in detection targets, including behavior, operations, and activities. For instance, one study intended to predict worker sleep deprivation on-site. The study reported high sleep deprivation prediction accuracy, which has been linked to increasing accidents (Sathvik et al., 2024). Other applications have included work and heavy machinery detection to automate site safety monitoring (Chi & Caldas, 2011). Another study explored deep learning models for injury prediction through historical reports, which aimed to aid practitioners in developing activity-specific safety interventions (Tixier et al., 2016). Fang et al. explored using deep learning for harness detection when working at heights to prevent falls (Fang et al., 2018b). These studies illustrate the range of deep-learning applications in construction safety, with additional opportunities for further research. Mei et al. developed a robust real-time object detection model capable of identifying unsafe behaviors on site and providing recommended procedures to support site safety and management (Mei et al., 2024).

Moreover, several studies have implemented deep learning approaches to detect PPE compliance on construction sites using various models (Ahmed et al., 2023; Al-Azani et al., 2024). These studies reported promising detection accuracy, proving the potential of deep learning applications to improve construction safety. However, some limitations remain, such as identifying non-compliant individuals and object tracking leading to reduced accuracy due to occlusions and abrupt motion (Fang et al., 2018a; Liu et al., 2021).

Study Purpose and Scope

While deep learning (DL) has demonstrated its effectiveness in detecting unsafe behaviors on construction sites, its full potential is not realized if detections are not properly recorded (Fang et al., 2018a). A major challenge hindering the adoption of such systems in the industry is the lack of integration between detection models and actionable reporting mechanisms. This study overcomes this limitation by integrating a PPE detection model with a facial recognition system, enabling automated identification and documentation of unsafe behaviors. Other reported limitations affecting detection were due to abrupt motion or occlusions (Liu et al., 2021), common occurrences on construction sites. Furthermore, some studies reported issues with decreased robustness where a single frame was used for detection (Karlsson et al., 2022). Therefore, this study employed an object-tracking approach to address these limitations and improve detection accuracy and robustness. The proposed approach ensures that detected unsafe behaviors are systematically and accurately recorded, allowing safety managers to take corrective actions and enhance overall site safety.

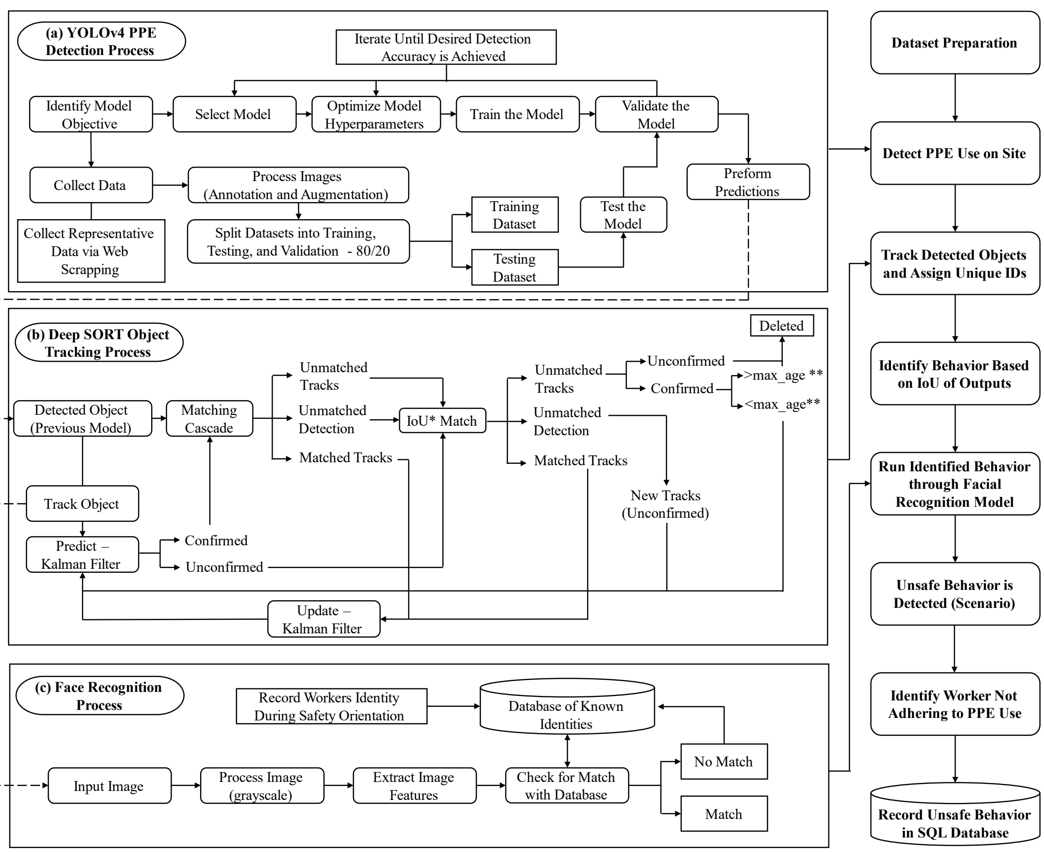
Methodology

Case Study Overview

Current safety management approaches often rely on post-fact interventions, which can be ineffective in delivering desired safety improvements. Therefore, this feasibility-oriented case study evaluates an integrated workflow for real-time PPE noncompliance management: (1) PPE detection, (2) object-tracking to improve detection robustness across frames, and (3) incident documentation by associating a worker ID and logging events in an SQL database. When an unsafe behavior is identified, the system logs detection details, which are then processed through a facial recognition model. Once the worker is identified, a record is generated, including the type of unsafe behavior, the time of occurrence, the detecting camera, and the incident location. By uniting these components, the framework not only detects unsafe behaviors but also systematically records them, enabling proactive safety interventions and contributing to a safer construction site environment.

Model Selection and Justification

The models selected to achieve the study's intended goal were: (1) YOLOv4 for object detection, chosen for its speed and accuracy (Protik, Rafi, & Siddique, 2021; Bochkovskiy, Wang, & Liao, 2020). (2) Deep SORT for tracking, which assigns unique IDs to maintain identity continuity across frames (Wojke et al., 2017). (3) face_recognition Library (Geitgey, 2025) with reported 99.38% benchmark accuracy to identify non-compliant individuals, matched against a pre-registered worker database. The system assumes a fixed camera view, sufficient visibility of PPE items for detection, and sufficient face visibility for identity association when violations are logged.



*IoU: Intersection over Union (IoU = Area of overlap/Area of Union)
** max_age: Number of frames before a track is terminated (adjustable)

Figure 1. Case study overview: (a) is the YOLOv4 model, (b) is the Deep SORT model, and (c) is the Facial Recognition model

Notably, YOLOv4 was selected to support real-time feasibility testing and because it provided acceptable accuracy at the time the system was implemented. The focus of this case study is end-to-end integration (detection–identification–recording), not detector benchmarking across architectures. Future iterations of the study intend to evaluate detection improvements using novel and recent models. The proposed approach, with details of each model's process, is highlighted in Figure 1.

Dataset Preparation

A total of 2,500 images were initially scraped from the web using Python's Selenium library. After manually filtering and cleaning (removing irrelevant images, watermarks, and disturbances), 1,600 images remained for training. The selected images were resized to 1920×1080 pixels. Data augmentation techniques, including flipping and rotation, were applied to increase variability and prevent overfitting, resulting in a final dataset of 2,200 images. Annotation was performed with the open-source Computer Vision Annotation Tool (CVAT), using bounding boxes for Person, Hardhat, and Vest labels. Finally, the dataset was divided into training (80%) and testing (20%) subsets to develop and validate the object detection models. The collected images were utilized to train the detection model (YOLOv4). A video from an active construction site was recorded to test the proposed deep learning integrated workflow.

The rationale behind employing web-scraped images for training was to enable rapid development of a labeled dataset with diverse PPE appearances and scene conditions for this pilot feasibility study. While this supports initial model training, site-specific conditions may differ. Thus, future works intend to incorporate larger field-collected datasets to strengthen generalizability. Note that this pilot study does not aim to develop or benchmark a highly generalized PPE-detection model, but to examine the feasibility of an integrated detection–tracking–identification–logging workflow under realistic site conditions. Accordingly, the dataset was considered sufficient for system-level feasibility testing, while comprehensive generalization evaluation is reserved for future work. The selected site video includes typical challenges such as occlusion, motion blur, and partial visibility, providing a conservative test case for evaluating workflow behavior.

YOLOv4 Model Application

As a single-stage detector, YOLOv4 supports real-time detection, making it appropriate for the feasibility focus of this study. The process of developing the YOLOv4 model is illustrated in Figure 1(a). Hyperparameters were tuned heuristically for optimal performance, and transfer learning was employed using pre-trained YOLOv4 weights, an effective strategy given the relatively small dataset. The training process, running on Google Colab Pro with an Nvidia Tesla P100 GPU (16 GB RAM), took 600 minutes to complete.

Deep SORT Application

Some challenges were encountered when testing the developed model on the recorded video feed, particularly false positives (misclassifying safe behavior as unsafe) caused by abrupt motion and occlusions. YOLOv4 provides frame-level PPE detections that are evaluated using a rolling average process, supported by Deep SORT (Figure 1(b)) to maintain consistent worker IDs across frames. As illustrated in Figure 2, PPE status is evaluated using a rolling average over consecutive frames. The model may briefly flag missing PPE in isolated intervals, but violations are logged only when unsafe detections persist beyond a preset threshold, reducing misclassifications. This is important in practice because frequent false alarms can reduce user trust and adoption. When a violation is confirmed, the system saves a snapshot of the worker's bounding box region for subsequent facial recognition.

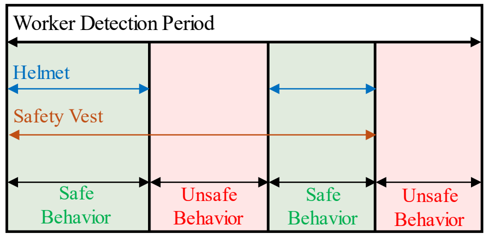


Figure 2. Rolling average worker detection process

Facial Recognition Application

Facial recognition is initiated only after a PPE violation is confirmed to associate the incident with a registered worker ID for documentation and targeted safety intervention. Faces are detected under varying orientations and lighting. Then, the unique facial features are extracted and matched against a database of known faces to establish identity (Figure 1(c)), employing the Python "Face_recognition" library (Geitgey, 2025).

For this study, images of workers' faces were sourced via web scraping, labeled, and assigned random IDs to protect individual anonymity. The library requires only three to five images per person, making it practical for construction environments where workers' faces can be captured during safety orientation. Nevertheless, strict data privacy measures are essential, including secure storage, limited distribution, and prompt data destruction once the safety objectives have been met.

Ethical, Privacy, and Legal Considerations

While the proposed detection-identification-recording framework is intended to enable targeted interventions that improve worker safety, the use of facial recognition raises ethical concerns. In particular, identity association can introduce privacy risks on construction sites that must be addressed prior to adoption.

First, informed worker consent should be obtained, supported by clear communication about what data are collected, how they are used, and why. Second, secondary uses unrelated to safety should be explicitly prohibited (e.g., productivity monitoring or disciplinary surveillance). Third, identity association should be purpose-limited and triggered only when a PPE violation is detected. In addition, robust data protection controls are required. Recommended safeguards include restricting access to authorized safety personnel, encrypting stored data, and enforcing minimal data retention. Violation records should be retained only as long as needed for safety review and corrective action and then securely deleted.

Legal requirements vary by jurisdiction and organizational context. Deployment should be reviewed against applicable workplace privacy and biometric-data regulations. These safeguards help preserve worker trust while enabling the intended safety benefits of the proposed framework.

Results and Discussion

YOLOv4 Model Performance

Deep learning models are typically evaluated using precision, recall, mean average precision (mAP), and the F1 score. Each metric captures a different performance aspect, providing a comprehensive overview of how well the model meets its intended purpose. Specifically, precision assesses the accuracy of the model's predictions, while recall measures the model's ability to correctly identify all relevant objects. The F1-score, the mean of precision and recall, balances both measures to provide a single metric for evaluating overall performance. Finally, mAP summarizes performance across multiple classes and thresholds, offering an overall view of detection quality.

In this study, object-detection performance metrics (precision, recall, and mAP) are reported to establish that the detector operates at a stability level sufficient to support downstream tracking and identification, rather than to claim state-of-the-art detection performance.

Figure 3(a) presents the training results of the model. The two main factors to consider from the figure are the training loss and the changes in the mAP over the training duration. The training loss dropped drastically initially and converged at around the 4000th iteration. From that point onward, changes in training loss were less drastic; this is attributed to the varying learning rate used to train the model and is typical for object detection training results attained by researchers in literature. Furthermore, the decreasing loss indicates a reduced possibility of overfitting. The mAP value stabilized at a much earlier point and fluctuated at around 92% for the rest of the duration of the training. Table 1 details model performance metrics and results. While the mAP calculated the average precision over all the categories, it is crucial to consider the model's performance over the three categories on which the model has been trained. The model precision and recall results are satisfactory; thus, the model is considered fit for the purpose of this study. Figure 3(b) presents the output of the detection model.

Table 1. YOLOv4 Model Performance

Parameter	Precision	Recall	F1-Score	mAP
Person	0.94	0.92	0.93	
Helmet	0.94	0.93	0.93	0.93
Vest	0.90	0.88	0.89	

(a) (b)

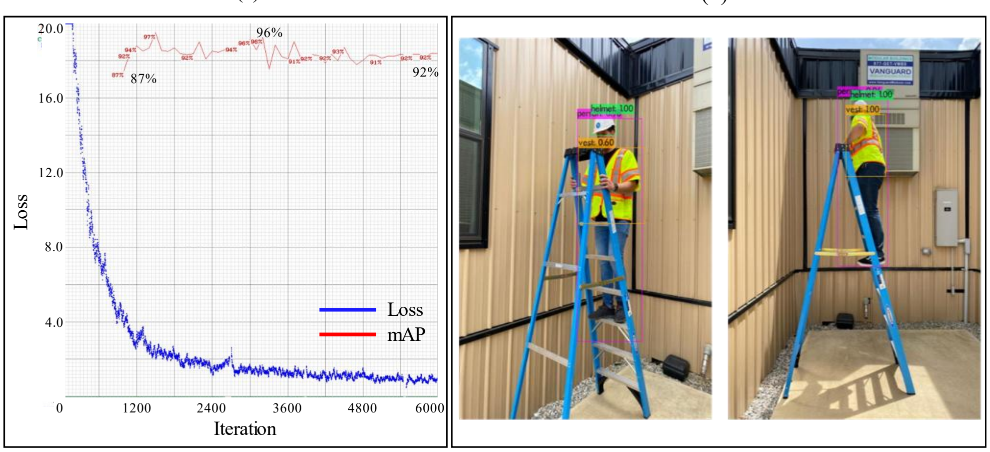


Figure 3. YOLOv4 PPE detection model: (a) mAP and training loss curves; (b) sample output

Deep SORT Model Performance

As previously noted, a rolling average approach (Figure 2) was utilized where Deep SORT tracked worker IDs over a specific duration in a live video feed to determine whether an individual complied with PPE requirements on site. The rolling average process is intended to limit false alarms by allowing the model to track the object over several frames to minimize the effects of abrupt motions and occlusions, which are common limitations of real-time detection. As presented in Figure 4, the person on the scene is complying with the PPE requirements; however, in this instance, the model detected the person and the helmet but failed to detect the vest, which caused it to flag the detection as unsafe behavior while it was not. Notable in the next frame presented at the right of Figure 4, the model detected both the helmet and vest and flagged the interaction as safe. Deep SORT enabled the rolling average process by tracking an object throughout the live feed, allowing the object detection model to observe the object of interest over several frames, improving model performance and accuracy.

FP – Unsafe Behavior | TN – Safe Behavior | FP – Unsafe Behavior | TN – Safe Behavior

Figure 4. Sample model false positive prediction during a live video feed test

Facial Recognition Model Performance

The model crops the bounding boxes of the person conducting unsafe behavior and saves the output in the “unknown” faces folder. Once the facial recognition model identifies the person’s identity, the unsafe behavior is recorded, and a new entry is added to the SQL database. While the detection model performed well, an explicit limitation became clear when running the facial recognition model. The model is very sensitive to the person’s orientation in the scene. In some instances, the facial recognition model could not identify the individual due to the workers not facing the camera or having a significant portion of their faces covered. Once the facial model successfully detected unsafe behavior, it exported the worker ID, timestamp, unsafe behavior category, identified unsafe behavior, and a link to the cropped image. Table 2 presents a sample of the developed SQL database.

Table 2. Sample of unsafe behavior SQL database

Employee ID	Category	Unsafe Behavior	Time
1010	PPE	Helmet = False	2022-07-20 12:45:56
1012	PPE	Vest = False	2022-04-14 10:34:42
1015	PPE	Helmet = False, Vest = False	2022-05-17 14:17:32

Conclusion

This study addresses a key gap in construction safety, recording and identifying PPE noncompliance, by integrating three components: a YOLOv4-based object detection model (93% mean average precision), Deep SORT tracking to reduce false positives, and facial recognition to identify individual violators. The result was an integrated detection-identification-recording deep learning framework that can be adopted on construction sites. The proposed model support faster interventions when unsafe behaviors are identified, improving overall site safety, a major concern in the construction industry. While the approach effectively flags unsafe behaviors and generates an actionable safety record, it raises ethical concerns. Recommendations include strict data protection measures, minimal data retention, and secure disposal.

Because the proposed framework operates as a multi-stage pipeline, the focus of evaluation in this study is on functional integration and operational feasibility rather than exhaustive statistical validation of each individual component. Beyond empirical results, this study contributes a reproducible system architecture and decision pipeline for integrating detection, tracking, identification, and logging, which can be adapted or extended in future construction safety implementations. While larger datasets and multi-site testing are necessary for quantifying generalization and bias, such analyses fall outside the scope of this feasibility-oriented case study.

As a result of this multi-stage detection-identification-recording pipeline, end-to-end reliability is influenced by error propagation across stages. In particular, construction-site conditions such as occlusion, abrupt motion, and partial PPE or face visibility can lead to short-lived false flags or prevent attribution of a violation to a specific worker ID. When facial visibility permits, the event is linked to a worker ID and logged in the SQL database (with timestamp and violation type), enabling follow-up actions and longer-term trend tracking.

While the pilot study achieved its intended goal of evaluating a combined detection, tracking, and identification deep learning model, some limitations remain. First, the model suffered from a small dataset posing a challenge in building a comprehensive and robust model, partly mitigated through augmentation. Also, testing was limited to a single site video, which limits generalizability and highlights the need for larger datasets in future studies. Finally, the technical requirements for model setup and operation may pose barriers, suggesting future work will include workflow-level false alarm rates, end-to-end processing latency, and multi-site validation. Such end-to-end metrics are more appropriately evaluated during pilot deployment or longitudinal field studies, rather than during initial feasibility validation.

References

Ahmed, M. I. B., et al. (2023). Personal protective equipment detection: A deep-learning-based sustainable approach. Sustainability, 15(18), Article 18. https://doi.org/10.3390/su151813990

Al-Azani, S., Luqman, H., Alfarraj, M., Sidig, A. A. I., Khan, A. H., & Al-Hamed, D. (2024). Real-time monitoring of personal protective equipment compliance in surveillance cameras. IEEE Access, 12, 121882–121895. https://doi.org/10.1109/ACCESS.2024.3451117

Balakreshnan, B., Richards, G., Nanda, G., Mao, H., Athinarayanan, R., & Zaccaria, J. (2020). PPE compliance detection using artificial intelligence in learning factories. Procedia Manufacturing, 45, 277–282. https://doi.org/10.1016/j.promfg.2020.04.017

Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020, April 22). YOLOv4: Optimal speed and accuracy of object detection [Preprint]. arXiv. http://arxiv.org/abs/2004.10934

Chi, S., & Caldas, C. H. (2011). Automated object identification using optical video cameras on construction sites. Computer-Aided Civil and Infrastructure Engineering, 26(5), 368–380. https://doi.org/10.1111/j.1467-8667.2010.00690.x

Fang, Q., et al. (2018). Detecting non-hardhat use by a deep learning method from far-field surveillance videos. Automation in Construction, 85, 1–9.

https://doi.org/10.1016/j.autcon.2017.09.018

Fang, W., Ding, L., Luo, H., & Love, P. E. D. (2018). Falls from heights: A computer vision-based approach for safety harness detection. Automation in Construction, 91, 53–61. https://doi.org/10.1016/j.autcon.2018.02.018

Geitgey, A. (2025). [Software]. https://face-recognition.readthedocs.io (accessed March 2025)

Karlsson, J., Strand, F., Bigun, J., Alonso-Fernandez, F., Hernandez-Diaz, K., & Nilsson, F. (2022, December 9). Visual detection of personal protective equipment and safety gear on industry workers [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2212.04794

Liu, W., Meng, Q., Li, Z., & Hu, X. (2021). Applications of computer vision in monitoring the unsafe behavior of construction workers: Current status and challenges. Buildings, 11(9), Article 9. https://doi.org/10.3390/buildings11090409

Mei, X., Xu, F., Zhang, Z., & Tao, Y. (2024). Unsafe behavior identification on construction sites by combining computer vision and knowledge graph–based reasoning. Engineering, Construction and Architectural Management, Advance online publication. https://doi.org/10.1108/ECAM-05-2024-0622

Occupational Safety and Health Administration. (2025, February 18). Commonly used statistics. U.S. Department of Labor. https://www.osha.gov/data/commonstats

Protik, A. A., Rafi, A. H., & Siddique, S. (2021, August). Real-time personal protective equipment (PPE) detection using YOLOv4 and TensorFlow. In 2021 IEEE Region 10 Symposium (TENSYMP) (pp. 1–6). https://doi.org/10.1109/TENSYMP52854.2021.9550808

Sathvik, S., Alsharef, A., Singh, A. K., Shah, M. A., & ShivaKumar, G. (2024). Enhancing construction safety: Predicting worker sleep deprivation using machine learning algorithms. Scientific Reports, 14(1), 15716. https://doi.org/10.1038/s41598-024-65568-2

Tixier, A. J.-P., Hallowell, M. R., Rajagopalan, B., & Bowman, D. (2016). Application of machine learning to construction injury prediction. Automation in Construction, 69, 102–114. https://doi.org/10.1016/j.autcon.2016.05.016

U.S. Department of Labor, Bureau of Labor Statistics. (2025, February 18). Census of fatal occupational injuries summary, 2023 (2023 A01 results). https://www.bls.gov/news.release/cfoi.nr0.htm

Wojke, N., Bewley, A., & Paulus, D. (2017, September). Simple online and realtime tracking with a deep association metric. In 2017 IEEE International Conference on Image Processing (ICIP) (pp.3645–3649). https://doi.org/10.1109/ICIP.2017.8296962