buildings

MDPI

Article

A Unified Computer Vision Framework for Safety Monitoring and Contractual Compliance Decision Support

Gasser Shawky *, Nabil Amer and Atef Ragab

Department of Building and Construction Management, College of Engineering and Technology, Arab Academy for Science, Technology and Maritime Transport (AASTMT), B 2401 Smart Village, Giza 12577, Egypt; nabil_amer63@yahoo.com (N.A.); atef5919@adj.aast.edu (A.R.)
* Correspondence: g.shawky30869@student.aast.edu

Abstract

Construction sites are highly dynamic and inherently hazardous environments where safety compliance is crucial yet challenging to enforce manually. Failing to wear essential personal protective equipment, such as helmets and safety vests, and unauthorized entry into restricted risk zones are major contributors to preventable injuries. This study introduces an intelligent automated safety monitoring system based on computer vision and deep learning. The proposed framework integrates the YOLOv8 object detection model with facial recognition to perform automated worker identification, real-time PPE compliance verification, and restricted zone monitoring. When a safety violation is detected, the system captures visual evidence, triggers localized alerts, and records metadata including worker identity, timestamp, location, and violation type. To support worker identification under varying environmental conditions, the framework employs a video-based enrollment process combined with attendance-based database filtering. The system was evaluated through a three-day field deployment at a real-world construction site in Egypt, demonstrating reliable performance under changing ambient conditions. The results demonstrate the feasibility of continuous automated safety monitoring under real-world construction-site conditions and show the framework's ability to detect, document, and associate safety violations with structured incident records. The proposed framework offers an AI-driven approach to support workplace safety monitoring and contractual compliance review.

Keywords: IOT in construction; construction management; safety management; safety control

Check for updates

Academic Editors: Chau Le, Tuyen Le and Rafiq Muhammad Choudhry

Received: 17 August 2026
Revised: 17 October 2026
Accepted: 27 October 2026
Published: 20 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

1.1. Background

The construction sector remains one of the most perilous industrial sectors worldwide, characterized by disproportionately high rates of occupational injuries and fatalities. A large proportion of construction-site accidents stem from human error, unsafe behavioral patterns, and systematic non-compliance with established site safety protocols. While personal protective equipment (PPE), particularly high-visibility reflective vests and safety helmets, serves as the primary line of defense against severe physical harm, maintaining consistent site-wide adherence is incredibly difficult. The dynamic, complex, and rapidly evolving landscape of a modern construction site significantly limits traditional safety-monitoring approaches, which rely on intermittent manual patrols by safety officers. Manual oversight is resource-intensive, prone to human fatigue, inherently subjective, and incapable of providing continuous, site-wide surveillance. Consequently, critical safety lapses frequently

pass unnoticed, exposing workers to unnecessary risks and contractors to substantial liability. Concurrently, recent advancements in artificial intelligence (AI) and deep learning have catalyzed a paradigm shift within industrial safety automation. Computer vision frameworks offer a viable mechanism to transform traditional video surveillance from passive recording tools into proactive, automated monitoring systems. By deploying deep-learning object detection architectures, these systems can continuously scan hazardous workspaces, accurately categorize compliance statuses, and instantly flag anomalies with minimal latency. Beyond improving workplace safety, construction projects require effective mechanisms to enforce contractual safety obligations. Most construction contracts include provisions that assign responsibility for safety violations and specify penalties or deductions for non-compliance. However, enforcing these contractual requirements typically relies on manual inspections, written reports, and subjective judgment, which may result in inconsistent documentation and delayed corrective actions. Integrating computer vision with contractual compliance offers an opportunity to automate the collection of objective evidence, improve transparency, and support fair and consistent enforcement of contractual safety provisions. This study reviews relevant previous research across several interconnected domains, including personal protective equipment (PPE) detection, unsafe behavior recognition, restricted-area intrusion monitoring, worker identity verification, corporate logo recognition, and fully unified site safety frameworks. Through this narrative review, we outline the current technological landscape, highlight critical research gaps in system integration, and establish the theoretical and practical foundation for an advanced, automated safety enforcement framework.

1.2. Literature Review

Previous research has focused on detecting PPE compliance using deep-learning models, particularly pretrained YOLO-based object detection algorithms. Wang et al. [1] proposed a pretrained YOLO model and compared the detection of helmets and vests with different colors, and developed YOLOv5x, which achieved the best performance (86.55% mAP). Ferdous and Ahsan [2] proposed a YOLO-based PPE detection framework for construction sites, demonstrating the effectiveness of deep-learning approaches for real-time safety monitoring. Lin [3] compared a YOLOv8-based model with several alternative architectures and reported superior helmet-detection accuracy, particularly after introducing the YOLOv8n-SLIM-CA enhancement, especially for small targets and complex backgrounds. Cheng [4] considered both detection accuracy and inference speed and developed a customized model (HRHD-YOLOv8); he improved the accuracy, but the speed was slowed down. Song et al. [5] improved YOLOv8 by adding modules (DWR, ASPP, NWD) to enhance helmet detection accuracy, especially for small objects and complex construction environments, and improved the model to achieve a mAP of 92.0%. J. Xing [6] proposed an improved YOLOv8-based model (YOLO-P2-Ghost-Dyhead) that incorporates a small-target detection layer, a lightweight backbone, an attention-based head, and an optimized loss function, achieving higher accuracy and significantly faster detection. Zhang et al. [7] introduced a dynamic focusing mechanism to balance the contribution of anchor samples during training and improved the FGP-YOLOv8 model, achieving 89.6%. Han et al. [8] enhanced YOLOv8 (YOLOv8s-SNC) using SPD-Conv, attention modules, and a small-object detection layer for better helmet detection. It achieved 92.4% mAP. Yang et al. [9] proposed SDCB-YOLO, an improved YOLOv8-based model using SE attention, DIOU loss, CARAFE upsampling, and BiFPN feature fusion. The model achieved a mAP of 97.1%. Barlybayev et al. [10] evaluated multiple YOLOv8 variants on benchmark datasets (CHV and SHEL5K) and found that larger models (YOLOv8l and YOLOv8x) achieved the highest accuracy. Jia et al. [11] proposed GeoIoU-SEA-YOLO, developing the GeoIoU

loss function for more accurate bounding-box localization and the Structural-Enhanced Attention (SEA) mechanism to improve feature extraction and attention to critical regions. Liu et al. [12] proposed an improved YOLOv5s-based framework with a Focal-EIoU loss function and a Coordinate Attention (CA) mechanism. The model achieved 89.3% mAP. Xu et al. [13] proposed DAF-YOLO, an improved YOLOv5-based model, especially for small targets like missing safety helmets or reflective vests in complex environments, achieving up to 95.7% mAP. Liu et al. [14] proposed a multi-task computer vision framework using YOLOv8 to improve construction safety monitoring through object detection, segmentation, pose estimation, and tracking in real time.

Tran et al. [15] proposed IDC4D, a hybrid safety-monitoring framework that combines 4D BIM (Building Information Modeling) and computer vision (YOLOv8) to automatically detect workers entering hazardous construction zones, and the system achieved about 93% detection accuracy. Zhao et al. [16] proposed a CNN-based real-time early warning system that detects construction workers entering hazardous zones using video surveillance, and the system achieved about 87% detection accuracy. Saeheaw [17] also proposed SC-YOLO, a real-time PPE detection model with multi-class for construction sites that combines the CSP Darknet backbone with the Sophia second-order optimizer to improve detection accuracy, achieving up to 97.6% mAP.

While detection of safety violations is essential, identifying the responsible individual is equally important for accountability. Face recognition technologies have been widely used for identity verification. Kwak and Kim [18] proposed a system that combines YOLO-based PPE detection with face recognition techniques for worker identification. This approach demonstrated the feasibility of linking detected violations to specific individuals. However, the system was evaluated under controlled conditions and may face challenges in real construction environments due to occlusion, lighting variations, distance, and the use of safety equipment.

In addition to identifying individual workers, recognizing their affiliated companies is important for enforcing contractual responsibilities. Logo detection has been widely studied in computer vision. Hou et al. [19] provided a comprehensive survey of deep learning approaches for logo recognition. López et al. [20] combined object detection with pose estimation to enhance PPE compliance monitoring. Other studies, such as X. Zhao et al. [16], have explored advanced AI-based systems for detecting hazards and unsafe behaviors in real time. Biswas and Hoque [21] demonstrated the use of YOLOv8 for detecting PPE and equipment simultaneously. Despite these advancements, most existing systems focus primarily on detection and monitoring, with less emphasis on connecting detection and worker identification to company attribution and contractual review.

1.3. Research Gap

Based on the reviewed literature, it is evident that significant progress has been made in PPE detection, unsafe behavior recognition, and restricted-area monitoring using deep learning techniques. YOLO-based models have proven highly effective for real-time object detection, while face recognition and logo detection methods offer potential for identifying individuals and organizations. However, existing studies already combine PPE detection and worker identification, as illustrated by Kwak and Kim [18]. The gap addressed here is the connection of these outputs to company attribution and documented contractual review. Specifically, this study proposes a workflow combining multi-class violation detection, worker identification, company attribution, and reviewable contractual incident records. Therefore, there is a clear need for an intelligent construction safety management system that integrates detection, identification, and incident documentation. This research aims to address this gap by developing a system intended to support management practices in

the construction sector from a safety perspective while linking detected events to potential contractual deductions for authorized review. The proposed system not only detects safety violations but also associates visual evidence with worker and company identification outputs to support reviewable contractual decisions. The methodological contribution is this application-level integration of task-specific detection and recognition modules, timestamped incident records, and contractual decision support. It extends the detection-and-identification workflow toward contract administration.

2. Methods

2.1. System Overview

This study proposes an intelligent safety management system designed to automate safety monitoring and enforce contractual compliance at construction sites. The system integrates multi-class object detection, worker identification, and an automated alert generation framework. The framework is used to identify contractors and apply contractual safety deductions to mitigate the risk of any accidents on site. Pretrained deep learning models are employed to detect personal protective equipment (PPE), including helmets and safety vests, as well as additional recognition modules (face recognition, vest ID detection, and logo detection) that are used to identify workers and their related companies. The system also uses an attendance-based database filter to improve the identification accuracy by limiting the search space to only the active workers on-site. This encourages construction companies to invest in advanced technologies to reduce the risk of safety accidents. By linking detected violations to contractual penalty mechanisms, the proposed system supports data-driven decision-making and enhances safety compliance in large-scale construction projects. Figure 1 summarizes the overall research procedures and methodological pathway.

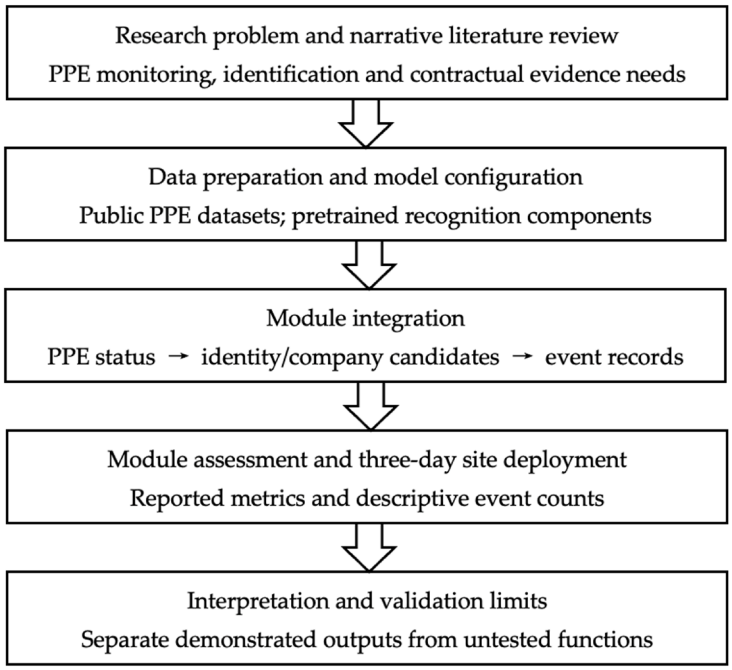


Figure 1. Overall research procedures and methodological pathway.

2.2. Hardware and System Setup

The proposed safety detection system was implemented using a workstation equipped with an Intel Core i7 CPU, 32 GB RAM, and an NVIDIA RTX 3060 GPU with 12 GB VRAM. High-resolution 4 MP surveillance cameras were used to capture real-time video streams

from the monitored worksite. The implementation was carried out using Python, PyTorch, OpenCV, and the Ultralytics YOLOv8 framework. The system uses two fundamentally different approaches to computer vision:

- Deep Learning (CNN): The system feeds images into neural networks that have been pre-trained on datasets. These networks have learned to recognize patterns automatically without being explicitly programmed with rules.
- Traditional Image Processing: The system applies mathematical operations directly to pixel values, filtering, thresholding, edge detection, and template comparison.

2.3. PPE Detection Module

The proposed system employs two separate detection models, each dedicated to a specific PPE detection task and detecting multiple classes. This architectural decision was made to reduce training complexity and enable more effective feature learning for each target class. By decomposing the detection task into two sequential stages, each model can focus on a narrower objective, with the aim of improving classification precision and reducing false detections. The sequential execution of the models facilitates a more robust detection pipeline.

2.3.1. Safety Helmet Model

The dataset used for training the helmet detection model was obtained from publicly available datasets hosted on the Kaggle platform. The helmet detection dataset consisted of approximately 18,500 annotated images across the relevant safety classes: Helmet, No-Helmet, and Person. Training was configured for 100 epochs at an input resolution of 800 × 800 pixels, with a batch size of 16 and an early-stopping patience of 10 epochs. Optimization was carried out using Stochastic Gradient Descent (SGD) with an initial learning rate of 0.01, a final learning-rate factor of 0.01, a momentum of 0.937, a weight decay of 0.001, and a warm-up period of 3 epochs.

2.3.2. Safety Vest Model

The model was trained on publicly available datasets hosted on Kaggle, containing approximately 16,200 annotated images across the relevant safety classes: Vest, No-Vest, and Person. Training was configured for 100 epochs at an input resolution of 640 × 640 pixels, with a batch size of 16 and an early-stopping patience of 10 epochs. Optimization was carried out using Stochastic Gradient Descent (SGD) with an initial learning rate of 0.01, a final learning-rate factor of 0.01, a momentum of 0.937, a weight decay of 0.001, and a warm-up period of 3 epochs.

2.4. Restricted Area Detection

The proposed system tackles the challenge of monitoring hazardous and restricted zones, which are very dynamic in construction environments, such as heavy machinery zones or excavation areas. These zones are defined within the system using camera-based mapping, allowing the safety officers to get instant alerts in case restricted areas are accessed. When a worker enters a designated hazardous zone, the system automatically generates an alarm to prevent potential accidents. Additionally, the system also evaluates PPE compliance within these zones, and in case of any violation of the safety equipment, the alert severity increases, reflecting the elevated risk of an accident. Restricted Area Monitoring. The restricted area is defined as a rectangular region of interest specified by two opposite corner coordinates, $((x_1, y_1), (x_2, y_2))$, which the operator configures either through the interface or by interactively drawing the zone on the video frame. Person localization is performed by the YOLOv8n detector, and for each detected individual, the centroid of the bounding

box is computed as $(c_x, c_y) = ((x_1 + x_2)/2, (y_1 + y_2)/2)$. An intrusion is registered when this centroid satisfies the point-in-rectangle condition $rx_1 \leq c_x \leq rx_2$ and $ry_1 \leq c_y \leq ry_2$, indicating that the person lies within the restricted boundary. Upon a positive detection, the system highlights the subject with a red bounding box, triggers a priority-based audible alarm, and logs the event with a timestamped screenshot and CSV record. Figure 2 illustrates restricted-area monitoring.



Figure 2. Showing restricted area.

2.5. Worker Identification Module

The proposed system automates the detection of safety violations and facilitates the identification of responsible workers through an alert and notification system. This approach aims to enhance monitoring efficiency and reduce reliance on manual supervision, potentially enabling safety officers to monitor larger areas with a reduced manual workload. Consequently, the system is intended to support more comprehensive and continuous site monitoring.

2.5.1. Face Recognition

Face Recognition Pipeline. Faces are first located using HOG (Histogram of Oriented Gradients), an image processing method that divides the image into 8 × 8 cells, computes gradient orientations, accumulates them into histograms, and classifies sliding windows as face/non-face. Each detected face is then identified by a pre-trained deep CNN that maps the cropped face to a fixed-length embedding vector, which is compared against all stored embeddings in the database to find the closest match falling below a distance threshold. The dlib ResNet-34 model is pre-trained on 3 million images to output a 128-D embedding, compared by Euclidean distance (threshold 0.6, ~0.1 s/face, 99.38% LFW accuracy).

2.5.2. Vest ID Recognition

To improve worker identification, particularly when face recognition is limited by camera position, viewing angle, distance, or occlusion, the system incorporates vest ID recognition as an additional identification method. When a worker is wearing an identifiable safety vest, the system extracts the identification number from the vest and compares it with the corresponding records in the worker database. Figure 3 illustrates vest ID recognition.



Figure 3. Vest ID recognition.

The vest ID recognition pipeline consists of the following steps:

- The vest image is enhanced through five image-processing variations (CLAHE, Otsu, morphological, sharpening, bilateral filtering).
- Tesseract (rule-based; OEM 3, PSM 6/7/8/11/13, digit whitelist 0–9 → 25 attempts).
- EasyOCR (deep learning; CRAFT-CNN text detection + CRNN recognition, English, digit whitelist, CPU inference).
- CRAFT (Character Region Awareness for Text Detection) is a text detector built on a VGG16-BN backbone, trained using the Adam optimizer with a learning rate of 1×10^{-4} and trained on 800,000 images.
- CRNN (Convolutional Recurrent Neural Network) follows a ResNet–BiLSTM–CTC architecture and is trained with the Adadelta optimizer ($\rho = 0.95$, $\varepsilon = 1 \times 10^{-8}$) at a learning rate of 1.0, a batch size of 192, approximately 300,000 iterations, and gradient clipping of 5.0, operating on 32 × 100 grayscale word images with a Connectionist Temporal Classification (CTC) loss, and trained on 9 million images.

2.5.3. Logo Detection and Recognition

The system identifies the company logo by combining four complementary methods, optionally guided by a YOLOv8 Nano detector that localizes the logo region. The detector was trained on 15,000 images with a batch size of 4, 50 epochs, 640 × 640 input, an SGD optimizer (momentum 0.937), a learning rate of 0.01, and mosaic + horizontal-flip augmentation.

Three image-processing methods then match each stored logo template against the vest image: Template Matching uses normalized cross-correlation across 11 scales; SSIM compares luminance, contrast, and structure and tolerates lighting changes; and ORB feature matching uses binary keypoint descriptors.

A fourth deep-learning method, EasyOCR, detects and reads visible text on the vest and fuzzy-matches it against known logo names. Figure 4 illustrates logo detection and recognition.



Figure 4. Logo detection and recognition.

2.6. Attendance-Based Filtering

To improve the accuracy and system efficiency, the proposed framework integrates an attendance-based database of workers and companies present on-site. This approach reduces the dataset that the modules search and filters for active individuals, decreasing computational complexity and enabling a lighter system that requires fewer hardware resources.

2.7. Analysis and Reporting System

When a violation is confirmed and the responsible individual is identified, the system performs several automated actions: It captures and stores the relevant image frame, logs associated metadata (including timestamp, camera ID, violation type), generates real-time alerts to safety personnel, and stores the incident data in a centralized database to support reporting, administrative review, and performance evaluation; it can also be integrated with ERP systems and sent via mobile communication platforms for easier and quicker actions.

2.8. Contractual Compliance Decision Support

Once a safety violation is detected and the responsible worker or subcontractor is identified, the system can associate the recorded violation with a predefined penalty category derived from the applicable construction contract. The system stores supporting evidence, including the image, timestamp, camera ID, violation type, worker identification, and contractor information, thereby creating an auditable record for review by authorized project personnel. Where a contractual penalty matrix is applicable, the corresponding potential deduction can be calculated and presented for administrative review. Any contractual deduction remains subject to verification and approval by authorized project management personnel in accordance with the applicable contract. After validation, the violation report can be exported to the ERP or contract management system where deductions are reflected in the subcontractor's monthly payment certificate.

2.9. System Workflow

For each frame, the PPE models detect helmet and vest classes. Face regions are encoded and matched to stored embeddings; vest regions are processed by OCR to recover identifiers; logo regions are compared with stored templates. The outputs are associated with timestamped incident evidence for identity, company, and contractual review. Figure 5 summarizes the system workflow.

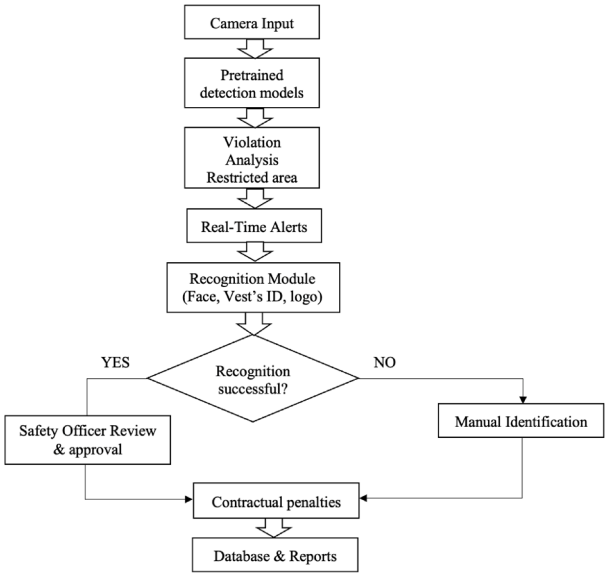


Figure 5. Workflow for automated construction safety monitoring and worker identification.

2.10. Evaluation Metrics and Selected Models

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$AP = \int_0^1 P(R)dR \quad (3)$$

$$mAP = \frac{\sum_{i=1}^{n} APi}{n} \quad (4)$$

$$F1 = \frac{2xPrecisionxRecall}{Precision + Recall} \quad (5)$$

where TP denotes the number of true positives, FP denotes the number of false positives, and FN denotes the number of false negatives. P(R) represents precision as a function of recall R. AP denotes the Average Precision for a given class, $APı$ denotes the Average Precision for the *i*-th class, and n denotes the total number of classes.

2.11. Methodological Limitations

Several limitations were identified during system development. First, the accuracy of face recognition is affected by camera positioning, distance, and occlusion, particularly in large-scale construction environments where faces appear small in captured frames. Second, lighting variations and dynamic site conditions may impact detection performance. Additionally, the system relies on camera-based monitoring, which may result in blind spots if camera coverage is not optimally designed. These limitations highlight the importance of proper camera placement and suggest opportunities for future improvements through enhanced imaging technologies and multi-sensor integration.

3. Results and Discussion

Evaluating the effectiveness of the proposed safety monitoring system, several performance metrics were employed, including Precision, Recall, F1-Score, and mAP@0.5.

Table 1 summarizes the performance of the developed detection and recognition modules under real-world construction site conditions.

Table 1. Comparative performance of the developed detection and recognition models.

Model	Precision	Recall	F1-Score	mAP@0.5
Helmet	92%	89%	90.5%	96%
Vest	91%	88%	89.5%	95%
Face recognition	80%	72%	75.8%	N/A *
Vest id recognition	68%	63%	65.4%	N/A *
logo recognition	74%	68%	70.9%	N/A *

mAP@0.5 is reported only for object detection models. * It does not apply to the face recognition, logo recognition, or vest ID recognition modules.

The helmet and safety vest detection modules demonstrated the strongest overall performance, achieving mAP@0.5 scores of 96% and 95%, respectively. Both models were developed as dedicated multi-class detectors based on the YOLOv8 architecture and trained on large, task-specific datasets. This design choice enabled each model to focus on learning the characteristics of a specific PPE detection task. The slightly superior performance of the helmet detector may be associated with its larger training dataset and higher input image resolution. Higher input resolutions are generally associated with enhanced detection of fine-grained visual features, contributing to improved localization and overall detection accuracy. The helmet and safety-vest datasets were each divided into 80% for training, 10% for validation, and 10% for testing. The ground-truth annotations used for evaluation were manually reviewed. For face recognition, a match was accepted when the Euclidean distance between the face embeddings was below 0.6.

The identification modules showed a clear performance gradient: the face recognition module achieved the highest F1-score (75.8%), followed by the logo recognition module (70.9%), while the vest ID recognition module recorded the lowest performance among all five modules (65.4%). This trend reflects the increasing complexity of the respective identification tasks under real-world construction site conditions.

Although the face recognition module was built upon a robust dlib ResNet-34 model that achieves 99.38% accuracy on the controlled Labeled Faces in the Wild (LFW) benchmark, its performance declined in the field due to factors such as low image resolution at long distances, partial occlusion caused by helmets and dust, and non-frontal viewing angles. Logo recognition presented additional challenges, as company logos typically occupy only a small number of pixels in wide-angle surveillance footage. The limited visual information available for these small targets constrained recognition accuracy. The vest ID recognition module proved to be the most challenging task because identification numbers are often small, partially occluded, blurred by motion, or distorted by folds in the safety vest, making reliable character extraction particularly difficult in unconstrained environments. Figure 6 summarizes the safety violations recorded during the three-day deployment.

During a three-day deployment at the Giza Zoo Project construction site, the proposed system autonomously detected and logged 126 safety violations, each accompanied by timestamped visual evidence. Four fixed cameras covered different areas within the project. All four cameras operated throughout the daily recording period, from 08:30 to 17:00, providing 8.5 h of recording per day. Recording was conducted during daytime working hours; no night work took place at the site. The total workforce present at the project was 660 workers on the first day, 680 on the second day, and 630 on the third day.

These results demonstrate the system's ability to operate continuously under real-world, uncontrolled construction site conditions while maintaining reliable violation detection and documentation. Compared with traditional manual inspections, which are inherently periodic and dependent on human availability, the proposed system provides continuous automated monitoring capable of identifying safety violations throughout the monitoring period. Consequently, it has the potential to detect a greater number of violations that might otherwise go unnoticed between routine inspections.

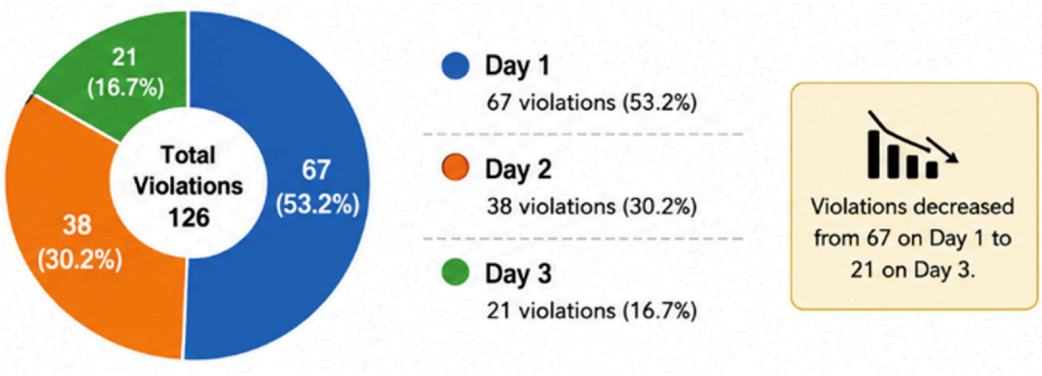


Figure 6. Safety violations detected during the three-day deployment.

Beyond improving safety monitoring, the proposed framework can support construction contract administration. The proposed system links verified violations to the relevant worker and subcontractor where identification is confirmed through manual review, supporting contractual compliance review based on applicable penalty clauses. By automatically generating timestamped visual evidence and maintaining incident records, the framework may help reduce subjectivity during safety inspections and support the verification of potential contractual deductions. Furthermore, the generated records can support payment certification, contractor performance evaluation, and dispute resolution. Future integration could extend beyond exporting records as Excel files to embedding the framework's functionality directly within Enterprise Resource Planning (ERP) systems, potentially enabling semi-automated processing of contractual penalties, subject to authorized review and approval. Actual contractual deductions, payment certification, and ERP integration were not independently validated during the present deployment. These potential applications extend the role of computer vision from safety monitoring to decision support for construction contract management. The field deployment results revealed a gradual reduction in the number of recorded safety violations over the three-day evaluation period. Figure 7 shows the distribution of violation types over the three days. Table 2 summarizes the daily counts and illustrative deductions.

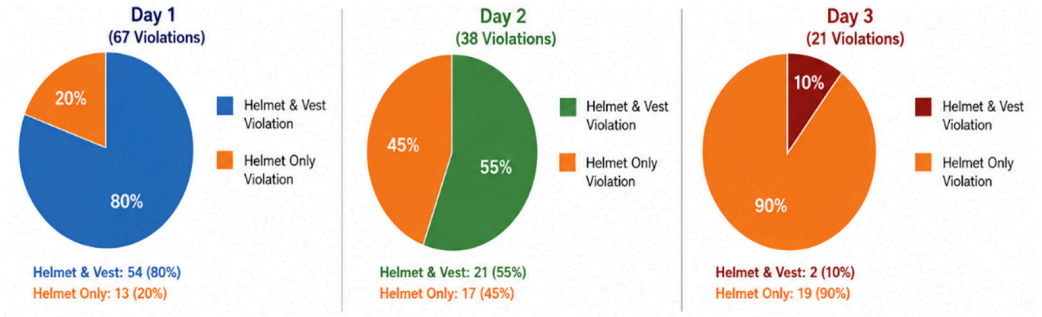


Figure 7. Distribution of violation types over three days.

Table 2. Results and classification of violations.

No of Days	Helmet	Vest	Violations	Deductions
First day	67	54	67	67,000
Second day	38	21	38	38,000
Third day	21	2	21	21,000
Total	126	77	126	126,000

Illustrative amounts use EGP 1000 per event and remain subject to contract verification and approval. Helmet and vest count overlap; the total number of events is 126. A total of 67 violations were detected on the first day, decreasing to 38 violations on the second day, with the remaining violations recorded on the third day. Further manual analysis of the attendance sheets showed that 66% of the violations recorded during the second and third days were associated with workers who had not been recorded as present during the previous monitoring period. This finding suggests that the observed reduction in violations was not solely attributable to improved compliance among previously monitored workers but was also influenced by the arrival of new workers at the construction site. Table 3 summarizes the methods used for each detection target.

Table 3. Methods by detection target.

Model	Method	Output
Helmet	YOLOv8, 800 × 800 input	Helmet/No-Helmet/Person
Vest	YOLOv8, 640 × 640 input	Vest/No-Vest/Person
Face recognition	HOG; 128-D embedding; Euclidean distance < 0.6	Worker identity match
Vest id recognition	Five preprocessing variants; Tesseract and EasyOCR	Vest number for worker lookup
Logo recognition	Optional YOLOv8n; template, SSIM, ORB and OCR matching	Company attribution

4. Conclusions

This study presented a unified computer vision framework for safety area monitoring and contractual compliance on construction sites. The proposed system integrates PPE detection, restricted-area monitoring, worker identification, logo recognition, and attendance-based filtering into a single automated platform. By combining real-time detection with worker identification and contractual review mechanisms, the framework supports proactive safety management and is intended to support accountability on construction projects. This research aimed to leverage recent advances in artificial intelligence to improve construction safety management, productivity, and operational efficiency in large-scale projects. The proposed system demonstrates how pretrained models can be implemented in real-time construction environments to assist project managers in classifying and evaluating contractors and workers, and supporting safety officers, with the aim of reducing manual monitoring workload and improving their performance. This helps in monitoring safety compliance and improving their performance. In addition, the proposed framework demonstrates the potential to support contractual compliance by linking verified safety violations with predefined contractual penalty categories. The resulting timestamped evidence and contractor-specific violation records can provide structured decision support for project management. However, automated contractual deductions, payment certification, and ERP integration were not independently validated in the present field deployment and should be investigated in future work. These capabilities extend the role of computer vision beyond safety monitoring to an intelligent decision-support tool for construction contract management.

The study identified several limitations affecting system performance. These include camera positioning, image quality, lighting conditions, and the accuracy of face detection in construction environments, where images often contain significant noise and faces may appear very small and difficult to detect in many cases due to distance from the camera. Future research should focus on improving the model's robustness under challenging conditions, enhancing detection and segmentation accuracy, and addressing the continuity of tracking IDs to achieve more reliable long-term monitoring.

The proposed system was deployed during the development of the Giza Zoo Project, where it recorded 126 safety violations over a three-day period. Helmet and vest detection achieved the highest performance among all evaluated modules, potentially reflecting the availability of a large, annotated training dataset and the robustness of the YOLOv8 architecture. In contrast, the face recognition, logo recognition, and vest ID recognition modules exhibited comparatively lower recall rates. This reduction in performance may be associated with variations in image quality, viewing angles, partial occlusions, and challenging lighting conditions encountered in real-world construction environments. Additionally, the relatively small size of logos and identification numbers within surveillance footage further increased the difficulty of accurate recognition. One of the primary challenges encountered was the small size of faces in captured images, particularly in large construction areas, which reduced identification accuracy. To address this limitation, future work should focus on collecting larger and more representative construction-site datasets containing low-resolution images, varying lighting conditions, company logos, vest identification numbers, and facial data. Additionally, improvements in camera resolution, camera placement, and model training are expected to enhance the system.

Author Contributions: Conceptualization, G.S., A.R. and N.A.; methodology G.S. and A.R.; software, G.S.; validation, A.R. and N.A.; formal analysis, G.S.; investigation, G.S.; resources, G.S.; data curation, G.S.; writing—original draft preparation, G.S.; writing—review and editing, G.S.; visualization, G.S.; supervision, A.R. and N.A.; project administration, G.S. and A.R.; funding acquisition, G.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The university required the research team to obtain formal written permission from the company to conduct the study at the construction site and written informed consent from all participating workers. Both requirements were fulfilled.

Informed Consent Statement: Workers and site management were informed of the camera-based safety-monitoring activities and their scientific and academic research purposes. As part of this procedure, written informed consent was obtained from all workers who entered the site during the three-day deployment through the participating contractors. Identifiable worker data were used internally for safety monitoring and research-related verification and are not included in the publicly available data.

Data Availability Statement: The publicly available datasets used for training the PPE detection models are described in the manuscript. Data collected during the construction-site field deployment are not publicly available because they contain potentially identifiable worker information and site-specific operational records. Aggregated results supporting the findings of this study are presented within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wang, Z.; Wu, Y.; Yang, L.; Thirunavukarasu, A.; Evison, C.; Zhao, Y. Fast Personal Protective Equipment Detection for Real Construction Sites Using Deep Learning Approaches. *Sensors* **2021**, *21*, 3478. [CrossRef] [PubMed]

2. Ferdous, M.; Ahsan, S.M.M. PPE detector: A YOLO-based architecture to detect personal protective equipment (PPE) for construction sites. *PeerJ Comput. Sci.* **2022**, *8*, e999. [CrossRef] [PubMed]
3. Lin, B. Safety Helmet Detection Based on Improved YOLOv8. *IEEE Access* **2024**, *12*, 28260–28272. [CrossRef]
4. Cheng, L. A Highly Robust Helmet Detection Algorithm Based on YOLO V8 and Transformer. *IEEE Access* **2024**, *12*, 130693–130705. [CrossRef]
5. Song, X.; Zhang, T.; Yi, W. An improved YOLOv8 safety helmet wearing detection network. *Sci. Rep.* **2024**, *14*, 17550. [CrossRef] [PubMed]
6. Xing, J.; Zhan, C.; Ma, J.; Chao, Z.; Liu, Y. Lightweight detection model for safe wear at worksites using GPD-YOLOv8 algorithm. *Sci. Rep.* **2025**, *15*, 1227. [CrossRef] [PubMed]
7. Zhang, L.; Ma, H.; Huang, J.; Zhang, C.; Gao, X. An Improved Lightweight Safety Helmet Detection Algorithm for YOLOv8. *Comput. Mater. Contin.* **2025**, *83*, 2245–2264. [CrossRef]
8. Han, D.; Ying, C.; Tian, Z.; Dong, Y.; Chen, L.; Wu, X.; Jiang, Z. YOLOv8s-SNC: An Improved Safety-Helmet-Wearing Detection Algorithm Based on YOLOv8. *Buildings* **2024**, *14*, 3883. [CrossRef]
9. Yang, X.; Wang, J.; Dong, M. SDCB-YOLO: A High-Precision Model for Detecting Safety Helmets and Reflective Clothing in Complex Environments. *Appl. Sci.* **2024**, *14*, 7267. [CrossRef]
10. Barlybayev, A.; Amangeldy, N.; Kurmetbek, B.; Krak, I.; Razakhova, B.; Tursynova, N.; Turebayeva, R. Personal protective equipment detection using YOLOv8 architecture on object detection benchmark datasets: A comparative study. *Cogent Eng.* **2024**, *11*, 2333209. [CrossRef]
11. Jia, X.; Zhou, X.; Shi, Z.; Xu, Q.; Zhang, G. GeoIoU-SEA-YOLO: An Advanced Model for Detecting Unsafe Behaviors on Construction Sites. *Sensors* **2025**, *25*, 1238. [CrossRef] [PubMed]
12. Liu, Y.; Wang, P.; Li, H. An Improved YOLOv5s-Based Algorithm for Unsafe Behavior Detection of Construction Workers in Construction Scenarios. *Appl. Sci.* **2025**, *15*, 1853. [CrossRef]
13. Xu, Q.; Cheng, X.; Zhou, X.; Jia, X.; Wang, X.; Shi, Z.; Huang, S.; Zhang, G. DAF-YOLO: Detection of Unsafe Behaviors on Construction Sites. *Sensors* **2025**, *25*, 7216. [CrossRef] [PubMed]
14. Liu, L.; Guo, Z.; Liu, Z.; Zhang, Y.; Cai, R.; Hu, X.; Yang, R.; Wang, G. Multi-Task Intelligent Monitoring of Construction Safety Based on Computer Vision. *Buildings* **2024**, *14*, 2429. [CrossRef]
15. Tran, S.V.-T.; Lee, D.; Bao, Q.L.; Yoo, T.; Khan, M.; Jo, J.; Park, C. A Human Detection Approach for Intrusion in Hazardous Areas Using 4D-BIM-Based Spatial-Temporal Analysis and Computer Vision. *Buildings* **2023**, *13*, 2313. [CrossRef]
16. Zhao, J.; Xu, Y.; Zhu, W.; Liu, M.; Zhao, J. Real-Time Early Safety Warning for Personnel Intrusion Behavior on Construction Sites Using a CNN Model. *Buildings* **2023**, *13*, 2206. [CrossRef]
17. Saeheaw, T. SC-YOLO: A Real-Time CSP-Based YOLOv11n Variant Optimized with Sophia for Accurate PPE Detection on Construction Sites. *Buildings* **2025**, *15*, 2854. [CrossRef]
18. Kwak, N.; Kim, D. Detection of Worker's Safety Helmet and Mask and Identification of Worker Using Deeplearning. *Comput. Mater. Contin.* **2023**, *75*, 1671–1686. [CrossRef]
19. Hou, S.; Li, J.; Min, W.; Hou, Q.; Zhao, Y.; Zheng, Y.; Jiang, S. Deep Learning for Logo Detection: A Survey. *ACM Trans. Multimed. Comput. Commun. Appl.* **2023**, *20*, 72. [CrossRef]
20. López, L.; Suárez-Ramírez, J.; Alemán-Flores, M.; Monzón, N. Automated PPE compliance monitoring in industrial environments using deep learning-based detection and pose estimation. *Autom. Constr.* **2025**, *176*, 106231. [CrossRef]
21. Biswas, M.; Hoque, R. Construction Site Risk Reduction via YOLOv8: Detection of PPE, Masks, and Heavy Vehicles. In Proceedings of the 2024 IEEE International Conference on Computing, Applications and Systems (COMPAS), Cox's Bazar, Bangladesh, 25–26 September 2024; pp. 1–6. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.