

Article

Integrated Construction-Site Hazard Detection System Using AI Algorithms in Support of Sustainable Occupational Safety Management

Zuzanna Woźniak ^{1,*} , Krzysztof Trybuszewski ² , Tomasz Nowobilski ¹ , Marta Stolarz ³  and Filip Šmalec ⁴

- ¹ Department of Materials Engineering and Building Processes, Faculty of Civil Engineering, Wrocław University of Science and Technology, 50-370 Wrocław, Poland; tomasz.nowobilski@pwr.edu.pl
² ITRIT, 89-100 Nakło nad Notecią, Poland; k.trybuszewski@gmail.com
³ Creoox AG, 9495 Triesen, Liechtenstein; marta.stolarz@creoox.com
⁴ Design, Technology & Business (Graphics), University College of Northern Denmark, 9200 Aalborg, Denmark; filipsmalec@gmail.com
* Correspondence: zuzanna.wozniak@pwr.edu.pl

Abstract

Despite preventive measures, the construction industry continues to exhibit high accident rates. In response, visual detection system was developed to support safety management on construction sites and promote sustainable working environments. The solution integrates the YOLOv8 algorithm with asynchronous video processing, incident registration, an open API, and a web-based interface. The system detects the absence of safety helmets (NHD) and worker falls (FD). Its low hardware requirements make it suitable for small and medium-sized construction enterprises, contributing to resource efficiency and digital transformation in line with sustainable development goals. This study advances practice by providing an integrated, low-resource solution that unites multi-hazard detection, event documentation, and system interoperability, addressing a key gap in existing research and implementations. The contribution includes an operational architecture proven to run in real time, addressing a gap between model-centred research and deployable, OHS applications. The system was validated using two independent test datasets, each comprising 100 images: one for NHD and one for FD. For NHD, the system achieved a precision of 0.93, an accuracy of 0.88, and an F1-score of 0.79. For FD, a precision of 1.00, though with a limited recall of 0.45. The results demonstrate the system's potential for sustainable construction site safety monitoring.

Keywords: construction; safety management; occupational health; sustainable development; accident at work; image recognition; personal protective equipment; YOLOv8; fall detection; machine learning



Academic Editor: Mario Fargnoli

Received: 1 October 2025

Revised: 8 November 2025

Accepted: 15 November 2025

Published: 26 November 2025

Citation: Woźniak, Z.; Trybuszewski, K.; Nowobilski, T.; Stolarz, M.; Šmalec, F. Integrated Construction-Site Hazard Detection System Using AI Algorithms in Support of Sustainable Occupational Safety Management. *Sustainability* **2025**, *17*, 10584. <https://doi.org/10.3390/su172310584>

Copyright: © 2025 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The construction sector has long been one of the most accident-prone areas of economic activity, with consistently higher accident rates than other industries [1,2]. This is due, among other things, to the diversity of professions, materials, and construction equipment [3]. According to data from the International Labour Organisation (ILO) [4], on average, there are approximately 60,000 fatal accidents at work on construction sites worldwide each year. On average, this means that one fatal accident occurs approximately every ten min. In industrialised countries, as many as 25–40% of all fatal accidents at work

occur on construction sites, even though this sector employs only 6–10% of the workforce in each country [4]. A similar pattern [5] can also be observed in countries around the world, including China, the United Kingdom, Singapore, Australia, and South Korea.

Although approximately USD 1.8 billion is spent annually on workplace safety audits and inspections, and the construction worker safety market is valued at around USD 3.5 billion in 2025 [6,7], accident rates remain high due to limited inspection frequency and human error, emphasising the urgent need for automated hazard detection systems.

Accidents at work and other dangerous incidents in construction generate significant material losses that burden not only companies but also entire societies. These costs include both direct expenses related to the treatment and rehabilitation of injured persons and indirect losses resulting from work stoppages, loss of productivity, and compensation payments [8,9]. The costs of accidents at work can place a significant burden on the budgets of companies and insurance institutions [10]. For example, in Poland alone, the estimated material losses caused by accidents at work in the period 2015–2022 amount to approximately USD 3.75 million [11]. However, it is worth noting that the losses associated with pain, suffering, and reduced quality of life for both the injured person and their family are practically impossible to estimate. In view of these devastating statistics, improving safety at work in the construction industry is now becoming a key priority [12]. One of the most important preventive measures aimed at improving the current unfavourable situation is the identification of the most common and most dangerous factors in the work process. These are the direct causes of accidents at work and near misses [13,14].

The results of previous studies indicate that one of the main causes of injuries and deaths in construction is falls by workers [15,16]. In a study conducted by Chi and Han [17], 42% of 9358 accidents at work were caused by falls from height. Similar results were observed when analysing data from the Polish branch of one of the largest construction companies [13], where it was found that 25.8% of accidents at work and 6.2% of near misses were incidents involving falls from height.

Another significant problem on construction sites is the failure of workers to use personal protective equipment [18]. The link between the lack or improper use of such equipment (including safety helmets) and the occurrence of accidents at work has been confirmed in studies [19]. An analysis of accident reports carried out by the authors clearly showed that the most dangerous incidents during work at height were directly related to the lack or incorrect use of protective equipment. Similar conclusions were presented by Polish researchers, who indicated that the failure to use personal protective equipment by employees was the cause of approximately 20% of accidents related to work on construction scaffolding [20].

In recent years, the automation of occupational safety monitoring through computer vision and artificial intelligence has gained significant attention. Numerous studies have focused on detecting the absence of personal protective equipment (PPE) or identifying unsafe worker behaviours using deep learning models, particularly convolutional neural networks (CNNs) and real-time object detection frameworks such as YOLO, SSD, or Faster R-CNN. These approaches have demonstrated high accuracy in controlled or simulated conditions; however, their implementation on real construction sites remains limited due to high computational demands, lack of integration capabilities, and difficulties in adapting to variable lighting and environmental conditions.

In response to the identified problems related to work safety on construction sites, this article presents an original approach using machine learning algorithms and advanced image and data analysis techniques. The aim of the research conducted and presented in this paper was to develop a comprehensive system integrating automatic detection of key hazards and dangerous events occurring on construction sites with the functionality of

monitoring them and generating alerts about identified events. The proposed solution, using machine learning algorithms, aims to identify health and safety violations, such as the failure of construction workers to wear protective helmets, and to detect potential hazards and incidents related to worker falls at an early stage. The proprietary system also enables automatic notification of the relevant supervisory services on the construction site about any dangerous events, which improves the response to hazards and supports rapid rescue operations.

The prototype developed in this study aims to address these challenges by employing a lightweight YOLOv8s model, which ensures real-time performance without the need for high-end hardware. Combined with a simple web-based interface and an open API, the system provides a flexible foundation for further research and development of affordable and easily deployable safety monitoring tools for small and medium-sized construction enterprises. According to the authors, the developed system can contribute to a significant reduction in the number of accidents at work, especially those with the most serious consequences. Consequently, the developed system not only addresses the practical need to improve safety on construction sites but also supports the implementation of the global Sustainable Development Goals (SDGs) defined in the United Nations 2030 Agenda [21]. In particular, it contributes to Goal 3 (Good Health and Well-being) by reducing the number of accidents and improving workers' well-being; to Goal 8 (Decent Work and Economic Growth) by minimising economic losses resulting from downtime and accident-related disruptions; and to Goal 9 (Industry, Innovation and Infrastructure) by promoting the adoption of modern information technologies in the construction sector [21]. In a broader perspective, the development of such AI- and image-analysis-based solutions supports the ongoing digital transformation of the construction industry and contributes to creating safer, more resilient, and sustainable working environments, consistent with the directions highlighted in recent reports on SDG progress [22].

Despite substantial progress in YOLO-based PPE and fall-detection models, existing studies typically deliver stand-alone detection modules without integration into end-to-end systems capable of alerting, event registration, or low-latency deployment on resource-constrained devices. This limits practical adoption, as the safety impact depends on complete supervisory workflows rather than detection alone. The present study addresses this gap by demonstrating a lightweight, fully integrated system architecture that connects detection, alerting, and incident traceability within a unified, API-driven framework.

Based on the identified research gap, the main objective of this study is to design and experimentally evaluate a prototype system for the automatic detection of key safety hazards on construction sites using computer vision and deep learning methods. Additionally, the study includes a comprehensive literature review aimed at consolidating current findings and positioning the proposed solution within the broader research landscape. From a technical perspective, the research addresses several important challenges related to the development of AI-based detection systems, including the need to ensure reliable real-time performance, achieve accurate classification with limited training data, and integrate detection, alerting, and event recording within a unified modular architecture. Accordingly, the study seeks to answer the following research questions:

RQ1: How effective is the proposed YOLOv8-based model in detecting the absence of safety helmets and worker falls within the designed prototype system?

RQ2: How can detection, alert generation, and event recording be conceptually integrated through an open API to support real-time safety monitoring?

RQ3: What are the main factors influencing detection accuracy and system performance under test conditions, and how can they inform further development of the model and its architecture?

The article is organised as follows: Section 2 provides an overview of the literature in the field of OHS hazard detection methods based on deep learning algorithms, with a particular focus on YOLO (You Only Look Once) models. Section 3 presents the proposed system architecture, including technological solutions, detection modules, and the API interface. Section 4 describes the experiments conducted, the test environment configuration, and the system effectiveness evaluation procedures. Section 5 contains an analysis of the results obtained, together with a comparison to existing solutions. Section 6 presents the conclusions, while the following Sections 7 and 8 discuss the limitations of the research conducted and indicate directions for further development of the system.

2. Literature Review

Work in the construction sector involves a high level of risk, which results, among other things, from the diversity of tasks performed, the technologies used, and the constantly changing conditions on the construction site [23–25]. The scale of hazards in construction and the consistently high number of accidents at work indicate an urgent need to implement modern, automated, and precise systems to increase the safety of workers. Existing methods, based mainly on periodic inspections and reporting incidents after they occur, are proving insufficient in the context of the growing dynamics and complexity of modern construction processes. In response to these challenges, digital technologies such as the Internet of Things (IoT), data analytics, and artificial intelligence (AI) algorithms are playing an increasingly important role, enabling the implementation of systems that provide continuous monitoring of working conditions and immediate detection of potential hazards.

2.1. Real-Time Image Analysis

Current scientific forecasts indicate that in the coming years, the development of predictive technologies and systems using real-time data analysis will play a key role in improving safety indicators in the construction sector [26]. Of particular importance in this context are advances in deep learning algorithms, including single-pass models such as YOLO, which have enabled significant improvements in the quality and performance of object detection in real-world environments.

The YOLO system is an advanced algorithm based on convolutional neural networks (CNN) for real-time object detection [27]. Its application is crucial in many fields, such as autonomous vehicles, robotics, video surveillance, and augmented reality [28]. YOLO algorithms are distinguished by their combination of high speed and precision, making them one of the most effective tools in this field [29].

Since the introduction of the first version of YOLO v1 by Redmon et al. [30] in 2016, the framework underwent several improvements [31]. Subsequent iterations have consistently improved, among other things, detection accuracy, speed, and processing efficiency on various hardware platforms [32]. Other important approaches have also been developed in the field of object detection, including R-CNN (Region-based Convolutional Neural Networks) and its newer versions, Fast R-CNN and Faster R-CNN. Although these methods are highly accurate, their computational complexity limits their use in systems where real-time detection is crucial. Alternatives include SSD (Single Shot MultiBox Detector) [33] and RetinaNet [34], algorithms that offer a compromise between speed and detection accuracy.

Despite growing competition, YOLO remains one of the most versatile solutions, widely used in both scientific research and industrial applications. Its architecture is easily adaptable to diverse environmental conditions, making it a suitable tool for hazard detection on construction sites and in occupational safety support systems.

2.2. The Use of Image Analysis Systems in Construction

In recent years, there has been a marked increase in interest in the use of modern technologies in the construction sector, including artificial intelligence algorithms [35]. An example of such an application is the detection of damage during the inspection of damaged wind turbines using the YOLO algorithm [36] or during bridge inspections to detect damage to bridge structures using migration learning techniques [37].

In the field of occupational safety, one of the most promising areas of research is the use of image analysis techniques for the automatic recognition of dangerous behaviour by construction workers and violations of occupational safety rules. Thanks to the use of deep learning algorithms, especially convolutional neural networks (CNN), it is possible to detect situations such as the lack of protective clothing or harnesses [27]. It is also worth noting research using YOLOv8 in combination with the ByteTrack algorithm to track and count objects on a construction site, which enables, among other things, automatic monitoring of the number of personnel, machines, and construction equipment [38].

Recent research on the use of artificial intelligence (AI) and advanced computer-vision techniques for detecting irregularities in the use of personal protective equipment (PPE) by construction workers has progressed significantly in recent years. We restrict our review to a selected set of representative vision-based studies on construction-site occupational health and safety (OHS), covering PPE compliance and fall detection. Table 1 summarises the detection tasks and system-level capabilities, while Table 2 extends the same entries with dataset sizes/splits, key performance metrics, applicable scenarios, and concise innovation notes.

Table 1. Selected vision-based studies on construction-site safety (OHS).

No	Reference	Technique	Detection Range	Recording and Permanent Storage of Events	API	RT
1	Wu et al. [39]	YOLOv8-CGS (CBAM, GAM, SiU)	No helmet	No	No	Yes
2	Lin [40]	YOLOv8n-SLIM-CA	No helmet	No	No	Yes
3	Qin et al. [41]	YOLOv8 + DyHead (ESD-YOLO)	Human fall	No	No	Yes
4	Wang et al. [42]	LFD-YOLO (lightweight YOLOv5)	Human fall	No	No	Yes
5	Huang et al. [43]	SDES-YOLO	Human fall	No	No	Yes
6	Pereira [44]	YOLOv8m	Human fall	No	No	Yes
7	Xing et al. [45]	YOLO-P2-Ghost-Dyhead (YOLOv8 + Ghost + Dyhead + EMASlideLoss)	Safety helmet, reflective clothing	No	No	Yes
8	Feng et al. [46]	YOLOv8s	helmets, vests; entry into dangerous areas	No	Yes	Yes

Details for each study (dataset size and split, metrics, scenarios, and innovation highlights) are provided in Table 2.

As presented in Table 2, recent YOLO-based architectures demonstrate mean Average Precision (mAP@0.5) values ranging from approximately 84.7% to 93.2%, while maintaining real-time inference performance between 60 and 90 frames per second. Wu et al. [39] proposed the most computationally efficient lightweight configuration, achieving 89 FPS. In contrast, studies such as those by Qin et al. [41] and Huang et al. [43] emphasised robustness in fall detection tasks, particularly under challenging visual conditions, including occlusion and variable illumination.

All the above approaches indicate the great potential of modern technologies in the context of preventive occupational safety management on construction sites, while emphasising the need for further improvement of algorithms and their adaptation to the specific nature of the construction environment. The above research also shows that the use of systems employing artificial intelligence algorithms can contribute not only to reducing the number of accidents at work, but also to improving the safety culture in construction organisations [47]. However, the implementation of such solutions requires appropriate

organisational preparation, including training for employees [48] and the adaptation of risk management procedures to new technologies [49].

Table 2. Detailed comparison of the studies listed in Table 1: datasets, performance, applicable scenarios, innovation, and key information.

Reference	Dataset	Performance	Applicable Scenarios	Innovation and Key Information
Wu et al. [39]	Safety Helmet Dataset (SHD)—45,200 imgs (31,640 train/6780 val/6780 test) and Safety Helmet Wearing Dataset (SHWD)—64,830 images	SHD: P 94.58%, F1 92.48%, mAP@0.5 93.18%, 89 FPS. SHWD: P 92.38%, F1 89.68%, mAP@0.5 90.98%, 87 FPS.	Real-time construction-site PPE monitoring under low light, occlusions, varied helmet colours/shapes; assessment of correct helmet wearing.	YOLOv8-CGS (YOLOv8 + GAM, CBAM, SiOU); lightweight (≈ 5.3 – 5.5 M params; 9.5–9.7 GFLOPs); ablations show stepwise gains
Lin [40]	Safety Helmet Wearing Dataset—7581 imgs; split 70/20/10; 2 classes (helmet/no-helmet)	mAP@0.5 94.361%, mAP@0.5:0.95 61.764%; vs. YOLOv8n: +1.46 pp P, +2.97 pp R, +2.15 pp mAP@0.5, +3.55 pp mAP@0.5:0.95; −6.98% params, −9.76% compute	Real-time helmet-wearing detection with small/distant targets, dense crowds, occlusion, varied lighting; validated across construction, power, mining, engineering, and traffic contexts; mobile/edge ready.	YOLOv8n-SLIM-CA: 9-image Mosaic, Coordinate Attention in backbone, Slim-Neck with GSConv/VoV-GSCSP, added small-object head (160×160); 4 heads total
Qin et al. [41]	Self-built fall-detection dataset (UR Fall + Fall Detection + web fall images): 4976 images, 5655 samples; split 70/20/10	ESD-YOLO: P 84.2%, R 82.5%, mAP@0.5 88.7%, mAP@0.5:0.95 62.5%; vs. YOLOv8s: +1.9/+4.1/+4.3/+2.8 pp, respectively.	Real-time human fall detection in complex scenes: scale changes, dense crowds, high background similarity, occlusion.	ESD-YOLO (YOLOv8s-based): C2Dv3 with DCNv3 in C2f blocks; DyHead dynamic head; EASlideloss (Slideloss + EMA) for hard samples/stable training
Wang et al. [42]	Falling Posture Image Dataset (FPID)—8416 imgs; Person/Down; split 70/10/20. Person Fall Detection Dataset (PFDD)—7859 imgs (web + volunteers); split 70/10/20.	LFD-YOLO: FPID mAP@0.5 84.7%, mAP@0.5:0.95 48.8%; PFDD mAP@0.5 92.7%, mAP@0.5:0.95 70.2%; P/R on FPID 83.1/77.3%, on PFDD 88.4/87.5% vs. YOLOv8s: +0.3–0.5 pp mAP@0.5.	Real-time fall detection for the elderly on resource-constrained edge devices; robust to low light, occlusion, scale changes, dense scenes.	LFD-YOLO (YOLOv5-based): CSRG (Split RepGhost) backbone, EMA attention, WFPN with weighted fusion + GSConv, Inner-WIoU loss
Huang et al. [43]	Human fall image dataset—4497 images	SDES-YOLO: P 84.3%, R 76.0%, mAP@0.5 85.1% vs. YOLOv8n: +3.41 pp mAP@0.5	Real-time fall detection for elderly care, workplace/public safety, assisted living, sports injuries; robust to occlusion, small targets, clutter, scale/pose changes.	SDES-YOLO (YOLOv8-based): SDFP (replaces SPPF), SEAM occlusion-aware attention, ES3 edge + spatial fusion, WIoU-Shape loss
Pereira [44]	Industrial Activity (Fall) Dataset—from 2560×1920 , 6 FPS mono video; 95 frames annotated (2 classes: Fall Detected/Human in Motion); after preprocessing total of 115 images	YOLOv8 v8n mAP@0.5 77.4%/mAP@0.5–0.95 61.2%; v8s	Industrial (lab-simulated) real-time fall detection and worker-activity monitoring; two-class setup for factory/warehouse-like scenes.	Custom augmentation for tiny dataset (resize, grayscale, blur, median-blur, CLAHE); systematic comparison of YOLOv8n/s/m/l/x; identifies v8m as optimal trade-off, v8l highest mAP@50.

Table 2. *Cont.*

Reference	Dataset	Performance	Applicable Scenarios	Innovation and Key Information
Xing et al. [45]	Roboflow construction-safety dataset—103,500 images, 9 classes (person, helmet, no-helmet, excavator, dump truck, mobile crane, bulldozer, safety vest, no-safety vest); 90k train/13.5k val	YOLOv8s: mAP@0.5 84.0%, P 85.0%, R 60.5% vs. YOLOv8n mAP@0.5 83.6%, YOLOv5s 84.5%, YOLOv5n 79.6%	Real-time construction-site safety: PPE detection (helmet/vest) and hazardous-zone intrusion via e-fence; OpenCV tracking; preliminary worker-to-machine distance estimation.	YOLOv8s pipeline for multi-class construction safety; add-ons (e-fence, tracking, distance design); class-wise error analysis highlights small-object/viewpoint challenges
Feng et al. [46]	No empirical dataset (concept/position paper).		Construction-site safety management: real-time monitoring, unsafe behaviour/condition detection, predictive risk identification, supervisor alerts, targeted training.	AI strategy framework

As seen across Tables 1 and 2, helmet-focused methods primarily exploit attention and small-object heads to cope with occlusion and low light, whereas fall-detection approaches favour lightweight backbones and tailored losses to recover recall in cluttered scenes. Dataset scales vary markedly (from 95 annotated frames to 103,500 images), limiting direct leaderboard-style comparisons and motivating a scenario-aware synthesis. Most prior studies optimise a single task and rarely report system-level features (event logging, open APIs). In contrast, our system jointly addresses two priority hazards (no-helmet and human falls) within a low-compute, asynchronous pipeline with a web UI, incident register, and open API, prioritising deployability on real construction sites.

Our decision to adopt YOLOv8 for construction-site safety analytics is motivated by peer-reviewed evidence. First, in a head-to-head comparison on identical helmet-detection data, Wang et al. [50] reported that YOLO attained both the highest accuracy (mAP 53.8%) and the fastest inference (10 FPS) relative to SSD and Faster R-CNN, while on the CHV dataset the YOLOv5 family scaled from real-time (YOLOv5s, 52 FPS on GPU) to high-accuracy (YOLOv5x, mAP 86.55%), with quantified robustness limits under face-blurring/occlusion perturbations (≈ 7 pp AP drop for helmets). At the same time, Fang et al. [51] demonstrated that two-stage Faster R-CNN is effective for non-hardhat-use detection on real sites, but with the characteristic throughput penalty of region-proposal pipelines. Beyond daylight conditions, Wang et al. [52] showed that an improved YOLOX variant maintains strong performance for small PPE targets in low-illumination tunnel environments, supporting the family's resilience to adverse lighting. Finally [53], lightweight YOLOv5 modifications (e.g., pruned/attention-augmented YOLOv5s) have been validated on safety-helmet tasks, indicating suitability for resource-constrained deployments without forfeiting accuracy. In Table 3, we summarise the comparative evidence and concise characteristics of the referenced methods.

Table 3. Comparison of YOLO, Faster R-CNN, SSD, and RetinaNet in building scenarios.

Detector Family	Representative Study	Scenario/Dataset	Accuracy (as Reported)	Speed (FPS) & Hardware (as Reported)	Hardware Demand	Site-Specific Robustness Notes
Faster R-CNN (two-stage)	Fang et al. [51]	Real construction CCTV (NHU)	High precision/recall for NHU (no unified mAP vs. YOLO in this paper)	FPS not reported; typical two-stage lower throughput	GPU recommended; higher compute (RPN + head)	Reliable on cluttered backgrounds; no dedicated low-light stress tests reported
SSD (one-stage, anchors)	Wang et al. [50]	Same helmet dataset as YOLO/Faster R-CNN	Lower than YOLO; YOLO best at mAP 53.8%	Slower than YOLO in this comparison	GPU; moderate compute	No detailed occlusion/illumination analysis in this comparison
YOLOv5/YOLO (one-stage)	Wang et al. [50]	CHV (helmet + vest + colours) and matched baselines	CHV: YOLOv5x mAP 86.55%	YOLOv5s $\approx$ 52 FPS (GPU); YOLO baseline 10 FPS in the comparative	Scales from edge (s) to server (x)	$\approx$ 7 pp helmet AP drop under face blurring; partial-occlusion FNs discussed
YOLOX (anchor-free YOLO family)	Wang et al. [52]	Tunnel construction (low-light, small PPE)	Improved accuracy vs. baselines in low light	Efficient; specific FPS not in abstract	GPU; efficient back-bone/head	Robust under illumination deficits and small-object regimes
Lightweight YOLOv5 (pruned/attention)	An et al. [53]	Helmet detection (construction)	Accuracy preserved/improved vs. base YOLOv5s	Real-time feasible on commodity GPUs	Reduced params/FLOPs (edge-friendly)	Suitable for embedded/low-resource deployments
YOLOv10 (one-stage, recent)	Wang [54]	Personal Protective Equipment	Higher PPE/object performance	Real-time on modern GPUs	GPU, efficient backbones	Robust across varied viewpoints
YOLOv9 (one-stage, recent)	Li et al. [55]	Power construction sites, adverse weather	Improved detection in fog/rain/snow	GPU, FPS varies	GPU, enhanced visibility modules	Robust in fog, rain, and snow; illumination not evaluated
YOLOv12 (one-stage, recent)	Ge et al. [56]	Safety-related behaviour detection	Improved accuracy via adaptive refinement	Real-time on GPU	GPU, deeper heads	Representative benchmark for the latest YOLO
CenterNet	Guo et al. [57]	Helmet detection	Improved mAP for small helmets	Real-time on GPU	GPU, moderate cost	Good for small PPE targets
Multiple	Cai et al. [58]	Safety CV baselines (PPE/unsafe behaviour)	High precision, lower complexity	Efficient, FPS varies	GPU, anchor-free efficiency	Stable across object-size variability
Mask R-CNN (complementary)	Fang et al. [59], Yang et al. [60]	Falls from height, crane safety	High segmentation accuracy	Sub-real-time, FPS not in abstract	High GPU demand	Strong under occlusion, slow on edge devices

2.3. Identified Research Gap and Research Objective

Despite the growing number of solutions using artificial intelligence algorithms to monitor occupational safety hazards, their practical implementation on construction sites faces numerous difficulties. The main barriers include the complexity of integration, the lack of standardised interfaces, and the limited availability of technology for smaller companies. The lack of simple and inexpensive systems that can be implemented in environments with limited technical resources is particularly acute.

Recent studies confirm that these challenges are not merely general observations but well-documented barriers in the implementation of computer vision systems for safety management in construction. Research to date [26,27] highlights the technical and organisational complexity of integrating AI-based detection modules with BIM, IoT, and site management platforms, largely due to the absence of standardised data exchange protocols and interoperability frameworks, which increases both the cost and the risk of deployment in heterogeneous environments. Furthermore, computational requirements and limited

hardware capabilities remain a significant obstacle for small and medium-sized enterprises, as many high-performance AI models are unsuitable for real-time inference on low-cost edge devices. However, recent works demonstrate that lightweight architectures such as YOLOv8s or MobileNet variants can maintain adequate detection accuracy while significantly reducing resource consumption, thus enabling feasible on-site deployment [43]. These findings reinforce the relevance of the proposed solution, which addresses the identified gaps by combining a lightweight detection model, open API integration, and event traceability mechanisms, ensuring both scalability and affordability for SMEs.

Unlike prior single-task YOLOv8 implementations, our system contributes a modular foundation in which detection, event logging, and alert propagation form a cohesive workflow. This foundation is required because current YOLOv8-based studies rarely address interoperability, real-time system latency, data persistence, or hardware-constrained deployment—all of which determine whether predictive or proactive safety modules can be practically introduced.

To address this research gap, a vision system based on the YOLO algorithm was designed to detect key hazards such as the absence of a safety helmet and employee falls. The developed system is also equipped with an open API and an incident recording module, enabling easy integration and documentation of events. Unlike previous single-task models, the proposed system integrates an open API, an incident-logging module, and a lightweight cloud architecture, forming a closed loop of ‘detection–alert–record–traceability’. Hardware requirements are reduced to ordinary edge devices, making the solution affordable for small and medium-sized contractors.

3. Materials and Methods

A schematic diagram of the research and development process for the developed system is presented in Figure 1.

The system development process began with defining the general design assumptions, including its functionality, technical requirements, and available hardware and software resources. At this stage, it was assumed that the system should enable automatic detection of two key hazards to the safety of workers on the construction site:

- Failure to wear a safety helmet by construction workers—no hardhat detection (NHD);
- Human fall detection—fall detection (FD).

If one of the above events is detected, the system should automatically record the incident in the database and immediately forward the relevant information to the supervisory team on the construction site (i.e., the site manager or health and safety inspector) by means of an appropriate alert. Additionally, it was assumed that the system would be able to integrate with the existing technical infrastructure via an API interface and operate via an intuitive web application.

In the next stage of the work, the system architecture was developed, covering both analytical components and technical infrastructure. The first step was to select a suitable real-time object detection algorithm. The YOLO (You Only Look Once) model was chosen, which, thanks to the use of deep neural networks, enables high detection accuracy with low computational delays. YOLOv8 was selected as the detection framework due to its stability and proven performance in industrial safety applications, which made it well-suited for the prototype implementation on low-cost edge devices [46,47].

For the system, functions were implemented to detect selected threats to work safety. These included the detection of the presence and colour of safety helmets (e.g., yellow, blue, white, or red) and the identification of situations indicating a person has fallen. These modules were integrated with the analytical part of the system, enabling automatic classification of events based on camera images.

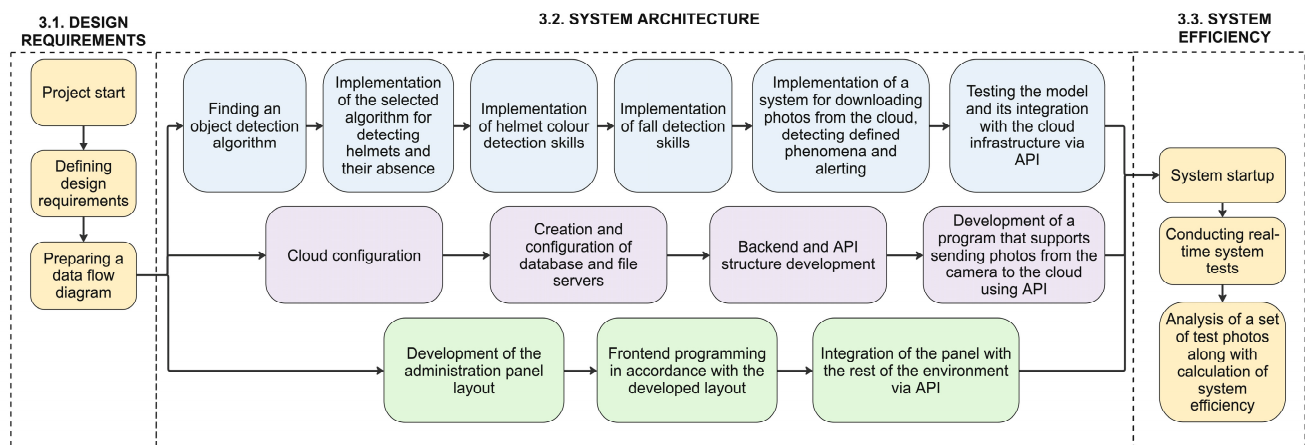


Figure 1. Stages of the design, implementation, and testing of the proprietary hazard detection system on a construction site.

At the same time, a cloud environment was configured, in which a database for recording detected incidents and disc space for storing video material and analysis results were launched. Communication between system components was also ensured through the implementation of a backend layer and the provision of an API interface. This interface enabled the transmission of images from surveillance cameras and integration with external administrative systems.

As part of the system development, an administrative panel accessible via a web browser has been envisaged. The user interface allowed for viewing recorded incidents, analysing statistics, and basic management of system parameters.

After implementation, functional tests were carried out in an environment similar to the real one. The tests included system initialisation, transfer of test images from cameras to the cloud, evaluation of the effectiveness of the detection algorithm, and the correctness of event recording.

3.1. Design Requirements

The design requirements for the proposed solution were divided into functional requirements (Table 4) and technological requirements (Table 5).

Table 4. Functional requirements of the system.

No	Task	Description
1	Object detection	Recognising the presence of people in an image, identifying whether employees are wearing safety helmets and, if so, the colour of the helmet, and detecting situations in which a person has fallen.
2	Alarm generation	Immediate notification of relevant persons in the event of undesirable events, such as the absence of a safety helmet or a person falling. Notifications can be sent via SMS, email, or dedicated applications.
3	Sharing results via API	Integration of external systems using REST API, or WebSocket for data and alarm transmission.
4	Event logging	Creation of an incident log with date, location, and type of event.

Table 5. System technology requirements.

No	Task	Description
1	AI framework	The system uses the YOLO algorithm. Python 3.12.0 with the OpenCV library can be used for implementation.
2	Server infrastructure	The solution can run on Linux, Windows, or macOS systems, and for greater flexibility, containerisation is recommended.
3	API	Choosing the right data storage system.
4	Database	Creating a detailed log containing information about detected events, their time, and location, which will enable incident analysis and reporting.
5	Optional user interface	An administration panel enabling system monitoring and event visualisation.
6	Video stream processing	The system operates in real time, analysing the data stream from the camera.

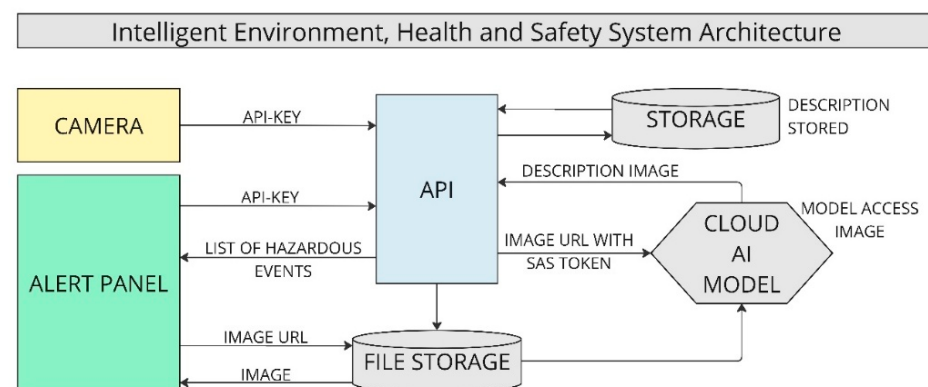
3.2. System Architecture

The data flow model developed in the proprietary construction site hazard detection system uses machine learning (ML) algorithms to analyse images in real time. The system aims to automatically monitor the working environment and identify potentially dangerous events based on defined behaviour patterns. The solution architecture includes key components such as a monitoring camera, an API programming interface, a data storage system, an ML model operating in a cloud environment, and an alarm panel. The data flow model is illustrated in Figure 2.

The process begins with the acquisition of image data from a camera observing a specific workspace. Communication between system components is provided by an API, which acts as a central transmission hub. The API also authorises access to data using a unique identification key (API Key), which increases transmission security.

The API interface transmits image data and provides so-called descriptors containing important information necessary for further processing, including:

- The location of files in the data store—images are stored on the server and accessible via unique URLs secured with a SAS (Shared Access Signature) token, enabling controlled and secure access;
- Processing status—descriptors contain HasAlert and ResolvedAlert flags, indicating whether a potential threat has been detected and whether a given frame has already been verified by the system operator.

**Figure 2.** Data flow model in the system architecture.

The developed system uses a RESTful API that enables communication between the monitoring module, the detection model, and the database. The communication between system components follows a sequence of asynchronous operations designed to ensure real-time performance and secure data handling.

The typical sequence of operations is as follows:

- (I) Image acquisition: the camera periodically captures an image frame and sends it to the backend server via an HTTP POST request.
- (II) Data validation: the API verifies the request parameters and stores the received image in the temporary data repository.
- (III) Detection request: the backend triggers the YOLOv8 detection module through an internal API call, passing the image reference and metadata.
- (IV) Analysis and classification: the detection module performs object detection (helmet absence/fall event) and returns a JSON response containing bounding box coordinates, class labels, and confidence scores.
- (V) Incident registration: if a safety violation is detected, the API automatically records the event in the incident database with time, camera ID, and event type.
- (VI) Alert notification: the system updates the web-based interface through a WebSocket channel, displaying the new event and enabling user acknowledgment.

Upon receiving a descriptor, the ML model retrieves images from the data warehouse and analyses them in a cloud environment. The model has been pre-trained on datasets containing patterns of dangerous behaviour (e.g., lack of a safety helmet). During the analysis, it compares new images with learned representations to identify undesirable situations.

Once a dangerous event has been identified, the developed programme flags the incident in the database, passing the appropriate status (e.g., “HasAlert”) via a POST request handled by the backend server. If they occur, a notification is generated on the panel, enabling immediate response by personnel. Automation of this process reduces the response time to potential incidents.

It is worth noting that the system also allows the collected data to be used to further improve the ML model, increasing its adaptability to the specific conditions of a given work environment. New cases can be used as additional training material, which has a positive impact on the accuracy of detection in the future.

The designed system uses two object detection models from the YOLO family:

- YOLO (keremberke/yolov8s-hard-hat-detection)—enabling the detection of the presence of safety helmets and their colour.
- YOLO (yolov8s.pt)—used to locate human silhouettes, which, in combination with geometric analysis, enables the identification of falls.
- A detailed block diagram of the detection algorithm is shown in Figure 3.

The video stream from the camera is analysed asynchronously. Every three seconds, the system extracts a single frame, which is then processed by the inference model using the *hardhat_processing()* and *falling_detecting()* functions.

The *hardhat_processing()* function identifies the presence of a safety helmet on the worker’s head. To do this, the image is subjected to object detection using the YOLO model. If a helmet is detected, its colour is analysed in the HSV space. If no helmet is detected, a “NO HARDHAT DETECTION” warning is generated. The algorithm allows for the simultaneous analysis of multiple figures within the camera’s field of view.

The *falling_detecting()* function focuses on detecting human falls. The width-to-height ratio threshold ($\text{width} \geq 2 \times \text{height}$) used to identify potential human falls was determined empirically during preliminary tests as a simplified heuristic for detecting horizontal body postures, aimed at ensuring computational efficiency in the prototype implementation. In

the first stage, all persons in the image are identified, and then the ratio of the width to the height of the area reserved for each of them is analysed. A heuristic has been adopted whereby if the width of the area is at least twice its height, this is interpreted as a lying position, potentially indicating a fall. This approach reduces the number of false alarms, e.g., those resulting from a kneeling position.

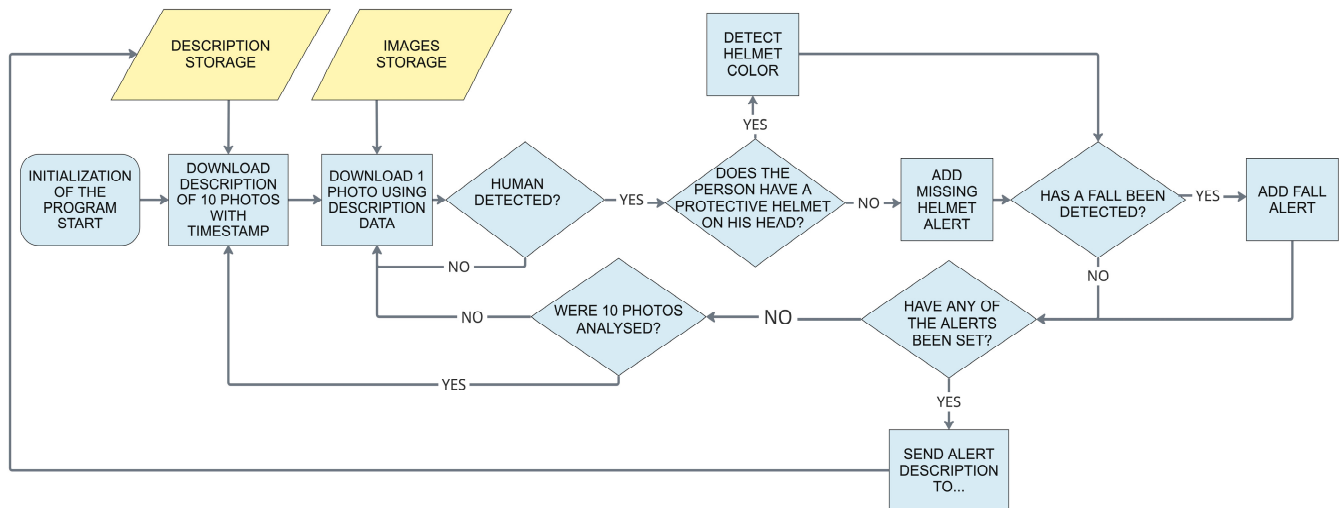


Figure 3. Block diagram of the system operation.

The system-level novelty lies in combining two independent YOLOv8-based modules with a low-overhead asynchronous inference loop, an incident-logging backend, and a real-time alerting interface, which collectively support deployment on ordinary edge devices. This distinguishes the present work from prior YOLOv8 studies focused exclusively on algorithmic accuracy without addressing deployment constraints, modularity, or interoperability.

3.3. System Efficiency

After developing the system, the key stage was its validation. For this purpose, a database was developed consisting of photos of people wearing protective helmets and people without protective helmets, as well as falls. The database included 100 photos, which were analysed individually. The dataset used in this study was collected in situ by the authors during controlled observation sessions conducted on an active construction site. All image data were gathered directly by the research team and manually annotated to identify the presence or absence of safety helmets and human falls.

The proprietary dataset used for model training and validation consisted of 200 RGB images (100 for helmet detection and 100 for fall detection) captured on an active construction site under varying lighting and weather conditions. The images were captured on an active construction site under varying lighting and weather conditions using a USB camera connected to a Raspberry Pi 3B+ microcomputer, which streamed 1080p video to the storage server. The camera used for data acquisition was a Razer Kiyo Full HD webcam. All images were manually annotated using the open-source Labelling 1.8.6. tool, and labels were saved in YOLO format (.txt files containing class ID and bounding box coordinates). The labelling process distinguished two main classes: *person_with_helmet/person_without_helmet* and *person_falling*. Given the exploratory scope of this study, the present experiments were conducted on the baseline dataset. In subsequent phases, we plan to extend the dataset and adopt a systematic augmentation pipeline (e.g., horizontal/vertical flips, small-angle rotations, photometric adjustments, and controlled noise injection) together with stratified train/validation/test protocols to preserve

class balance. This staged approach is intended to improve generalisation and robustness of the YOLOv8s model while maintaining methodological transparency to support reproducibility.

The significance of the classifiers in the analysed test is shown in Figure 4. The high number of true positive (TP) and true negative (TN) results indicates the high effectiveness of the model. On the other hand, the significant number of false positive (FP) and false negative (FN) results indicates the need for further optimisation of the model.

Classification by algorithm:

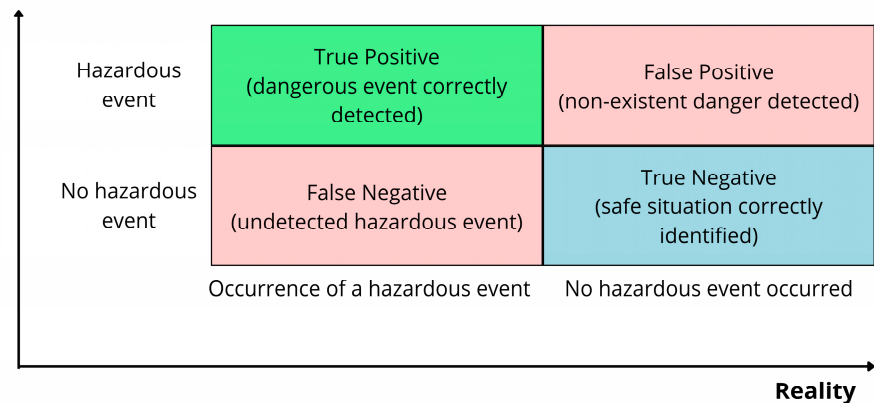


Figure 4. Classification matrix with definitions of effectiveness metrics.

The quality measures used are summarised in Table 6.

Table 6. Classification quality measures are used to evaluate the system.

No.	Measure	Description	Formula
1	Accuracy	The proportion of correctly classified cases (incidents and no incidents) in the entire set.	$\frac{TP + TN}{TP + FP + FN + TN}$
2	Precision	This is a measure of the ratio of correctly classified cases of helmet non-compliance to the total number of violations detected by the system.	$\frac{TP}{TP + FP}$
3	Recall	The ratio of correctly defined incidents of helmet non-use to all actual violations (detected and undetected).	$\frac{TP}{TP + FN}$
4	F1-score	Harmonic mean of precision and sensitivity.	$\frac{Precision \cdot Recall}{Precision + Recall}$
5	Specificity	Proportion of correctly recognised cases of no incident.	$\frac{TN}{TN + FP}$

4. Results

4.1. Detection of Protective Helmets

The results of alerts generated by the system are presented in Figure 5.

Each colour in the system corresponds to a specific category of objects. A blue field indicates the detection of a person. A green field indicates the detection of a safety helmet on an employee's head. In addition, the colour description of the identified safety helmet is displayed above the green field. A red field indicates the detection of a person not wearing a safety helmet, with a warning in the form of the message "NO HARDHAT DETECTION".



Figure 5. Image analysis with marked elements of the model's operation: (a) scene from an outdoor construction site; (b) scene from an outdoor construction site.

4.2. Human Fall Detection

Figure 6 shows the correct detection of a person's fall by the system.

In order to verify the effectiveness of the algorithm, a detailed analysis of one event from two different perspectives was carried out, as shown in Figure 7. The study aimed to assess the impact of perspective on the correctness of the algorithm and to identify potential limitations in its functioning.

In Figure 7a, the algorithm correctly classified the fall based on the position of the worker's body. However, in Figure 7b, showing the same moment of the event from a different perspective, the algorithm did not correctly classify the human fall.



Figure 6. Video output frames showing the expected classification change in the event of a human fall detection: (a) scene from an outdoor construction site; (b) scene from an outdoor construction site.



Figure 7. Analysis of the event detection system in the human fall category: (a) Correct event detection; (b) Failure to detect a human fall, even though an accident occurred.

4.3. User Interface

The developed intelligent health and safety monitoring system is an integrated demonstration solution combining image analysis, alert generation, and incident recording modules. Unlike approaches limited to passive image analysis, the system has been implemented as a coherent web platform based on cloud architecture. The designed infrastructure enables the launch of pre-trained detection models that identify people without protective helmets and falls on the construction site. All elements form a compatible whole. The final element of the solution is a user interface designed based on best practices in User Interface and User Experience. On a single screen, the user can access all alerts related to the recognition of safety helmets and their colours, which allows for the identification of the role of the employee on the construction site, as well as reports of falls. This allows the person verifying the information to easily and effectively analyse the alerts that have occurred on the construction site. It is an ideal solution for organisations that want to monitor work safety in real time.

The home page contains simple tabs and visual data designed to be accessible to the construction manager. The IoT interface provides access to both PCs and smartphones. A visualisation of the web application user interface is shown in Figure 8.

4.4. Evaluation Metrics

To quantitatively assess the effectiveness of the hazard detection system, a comparative analysis of the classification results with actual image data was performed. The verification was carried out on two independent test sets:

- No hardhat detection (NHD): 100 images;
- Fall detection (FD): 100 images.

Based on the classification results for each module, basic performance metrics were calculated: sensitivity (Recall), precision (Precision), accuracy (Accuracy), and F1-score. The calculations were based on the number of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). The evaluation focuses not only on detection accuracy but also on validating the end-to-end latency and operational behaviour of the integrated pipeline, which remains largely undocumented in previous YOLOv8-based OHS research. The results are summarised in Table 7.

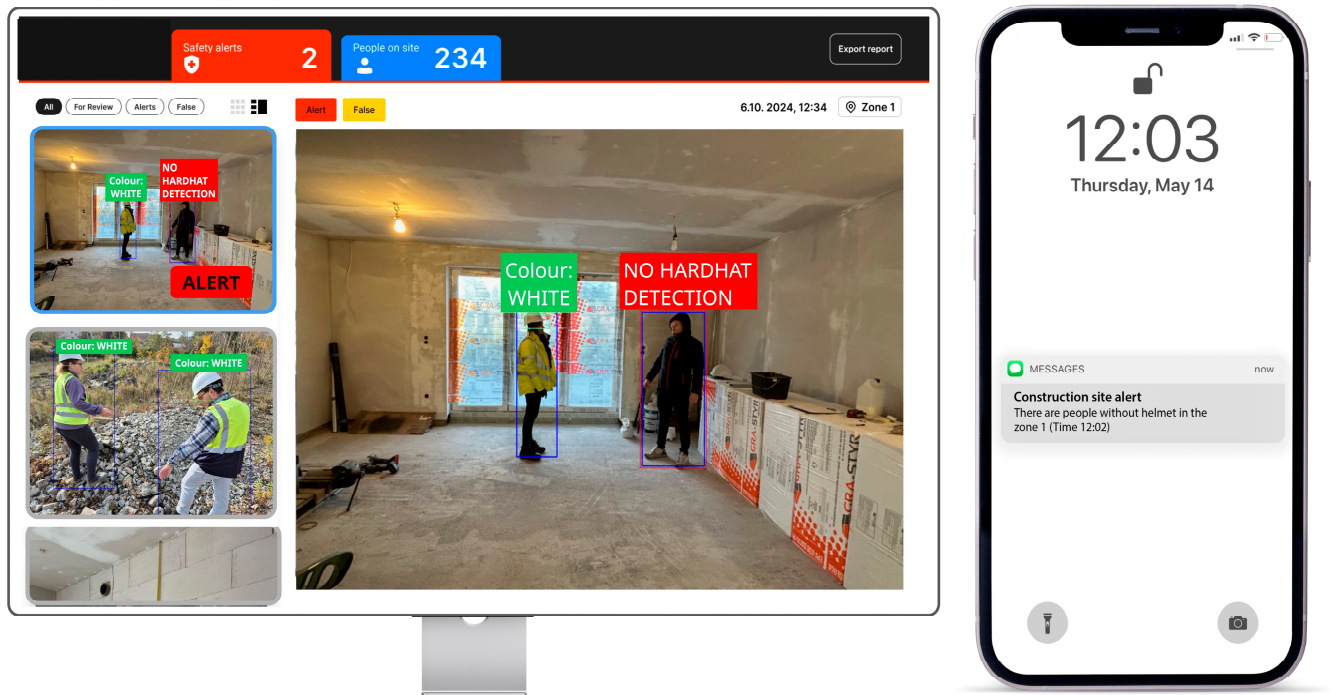


Figure 8. Visualisation of the user interface of the web application of the safety helmet and fall detection system on a construction site.

Table 7. Classification report for the test dataset.

	System Response Classification Results				System Quality Measures			
	TP	FP	TN	FN	Recall	Precision	Accuracy	F1-Score
NHD	39	3	108	18	0.68	0.93	0.88	0.79
FD	44	0	9	55	0.45	1.00	0.62	0.62

For the fall detection task, 87% of the samples represented actual fall events, while 13% depicted people standing or crouching, which occasionally resulted in misclassification. In the helmet detection task, 60% of the images showed workers without helmets, and 40% showed workers wearing proper protective equipment. The relatively small number of fall cases compared to the total number of frames introduced a class imbalance, which may have contributed to the lower recall observed in the fall-detection module.

All images in the test dataset were captured under daylight conditions; however, due to the relatively low recall observed in the fall-detection task, an additional analysis was conducted to identify potential sources of error. To assess the impact of lighting on detection accuracy, two illumination levels were defined: normal and low/uneven. The test dataset included samples captured from multiple viewpoints under varied illumination conditions to evaluate the model's robustness in real construction-site environments. The analysis of misclassified fall-detection cases revealed a relationship between detection accuracy, camera viewpoint, and lighting conditions. Table 8 summarises the number of misclassified samples grouped by these two factors.

The highest error rate was observed for oblique camera angles under uneven lighting, confirming the model's sensitivity to complex visual conditions. Beyond the counts in Table 8, misclassifications cluster into three recurring failure modes: (i) geometric ambiguity under oblique viewpoints, where the width–height proxy becomes non-discriminative; (ii) appearance degradation due to low/uneven illumination that suppresses texture/edge cues for small human silhouettes; and (iii) partial occlusion by materials or equipment.

These modes are consistent with construction-site vision literature [53,61,62], where attention mechanisms and small-object heads mitigate occlusion and scale issues, while low-light enhancements recover contrast for PPE/fall cues. In our prototype, these patterns primarily depress recall in FD while keeping precision high, aligning with the conservative alerting behaviour intended for early deployments.

Table 8. Number of misclassified samples grouped by camera viewpoint and illumination conditions.

Viewpoint	Illumination	
	Normal	Low/Uneven
Front	2	0
Side	6	4
Top	11	2
Oblique	18	13

The system performs frame acquisition, YOLO-based inference, and output rendering in parallel threads. Incoming frames are buffered locally when temporary delays or network interruptions occur, ensuring continuous operation without frame loss. To quantitatively verify the lightweight and real-time characteristics, average inference times were measured for 100 test images using both detection modules. The results are presented in Table 9.

Table 9. Average processing times for the asynchronous inference pipeline.

Task	Number of Samples	Mean Time [ms]	Min Time [ms]	Max Time [ms]
NHD	100	123.88	95.88	383.46
FD	100	131.84	88.46	363.13

5. Discussion

The evaluation showed that the developed hazard detection system on construction sites has varying effectiveness depending on the type of event analysed. In the safety helmet detection module (NHD), the system achieved high precision (0.93), which indicates a low level of false positives, and moderate sensitivity (0.68), indicating that some violations were missed. The F1-score (0.79) confirms that the model is well-balanced and effective in recognising irregularities related to the non-use of personal protective equipment. These effectiveness indicators directly address RQ1, confirming that the YOLOv8-based prototype achieves competitive precision and balanced performance for helmet-absence detection under the tested conditions. Similar relationships were observed in studies by Xing et al. [45] and Feng et al. [46], where high precision values also co-occurred with slightly lower sensitivity, especially in complex lighting conditions and with varying visibility of protective helmets.

The oblique-view and low-light sensitivities observed here translate into actionable controls: (i) prioritise camera placement with reduced obliqueness and stable heights to preserve discriminative body geometry; (ii) ensure illumination provisioning (task lighting or exposure control) at high-risk zones; and (iii) for fall detection, consider multi-camera overlap or short-clip analysis to recover temporal cues and suppress angle-specific misses. These measures are consistent with prior evidence [27] on construction-site vision systems and can be implemented without high-end hardware.

In the fall detection module (FD), the system demonstrated very high precision (1.00), which means a complete absence of false alarms. However, the relatively low sensitivity (0.45) indicates that a significant proportion of actual events were not detected correctly. The overall classification accuracy for this category was 0.62, which is partly due to the limited

size of the test dataset ($n = 100$). This phenomenon is consistent with the observations reported in Alam et al. [63], where the authors pointed out the low representation of rare classes (e.g., falls) as an important factor leading to the instability of recall metrics and potential overestimation of model performance. Taken together, the high precision but modest recall observed for falls constitutes a qualified answer to RQ1: the prototype avoids false alarms but misses a portion of true fall events, mainly due to data scarcity and viewpoint sensitivity. Although newer YOLO variants are emerging, YOLOv8 remains the most widely validated architecture with publicly replicable baselines in OHS-related tasks. More importantly, the novelty of this work is system-oriented: the method demonstrates how a lightweight YOLOv8 module can be embedded in an integrated safety pipeline under real deployment constraints. The detection stage thus serves not as a stand-alone contribution but as the enabling component of a full supervisory workflow.

In relation to RQ3, the heuristic assumption used in the system, according to which a fall is identified when the proportions of the bounding box around the human silhouette satisfy the condition that its width is at least twice its height, has a significant impact on the classification effectiveness. In situations where a falling person is recorded, e.g., from the front, the required proportions may not be met even though the event took place. An example of such a situation is shown in Figure 7b. In this case, the system did not generate an alert despite the actual fall of a person. This is a significant limitation of the currently used approach, which indicates the need for further development work.

Addressing RQ3, the analysis of single image frames is prone to errors resulting from a limited perspective, obstruction of the human silhouette, or an unusual camera angle. This phenomenon has also been described in other works, e.g., Wang et al. [42] noted that the uniqueness of human body position classification significantly increases when using multi-camera detection, which enables triangulation and a more complete analysis of posture. Therefore, extending the system with data from multiple image sources could significantly increase its effectiveness in engineering practice.

The identified limitations indicate potential directions for further optimisation. In particular, the use of more advanced spatio-temporal analysis methods, such as video frame sequence analysis or the implementation of recurrent neural architectures (e.g., LSTM, GRU), may enable better representation of posture dynamics and improve fall detection performance [64]. At the same time, expanding the training dataset, especially with different fall scenarios and camera perspectives, could improve the model's generalisation ability. These observations collectively answer RQ3, pinpointing camera geometry, temporal context, and class imbalance as the dominant levers for improving both accuracy and system throughput in subsequent iterations.

In summary, the system demonstrates high stability and practical usability in detecting static health and safety violations, such as the failure of construction workers to wear protective helmets. Its limitations in detecting dynamic events such as falls are mainly due to simplified detection assumptions and the limited scope of the test data. Despite these limitations, the high accuracy and operational readiness of the system confirm its implementation potential, especially in the context of real-time monitoring of working conditions on construction sites.

Prior studies [53,62,65] report that attention modules (e.g., CBAM/coordinate attention), small-object detection heads, and low-light enhancement routinely improve recall for PPE and human-posture cues in construction imagery while keeping real-time operation feasible on commodity GPUs. These modifications are orthogonal to our pipeline and can be incrementally adopted in subsequent iterations.

6. Conclusions

Based on the design and research work carried out and the analysis of the functional test results, the following conclusions were drawn:

- The developed system enables automatic detection of two key safety hazards on construction sites: failure to wear a safety helmet (NHD) and human falls (FD). For the NHD class, a classification accuracy of 0.88, precision of 0.93, and an F1 score of 0.79 were achieved, which indicates high effectiveness in detecting health and safety violations with a low false alarm rate. The FD class was characterised by full precision (1.00), but with low sensitivity (0.45), which indicates the system's limited ability to detect all falls and requires further development.
- The effectiveness of the system is significantly higher in the case of detecting static violations and clearly recognisable visual features, such as the absence of a safety helmet, than in the case of dynamic and difficult to capture phenomena, such as human falls. The results confirm that the effectiveness of fall detection strongly depends on the camera's viewing angle and the way the worker's body position is represented. Limiting the analysis to single image frames and using simplified classification heuristics (based on bounding box proportions) can lead to false negative results. The use of multi-camera systems or integration with more advanced image analysis methods (e.g., time sequence analysis, 3D positions estimation) can significantly improve the effectiveness of dynamic event detection.
- Thanks to its modular architecture and open API, the developed solution allows for flexible integration with various IT environments, which facilitates its adaptation to different implementation contexts. Thanks to the use of an API, a web application, and an automatic incident recording mechanism, the system can provide real-time support for health and safety supervision. This operational integration of detection, alerting, and event logging via an open API provides a concise, affirmative answer to RQ2, demonstrating near-real-time support for site supervision within the prototype.
- The results of the tests indicate significant implementation potential for the developed solution. In particular, the development of the system with video sequence analysis, the use of data from multiple image sources (e.g., panoramic cameras, drones), and integration with IoT sensors (e.g., employee positioning, overload detection) can significantly increase its functionality, adaptability, and effectiveness in real-world conditions.

From a theoretical standpoint, the conducted research contributes to the advancement of adaptable frameworks for real-time hazard detection based on deep learning and open API integration. The proposed architecture demonstrates how convolutional neural networks can be effectively utilised to identify occupational risks in dynamic and unstructured construction environments. From a practical perspective, the developed system confirms the feasibility of deploying AI-driven monitoring solutions in conditions characterised by limited computational resources and constrained infrastructure. Furthermore, the modular design of the platform provides a foundation for future extensions aimed at improving proactive safety management and reducing accident rates in the construction industry. In contrast to prior YOLOv8-based studies, the proposed solution provides an integrated and deployable pipeline combining detection, alert generation, and persistent event documentation through an open API, forming a system-level contribution that extends beyond algorithmic performance.

Regarding RQ1, the prototype met real-time constraints and achieved competitive effectiveness on helmet-absence while exhibiting conservative (high-precision, lower-recall) behaviour for falls under the tested conditions. Concerning RQ2, the system integrates detection, alerting, and event recording through an open REST API, enabling end-to-end monitoring and traceable incident management. For RQ3, accuracy and performance were chiefly driven by training data balance, camera viewpoint/coverage, illumination variabil-

ity, and single-frame inference limits, which directly motivate sequence-level modelling and multi-source sensing in future work.

7. Limitations

Despite the results confirming the functionality and usefulness of the developed system in engineering practice, this research has several significant limitations that should be taken into account in the further development of the solution.

Firstly, the use of single image frames as the basis for event classification makes it difficult to detect situations that require consideration of the temporal context, such as a sequence of movements leading to a fall. The lack of video sequence analysis reduces the accuracy of posture interpretation and makes it impossible to distinguish between potentially dangerous and harmless situations. This limitation also motivates a systems-level layer: disaster-chain analysis can prioritise cameras, sensors, and procedural barriers located at high-influence nodes of accident propagation, improving overall risk reduction despite per-event detection constraints [66].

Second, the limited number of examples of the “fall” class in the training and test sets significantly affects the model’s ability to generalise and leads to reduced detection performance in real-world situations. The uneven distribution of data across classes increases the risk of overfitting and generating unstable performance metrics such as recall or F1-score, a phenomenon widely described in the literature on rare event detection.

Another limitation is the system’s dependence on a stable and fast Internet connection, which is required to transfer data to the cloud environment. Transmission disruptions can lead to delays in incident detection and recording, which in critical situations can reduce the effectiveness of the response of construction supervision services. The system has limited resistance to environmental disturbances, in particular changing lighting conditions and dynamic backgrounds (e.g., machines, moving elements of the environment), which may affect the quality of object detection. The developed detection model does not take into account adaptive or compensatory mechanisms that would allow automatic adjustment of analysis parameters to external conditions.

8. Directions for Future Research

In the context of the limitations identified above, directions for further research on the development of the system have also been identified:

- Integration of time sequence analysis: The use of approaches based on video stream analysis (e.g., short clips instead of single frames) would allow for more accurate mapping of motion dynamics and improved detection of events such as falls. The use of recurrent architectures, such as LSTM (Long Short-Term Memory) or GRU (Gated Recurrent Unit), may increase detection sensitivity without compromising accuracy.
- Expansion of the dataset: Increasing the number of representative examples of events, especially in the FD class, and diversifying them in terms of scenarios, perspectives, and environmental conditions can significantly improve the model’s generalisation ability and reduce the risk of classification errors.
- Use of multi-camera systems and data from IoT sensors: Integrating data from multiple cameras and additional sensors (e.g., accelerometers, GPS locators, position sensors) would allow for triangulation of events, improved spatial detection accuracy, and increased system resilience to interference.
- Increased resilience to environmental factors: The implementation of adaptive mechanisms, such as dynamic exposure compensation, automatic detection threshold adjustment, or preliminary background filtering, could improve the system’s performance in changing lighting and environmental conditions.

- **Optimisation of processing infrastructure:** The use of a hybrid computing model would allow some detection operations to be shifted to local edge devices, reducing dependence on data transmission continuity and increasing system reliability in field conditions.

The above directions for development form the basis for further improvement of the system and its adaptation to more complex, dynamic, and demanding working environments characteristic of modern construction. According to the authors, the development of automatic hazard detection systems based on image analysis and artificial intelligence algorithms is an important direction of research in the context of occupational health and safety on construction sites. Such solutions can directly contribute to increasing the level of worker safety by detecting hazardous situations earlier, shortening the response time of supervisory services, and providing objective documentation of events. In the long term, their implementation may improve the safety culture in the construction industry and reduce the number of accidents at work.

Author Contributions: Conceptualization, Z.W., K.T., T.N., M.S. and F.Š.; methodology, Z.W., K.T.; software, Z.W., K.T. and F.Š.; validation, Z.W., K.T., M.S. and T.N.; formal analysis, Z.W., K.T., T.N. and M.S.; investigation, Z.W., K.T., T.N., M.S. and F.Š.; resources, Z.W., K.T., T.N., M.S. and F.Š.; data curation, Z.W., K.T., T.N. and M.S.; writing—original draft preparation, Z.W., K.T., T.N. and M.S.; writing—review and editing, Z.W., K.T., T.N. and M.S.; visualisation, Z.W., K.T., T.N., M.S. and F.Š.; supervision, Z.W. and T.N.; project administration, Z.W.; funding acquisition, Z.W. All authors have read and agreed to the published version of the manuscript.

Funding: The work was supported by the project Minigrants for doctoral students of the Wrocław University of Science and Technology.

Institutional Review Board Statement: Not Applicable.

Informed Consent Statement: Not Applicable.

Data Availability Statement: The data used in this study contain identifiable images of individuals and therefore cannot be made publicly available due to legal and ethical restrictions. Data may be provided for scientific purposes upon reasonable request to the corresponding author.

Acknowledgments: The authors of this article would like to thank the organisers of the AEC Hackathon Wrocław Edition 2024 event and the entire Whistleblowers Team, who participated in the development of the concept for an automatic work safety monitoring system on construction sites and in the creation of its initial prototype. The current version of the solution is still being developed by part of the original team with a view to implementation in real construction environment conditions.

Conflicts of Interest: Author Marta Stolarz was employed by the company Creoox AG and author Krzysztof Trybuszewski was employed by the company ITRIT. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Winge, S.; Albrechtsen, E.; Mostue, B.A. Causal Factors and Connections in Construction Accidents. *Saf. Sci.* **2019**, *112*, 130–141. [\[CrossRef\]](#)
2. International Labour Organization. *Safety and Health at Work: A Vision for Sustainable Prevention*; International Labour Organization: Frankfurt, Germany, 2014; ISBN 9789221289081.
3. Ammad, S.; Saad, S.; Bashir, M.T.; Qureshi, A.H.; Altaf, M.; Rasheed, K. Construction Accidents via Integrating Building Information Modelling (BIM) with Emerging Digital Technologies a Review. In Proceedings of the 2021 3rd International Sustainability and Resilience Conference: Climate Change, Online, 15–17 November 2021; pp. 460–463. [\[CrossRef\]](#)
4. Samovia, J. *Facts on Safety at Work*; International Labor Office (ILO): Geneva, Switzerland, 2021.
5. Zhou, Z.; Goh, Y.M.; Li, Q. Overview and Analysis of Safety Management Studies in the Construction Industry. *Saf. Sci.* **2015**, *72*, 337–350. [\[CrossRef\]](#)

6. Safety Audits. Inspections—Workplace Safety Market Outlook. Available online: <https://www.grandviewresearch.com/horizon/statistics/workplace-safety-market-outlook/workplace-safety-services/safety-audits-inspections/global> (accessed on 27 October 2025).
7. Construction Worker Safety Market. Global Market Analysis Report—2035. Available online: <https://www.futuremarketinsights.com/reports/construction-worker-safety-market> (accessed on 27 October 2025).
8. Park, H.J.; Lee, J.S.; Seo, M.B.; Lee, S.Y. Risk Analysis by Activity Based on Construction Safety Accident Cases. *J. Archit. Inst. Korea* **2024**, *40*, 287–295. [\[CrossRef\]](#)
9. Haupt, T.C.; Pillay, K. Investigating the True Costs of Construction Accidents. *J. Eng. Des. Technol.* **2016**, *14*, 373–419. [\[CrossRef\]](#)
10. Dobrowolska, E.; Gaca, B. Costs of Benefits for Accidents at Work and Occupational Diseases. Prevention as a Part of Their Cost Management. *Ubezpieczenia Społeczne. Teor. I Prakt.* **2023**, *157*, 1–16. [\[CrossRef\]](#)
11. Central Statistical Office. Accidents at Work. Available online: <https://stat.gov.pl/> (accessed on 15 October 2025).
12. Choudhry, R.M.; Fang, D.; Ahmed, S.M. Safety Management in Construction: Best Practices in Hong Kong. *J. Prof. Issues Eng. Educ. Pract.* **2008**, *134*, 20–32. [\[CrossRef\]](#)
13. Woźniak, Z.; Hoła, B. The Structure of near Misses and Occupational Accidents in the Polish Construction Industry. *Heliyon* **2024**, *10*, e26410. [\[CrossRef\]](#)
14. Woźniak, Z.; Hoła, B. Quantitative and Qualitative Analysis of Hazardous Events in the Construction Sector: Causes, Classification and Implications. *Bull. Pol. Acad. Sci. Tech. Sci.* **2025**, *73*, 154729. [\[CrossRef\]](#)
15. Olatoye, O.G.; Arewa, A.O.; Tann, D. Role of Human Factors in Fall From Height Fatalities in the UK Construction Industry. *Saf. Manag. Human Factors* **2024**, *151*, 136–146.
16. Park, M.; Kulinan, A.S.; Dai, T.Q.; Bak, J.; Park, S. Preventing Falls from Floor Openings Using Quadrilateral Detection and Construction Worker Pose-Estimation. *Autom. Constr.* **2024**, *165*, 105536. [\[CrossRef\]](#)
17. Chi, S.; Han, S. Analyses of Systems Theory for Construction Accident Prevention with Specific Reference to OSHA Accident Reports. *Int. J. Proj. Manag.* **2013**, *31*, 1027–1041. [\[CrossRef\]](#)
18. Atasoy, M.; Temel, B.A.; Basaga, H.B. A Study on the Use of Personal Protective Equipment among Construction Workers in Türkiye. *Buildings* **2024**, *14*, 2430. [\[CrossRef\]](#)
19. Oo, B.L.; Lim, B.T.H. Women Workforces' Satisfaction with Personal Protective Equipment: A Case of the Australian Construction Industry. *Buildings* **2023**, *13*, 959. [\[CrossRef\]](#)
20. Hoła, B.; Nowobilski, T.; Woźniak, Z.; Białko, M. Qualitative and Quantitative Analysis of the Causes of Occupational Accidents Related to the Use of Construction Scaffoldings. *Appl. Sci.* **2022**, *12*, 5514. [\[CrossRef\]](#)
21. United Nations. Transforming Our World: The 2030 Agenda for Sustainable Development Preamble. 2015. Available online: <https://sdgs.un.org/goals> (accessed on 15 October 2025).
22. Sachs, J.D.; Lafortune, G.; Kroll, C.; Fuller, G.; Woelm, F. Includes the SDG Index and Dashboards. In *Sustainable Development Report 2022*; Cambridge University Press: Cambridge, UK, 2022. [\[CrossRef\]](#)
23. Szer, I.; Błazik Borowa, E.; Szer, J. The Influence of Environmental Factors on Employee Comfort Based on an Example of Location Temperature. *Arch. Civ. Eng.* **2017**, *63*, 163–174. [\[CrossRef\]](#)
24. Jabłoński, M.; Szer, I.; Szer, J. Probability of Occurrence of Health and Safety Risks on Scaffolding Caused by Noise Exposure. *J. Civ. Eng. Manag.* **2018**, *24*, 437–443. [\[CrossRef\]](#)
25. Szer, I.; Szer, J.; Czarnocki, K.; Błazik Borowa, E. Apparent Temperature Distribution on Scaffoldings during Construction Works. *Int. J. Med. Health Sci.* **2018**, *5*, 81–87.
26. Fang, W.; Love, P.E.D.; Luo, H.; Ding, L. Computer Vision for Behaviour-Based Safety in Construction: A Review and Future Directions. *Adv. Eng. Inform.* **2020**, *43*, 100980. [\[CrossRef\]](#)
27. Jiang, P.; Ergu, D.; Liu, F.; Cai, Y.; Ma, B. A Review of Yolo Algorithm Developments. *Procedia Comput. Sci.* **2022**, *199*, 1066–1073. [\[CrossRef\]](#)
28. Uma, M.; Abirami, S.; Ambika, M.; Kavitha, M.; Sureshkumar, S.; Kaviyaraj, R. A Review on Augmented Reality and YOLO. In *Proceedings of the 4th International Conference on Smart Electronics and Communication, ICOSEC 2023, Trichy, India, 20–22 September 2023*; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2023; pp. 1025–1030. [\[CrossRef\]](#)
29. Liang, J. A Review of the Development of YOLO Object Detection Algorithm. *Appl. Comput. Eng.* **2024**, *71*, 39–46. [\[CrossRef\]](#)
30. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016*; pp. 779–788. [\[CrossRef\]](#)
31. Wei, J.; As'array, A.; Anas Md Rezali, K.; Zuhri Mohamed Yusoff, M.; Ma, H.; Zhang, K. A Review of YOLO Algorithm and Its Applications in Autonomous Driving Object Detection. *IEEE Access* **2025**, *13*, 93688–93711. [\[CrossRef\]](#)
32. Terven, J.; Córdova-Esparza, D.M.; Romero-González, J.A. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Mach. Learn. Knowl. Extr.* **2023**, *5*, 1680–1716. [\[CrossRef\]](#)

33. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single Shot Multibox Detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Volume 9905, pp. 21–37. [\[CrossRef\]](#)
34. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal Loss for Dense Object Detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2999–3007. [\[CrossRef\]](#)
35. Woźniak, Z. Current Research Directions on the Application of Artificial Intelligence in Construction in the Context of Occupational Safety. *Mater. Bud.* **2025**, *1*, 21–28. [\[CrossRef\]](#)
36. Dai, Z. Image Acquisition Technology for Unmanned Aerial Vehicles Based on YOLO—Illustrated by the Case of Wind Turbine Blade Inspection. *Syst. Soft Comput.* **2024**, *6*, 200126. [\[CrossRef\]](#)
37. Chen, L.; Chen, W.; Wang, L.; Zhai, C.; Hu, X.; Sun, L.; Tian, Y.; Huang, X.; Jiang, L. Convolutional Neural Networks (CNNs)-Based Multi-Category Damage Detection and Recognition of High-Speed Rail (HSR) Reinforced Concrete (RC) Bridges Using Test Images. *Eng. Struct.* **2023**, *276*, 115306. [\[CrossRef\]](#)
38. Sirisha, M. Automated Object Detection and Tracking for Construction Site Safety. *Int. J. Res. Appl. Sci. Eng. Technol.* **2024**, *12*, 23–29. [\[CrossRef\]](#)
39. Wu, Z.; Lei, X.; Kumar, M. Advancing Construction Safety: YOLOv8-CGS Helmet Detection Model. *PLoS ONE* **2025**, *20*, e0321713. [\[CrossRef\]](#)
40. Lin, B. Safety Helmet Detection Based on Improved YOLOv8. *IEEE Access* **2024**, *12*, 28260–28272. [\[CrossRef\]](#)
41. Qin, Y.; Miao, W.; Qian, C. A High-Precision Fall Detection Model Based on Dynamic Convolution in Complex Scenes. *Electronics* **2024**, *13*, 1141. [\[CrossRef\]](#)
42. Wang, H.; Xu, S.; Chen, Y.; Su, C. LFD-YOLO: A Lightweight Fall Detection Network with Enhanced Feature Extraction and Fusion. *Sci. Rep.* **2025**, *15*, 5069. [\[CrossRef\]](#)
43. Huang, X.; Li, X.; Yuan, L.; Jiang, Z.; Jin, H.; Wu, W.; Cai, R.; Zheng, M.; Bai, H. SDES-YOLO: A High-Precision and Lightweight Model for Fall Detection in Complex Environments. *Sci. Rep.* **2025**, *15*, 2026. [\[CrossRef\]](#)
44. Pereira, G.A. Fall Detection for Industrial Setups Using YOLOv8 Variants. *arXiv* **2024**, arXiv:2408.04605. [\[CrossRef\]](#)
45. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Feng, R.; Miao, Y.; Zheng, J. A YOLO-Based Intelligent Detection Algorithm for Risk Assessment of Construction Sites. *J. Intell. Constr.* **2024**, *2*, 1–18. [\[CrossRef\]](#)
47. Sina, F. *Enhancing Construction Safety Behavior: AI-Based Strategies for Workplace Hazard Prevention and Risk Mitigation*; Research Gate: Berlin, Germany, 2025. [\[CrossRef\]](#)
48. Sabir, A.; Hussain, R.; Pedro, A.; Park, C. Personalized Construction Safety Training System Using Conversational AI-Based Virtual Reality. 2024. Available online: <https://ssrn.com/abstract=5043158> (accessed on 10 November 2025).
49. Mohamed, M.A.H.; Al-Mhdawi, M.K.S.; Dacre, N.; Ojiako, U.; Qazi, A.; Rahimian, F. Theoretical and Practical Instantiations of Generative AI in Construction Risk Management: An Analytical Exposition of Its Latent Benefits and Inherent Risks. 2024. Available online: <https://ssrn.com/abstract=5007208> (accessed on 10 November 2025).
50. 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. [\[CrossRef\]](#)
51. Fang, Q.; Li, H.; Luo, X.; Ding, L.; Luo, H.; Rose, T.M.; An, W. Detecting Non-Hardhat-Use by a Deep Learning Method from Far-Field Surveillance Videos. *Autom. Constr.* **2018**, *85*, 1–9. [\[CrossRef\]](#)
52. Wang, Z.; Cai, Z.; Wu, Y. An Improved YOLOX Approach for Low-Light and Small Object Detection: PPE on Tunnel Construction Sites. *J. Comput. Des. Eng.* **2023**, *10*, 1158–1175. [\[CrossRef\]](#)
53. An, Q.; Xu, Y.; Yu, J.; Tang, M.; Liu, T.; Xu, F. Research on Safety Helmet Detection Algorithm Based on Improved YOLOv5s. *Sensors* **2023**, *23*, 5824. [\[CrossRef\]](#)
54. Wang, S. Automated Non-PPE Detection on Construction Sites Using YOLOv10 and Transformer Architectures for Surveillance and Body Worn Cameras with Benchmark Datasets. *Sci. Rep.* **2025**, *15*, 27043. [\[CrossRef\]](#)
55. Li, J.; Feng, Y.; Shao, Y.; Liu, F. IDP-YOLOV9: Improvement of Object Detection Model in Severe Weather Scenarios from Drone Perspective. *Appl. Sci.* **2024**, *14*, 5277. [\[CrossRef\]](#)
56. Ge, T.; Ning, B.; Xie, Y. YOLO-AFR: An Improved YOLOv12-Based Model for Accurate and Real-Time Dangerous Driving Behavior Detection. *Appl. Sci.* **2025**, *15*, 6090. [\[CrossRef\]](#)
57. Guo, Y.; Lu, X. ST-CenterNet: Small Target Detection Algorithm with Adaptive Data Enhancement. *Entropy* **2023**, *25*, 509. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Cai, R.; Li, J.; Tan, Y.; Tang, J.; Chen, X. Convolutional Neural Networks for Construction Safety: A Technical Review of Computer Vision Applications. *Appl. Soft Comput.* **2025**, *180*, 113374. [\[CrossRef\]](#)
59. Fang, W.; Zhong, B.; Zhao, N.; Love, P.E.D.; Luo, H.; Xue, J.; Xu, S. A Deep Learning-Based Approach for Mitigating Falls from Height with Computer Vision: Convolutional Neural Network. *Adv. Eng. Inform.* **2019**, *39*, 170–177. [\[CrossRef\]](#)

60. Yang, Z.; Yuan, Y.; Zhang, M.; Zhao, X.; Zhang, Y.; Tian, B. Safety Distance Identification for Crane Drivers Based on Mask R-CNN. *Sensors* **2019**, *19*, 2789. [[CrossRef](#)]
61. Park, M.; Tran, D.Q.; Bak, J.; Park, S. Small and Overlapping Worker Detection at Construction Sites. *Autom. Constr.* **2023**, *151*, 104856. [[CrossRef](#)]
62. Liu, Y.; Jiang, B.; He, H.; Chen, Z.; Xu, Z. Helmet Wearing Detection Algorithm Based on Improved YOLOv5. *Sci. Rep.* **2024**, *14*, 8768. [[CrossRef](#)]
63. Alam, E.; Sufian, A.; Dutta, P.; Leo, M. Vision-Based Human Fall Detection Systems Using Deep Learning: A Review. *Comput. Biol. Med.* **2022**, *146*, 105626. [[CrossRef](#)]
64. Lin, C.B.; Dong, Z.; Kuan, W.K.; Huang, Y.F. A Framework for Fall Detection Based on OpenPose Skeleton and LSTM/GRU Models. *Appl. Sci.* **2021**, *11*, 329. [[CrossRef](#)]
65. Ren, H.; Fan, A.; Zhao, J.; Song, H.; Liang, X. Lightweight Safety Helmet Detection Algorithm Using Improved YOLOv5. *J. Real. Time Image Process* **2024**, *21*, 125. [[CrossRef](#)]
66. Su, C.; Ma, J.; Wang, C.; Deng, J.; Chen, W. Flood-Induced Coal Mine Disaster Chain Evolution and Risk Analysis. *Nat. Hazards* **2025**, *121*, 21031–21058. [[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.