INTERNATIONAL JOURNAL OF INFORMATIVE & FUTURISTIC RESEARCH (IJIFR)
ISSN: 2347-1697
VOLUME 13||ISSUE 8|| APRIL 2026
An Open Access, Peer Reviewed, Cross Reference DOI & Google Scholar Indexed International Research Journal
UGC Approved || IJIFR Impact Factor= 8.057
Estd Year: 2013 || Email: editors@ijifr.org
Availaible Online at: https://www.ijifr.org/

DOI: https://doi.org/10.64672/IJIFR/26.04.13.08.038 PUBLISHED ON: APRIL 20, 2026

SAFEHELMET VISION AI: REAL-TIME CONSTRUCTION SITE SAFETY HELMET DETECTION USING YOLOV8 TRANSFER LEARNING AND STREAMLIT DEPLOYMENT

Kondakavali Vani [1], V.Vijayalakshmi [2] , S. Usharani [3]
[1]M.C.A. Student, [2] Assistant Professor, [3] Professor
[1,2,] Department of Computer Applications,
Viswam Engineering College, Madanapalle, Andhra Pradesh, India

ABSTRACT

Construction sites consistently rank among the most hazardous occupational environments worldwide, with head injuries from falling or flying objects identified as a primary contributor to construction fatalities in every major market. Despite the universal regulatory mandate for safety helmet usage, non-compliance remains pervasive owing to the practical impossibility of maintaining continuous manual supervision across large, complex sites. Traditional automated monitoring approaches based on conventional computer vision techniques have demonstrated insufficient accuracy for reliable deployment in the visually complex conditions typical of active construction environments, while commercial AI-based platforms impose subscription costs prohibitive to small and medium contractors.This paper presents SafeHelmet Vision AI, a deep learning-based industrial safety monitoring system designed to automate helmet compliance detection at construction sites and related industrial workplaces. The proposed system employs a YOLOv8n (nano) object detection model trained via transfer learning from COCO-pretrained weights on a domain-specific dataset of 4,200 annotated construction site images encompassing 9,800 helmet and 8,200 worker bounding box instances. Training was conducted for 100 epochs with AdamW optimisation (lr_0 = 0.001), comprehensive data augmentation including mosaic, HSV perturbation, random flip, rotation (±10°), and scale variation (±50%), yielding a validation mean Average Precision at IoU = 0.50 (mAP50) of 97.8%, a precision of 95.2%, and a recall of 94.9%.The trained model is integrated into a Streamlit web application that accepts uploaded construction site images in JPG, PNG, or BMP formats and returns annotated detection results with bounding boxes, class labels, and confidence scores within a mean inference latency of 165 milliseconds on standard CPU hardware. An automated safety compliance assessment engine evaluates helmet-to-person count ratios and generates colour-coded violation alerts. Comparative evaluation demonstrates a 27.2 percentage-point precision advantage and a 33.9 percentage-point recall advantage over a traditional Haar cascade baseline. The complete system requires no client-side installation and is deployable to Streamlit Cloud from a GitHub repository with a single configuration step, making enterprise-grade safety monitoring accessible to safety personnel without specialised technical training..

KEYWORDS ***Safety helmet detection; YOLOv8; construction site safety; personal protective equipment; transfer learning; real-time object detection; Streamlit deployment***

PAPER CITATION:
Vani, K., Vijayalakshmi, V., Usharani, S..: " Safehelmet Vision Ai: Real-Time Construction Site Safety Helmet Detection Using YOLOV8 Transfer Learning And Streamlit Deployment", International Journal of Informative & Futuristic Research (IJIFR), Vol. (13) (8), April 2026, pp. 1189-1196,https://doi.org/10.64672/IJIFR/26.04.13.08.038

1. INTRODUCTION

According to the International Labour Organization [10], construction is consistently ranked among the three most hazardous industries globally in terms of occupational fatalities and serious injuries. Head injuries from falling or flying objects represent the leading cause of fatal construction incidents, accounting for a disproportionate share of the approximately 60,000 fatal occupational accidents occurring on construction sites worldwide each year. The safety helmet, universally mandated by occupational health and safety regulations across all major jurisdictions, provides critical protection against this primary fatality mechanism — yet non-compliance remains a persistent, pervasive, and largely unsolved problem across the global construction industry.The fundamental inadequacy of existing compliance monitoring approaches stems from their reliance on manual supervision by safety officers conducting periodic site walkthroughs. The practical impossibility of maintaining continuous visual monitoring across all workers on large, complex construction sites creates systematic gaps in coverage that workers readily exploit. Research consistently demonstrates that helmet compliance rates drop markedly when workers believe they are unobserved, suggesting that only the credible threat of continuous, comprehensive monitoring can achieve sustained compliance at the population level.

The maturation of deep learning-based object detection algorithms, culminating in the YOLO family of single-stage detectors, has created the technical foundation for automated visual monitoring systems capable of continuously and reliably detecting safety equipment compliance across entire construction sites through camera feed analysis. The YOLOv8 architecture [2], the current state of the art in real-time object detection, achieves exceptional accuracy-speed trade-offs through its anchor-free detection head, C2f backbone module, and decoupled classification-regression design — properties that are precisely suited to the requirements of construction site safety monitoring.Existing AI-based safety monitoring platforms from companies such as Smartvid.io, Intenseye, and Voxel AI demonstrate the commercial viability of this approach but are priced as enterprise software with annual subscription costs prohibitive to the majority of small and medium-sized construction contractors that collectively account for the largest share of construction site accidents. A freely deployable, open-source alternative would democratise access to AI-powered safety monitoring and could achieve substantially greater population-level safety impact than expensive proprietary platforms accessible only to large organisations.

SafeHelmet Vision AI addresses this opportunity by delivering a high-performance YOLOv8-based helmet detection system in a simple, accessible Streamlit web application that can be deployed and operated by safety personnel without specialised technical training. The contribution of this paper is threefold: (1) the training and empirical validation of a YOLOv8n transfer learning model achieving mAP50 = 97.8% on a 4,200-image domain-specific construction site dataset; (2) the design of a four-module Streamlit application architecture integrating model inference, safety compliance assessment, and multi-format result presentation within a sub-200 ms end-to-end response time on CPU hardware; and (3) a comprehensive comparative evaluation against traditional computer vision and competing deep learning baselines quantifying the performance advantages of the proposed approach.

2. LITERATURE SURVEY

The automated detection of construction site safety equipment has attracted substantial research attention over the past decade, driven by the recognised inadequacy of manual monitoring and the progressive maturation of deep learning tools accessible to applied researchers.

Redmon et al. [1] introduced the foundational You Only Look Once (YOLO) architecture in 2016, establishing the single-stage detection paradigm that fundamentally transformed the accuracy-speed trade-off landscape for real-time object detection. By reformulating detection as a single regression problem predicting class probabilities and bounding box coordinates from full-image convolutional features in a single forward pass, YOLO eliminated the computational overhead of the region proposal step inherent in two-stage detectors such as Faster R-CNN. This architectural innovation established the technical foundation upon which all subsequent YOLO generations, including YOLOv8, are built.

Nath et al. [4] provided a comprehensive empirical evaluation of deep learning approaches for construction site PPE detection, training and comparing Faster R-CNN, SSD, and YOLOv3 models on a dataset of 1,237 construction site images. Their results demonstrated clear superiority of deep learning over traditional computer vision across all evaluation conditions, with YOLOv3 achieving the best accuracy-speed balance at 74.3% mAP. The authors identified dataset size and diversity as the primary bottleneck for further performance improvement, motivating the larger 4,200-image training corpus assembled for SafeHelmet Vision AI.

Wu et al. [5] made a specific foundational contribution by constructing and publicly releasing a large annotated benchmark dataset for hardhat detection containing over 7,000 images from diverse construction environments, establishing a common evaluation standard for subsequent work. Their YOLOv3-based baseline achieved 86.2% mAP50 on this dataset. Shen et al. [6] subsequently demonstrated that architectural improvements introduced in YOLOv5, including the Path Aggregation Network neck, improved anchor assignment, and enhanced augmentation strategies, yielded a further 8.4 percentage-point mAP50 improvement over the YOLOv3 baseline on equivalent evaluation data.

Jocher et al. [2] introduced YOLOv8 in January 2023 as a comprehensive redesign incorporating anchor-free detection, C2f backbone modules providing richer gradient flow, and a decoupled head separating classification and regression branches. On the COCO object detection benchmark, YOLOv8n achieves 37.3 mAP at 80.4 fps on GPU — a 2.1 mAP improvement over YOLOv5n at comparable speed — establishing it as the appropriate architecture choice for the SafeHelmet Vision AI application where both accuracy and inference speed are critical.

Lin et al. [3] introduced the Microsoft COCO dataset and evaluation metrics, including the mAP50 measure used as the primary performance indicator in this work. The COCO pre-trained weights used as the initialisation for SafeHelmet Vision AI's transfer learning incorporate feature representations learned from 330,000 images spanning 80 object categories, providing a rich foundation of general visual features that substantially reduces the domain-specific training data requirement. He et al. [12] provided the theoretical basis for deep residual learning that underpins the backbone architectures of the YOLO family, demonstrating that skip connections enable training of much deeper networks by mitigating the vanishing gradient problem.

The surveyed literature establishes three key findings that directly motivate the design of SafeHelmet Vision AI: (i) YOLOv8 represents the current state-of-the-art architecture for real-time object detection, offering the best available accuracy-speed trade-off; (ii) transfer learning from COCO-pretrained weights dramatically reduces the training data requirement for domain-specific detection tasks, enabling high-performance models with moderately sized datasets; and (iii) the accessibility gap between research-grade implementations and practically deployable tools for safety personnel remains a significant barrier to real-world impact that the Streamlit-based deployment strategy of SafeHelmet Vision AI directly addresses.

3. PROPOSED WORK AND METHODOLOGY

The SafeHelmet Vision AI system is organised as a three-tier architecture — User Interaction Tier, Application Logic Tier, and AI Inference Tier — implemented within a Streamlit web application backed by a pre-trained YOLOv8n model. The system design prioritises the simultaneous achievement of high detection performance, sub-200 ms end-to-end response time on CPU hardware, and zero-configuration accessibility for non-technical safety personnel.

3.1 System Architecture

The proposed architecture leverages a hierarchically decomposed three-tier design with the following layer responsibilities:

Layer 1 — User Interaction Tier: A Streamlit web application rendered in the user's browser provides the image upload interface, progress feedback, and results display. The reactive execution model of Streamlit automatically re-renders affected UI components on state changes, eliminating explicit event handling code. The application is accessible from any modern browser without client-side installation, supporting deployment both on local machines (localhost:8502) and cloud-hosted environments via Streamlit Cloud over HTTPS.

Layer 2 — Application Logic Tier: The Python backend orchestrates the complete image analysis workflow. Uploaded files are decoded via PIL.Image.open() and converted to NumPy arrays (H×W×3, uint8). The @st.cache_resource decorator caches the loaded YOLO model object across script reruns, ensuring the 2-second model loading operation occurs only once per session. The compliance assessment engine processes detection results and generates structured safety status outputs.

Layer 3 — AI Inference Tier: The YOLOv8n model stored as best.pt (≈6 MB, PyTorch checkpoint format) accepts NumPy image arrays, internally handles resizing to 640×640 pixels and pixel normalisation, executes a single-pass forward inference through the neural network, applies non-maximum suppression (IoU threshold = 0.45, confidence threshold = 0.25), and returns a Results object containing bounding box coordinates, class identifiers, and confidence scores alongside an annotated image generated by the plot() method.

Layer 4 — Model Storage and Distribution: The best.pt checkpoint is hosted on GitHub Releases and downloaded automatically by the load_model function on first application startup if not present locally. This distribution architecture enables zero-configuration cloud deployment on Streamlit Cloud while maintaining support for fully local, air-gapped deployment in organisations with strict data privacy requirements.

3.2 YOLOv8 Architecture and Training Methodology

The YOLOv8n backbone employs C2f (Cross Stage Partial with 2 convolutions and feature fusion) modules that provide denser gradient connections during training compared to the C3 modules of YOLOv5. Multi-scale feature extraction at strides of 8, 16, and 32 pixels relative to the 640×640 input produces feature maps at three spatial resolutions capturing fine-grained texture details through high-level semantic features. A Path Aggregation Network (PAN) neck aggregates top-down feature pyramid connections with bottom-up path augmentation, ensuring both fine-grained and high-level features are available at each detection scale. The anchor-free detection head directly predicts centre coordinates, width, and height as continuous values relative to grid cell locations, eliminating the anchor hyperparameter tuning required by earlier YOLO versions.

Transfer learning from COCO-pretrained YOLOv8n weights was employed to initialise the backbone feature extraction layers. Fine-tuning on the SafeHelmet dataset adapted these general feature representations to the specific visual characteristics of construction site environments. Training ran for 100 epochs with early stopping (patience = 20 epochs) using AdamW optimisation ($lr_0 = 0.001$, lrf = 0.01, momentum = 0.937, weight_decay = 0.0005) and 3 warmup epochs. Input resolution was fixed at 640×640 pixels with batch size 16.

Data augmentation applied during training included: mosaic augmentation (probability = 1.0) combining four training images into composite inputs; random horizontal and vertical flipping (p = 0.5); HSV colour jitter (H ± 0.015, S ± 0.7, V ± 0.4); random rotation (±10°); random translation (±10% per axis); and random scaling (±50%). This augmentation strategy artificially expanded the effective diversity of the 3,500-image training split, substantially improving generalisation to unseen site conditions.

3.3 Safety Compliance Assessment Logic

The compliance assessment engine implements a deterministic rule over the helmet count H and person count P extracted from the YOLOv8 detection results. Three mutually exclusive outcomes are evaluated: (i) if $H < P$, at least $(P - H)$ workers are detected without helmets, triggering a critical violation alert rendered with red error styling; (ii) if H ≥ P and P > 0, all detected workers are wearing helmets, triggering a compliance success message rendered with green styling; (iii) if P = 0, no workers are detected and an informational message is displayed. This logic is designed to be conservative: the system errs toward false positive alerts (over-reporting violations) rather than false negatives (missing violations), reflecting the asymmetric cost of missing a safety-critical event.

Algorithm 1: SafeHelmet Vision AI — End-to-End Inference and Compliance Assessment Pipeline

```
INPUT:  uploaded_image_file (JPG / PNG / BMP)
OUTPUT: annotated_image, helmet_count, person_count, compliance_status
// ── Model Initialisation (once per session, cached) ──────────────
IF model NOT in cache THEN
    IF best.pt NOT on disk THEN download best.pt from GitHub Releases
    model ← YOLO('best.pt')            // load YOLOv8n checkpoint
    cache(model)                       // @st.cache_resource
// ── Per-Image Inference
──────────────────────────────────────────────────────────────────
pil_img   ← PIL.Image.open(uploaded_file)
img_array ← numpy.array(pil_img)        // shape: [H, W, 3] uint8
results   ← model(img_array,
                  conf=0.25, iou=0.45)  // single-pass YOLOv8 inference
annotated ← results[0].plot()           // BGR annotated NumPy array
// ── Detection Count Extraction
──────────────────────────────────────────────────────────
helmet_count ← 0;  person_count ← 0
FOR each box IN results[0].boxes DO
    cls_id ← int(box.cls[0])
    IF cls_id == 0: helmet_count += 1    // class 0 → helmet
    IF cls_id == 1: person_count += 1    // class 1 → person
// ── Compliance Assessment
──────────────────────────────────────────────────────────────
IF person_count == 0:
    status ← 'NO_PERSONS_DETECTED'       // informational
ELSE IF helmet_count < person_count:
    violations ← person_count − helmet_count
    status ← 'VIOLATION — ' + violations + ' worker(s) without helmet'
ELSE:
    status ← 'COMPLIANT — all detected workers wearing helmets'
RENDER two-column display: [original image | annotated image]
RENDER detection summary: helmet_count, person_count, status
RETURN annotated, helmet_count, person_count, status
```

4. RESULTS AND DISCUSSION

The trained SafeHelmet Vision AI model was evaluated on a held-out validation dataset of 500 construction site images withheld from training, supplemented by a 200-image test set drawn from construction environments not represented in the training corpus. The validation and test sets collectively encompassed diverse lighting conditions (natural daylight, artificial site lighting, overcast conditions, and direct sun glare), camera elevations (ground level, elevated platform, overhead crane-mounted), helmet colours (yellow, white, orange, blue, red, black, silver), and scene complexity levels ranging from single workers in open areas to crowded scaffolding scenes with partial occlusions.

Performance metrics were computed per class (helmet, person) and as overall mAP50 across both classes. Comparative evaluation was conducted against three baseline systems: a traditional Haar cascade head detector with HSV-based helmet colour filtering, a YOLOv5n model trained on the same dataset under identical conditions, and a commercial cloud-based PPE detection API evaluated on the same test images.

Table 1: Detection Performance Comparison — SafeHelmet Vision AI vs. Baseline Methods

Method	mAP50 (%)	Precision (%)	Recall (%)	F1-Score (%)	Inference (ms/img)
Haar Cascade + HSV Filter (Traditional)	61.4	68.0	61.0	64.3	~18
YOLOv5n (Same Dataset)	93.2	90.8	89.7	90.2	~140
Commercial Cloud API (Proprietary)	95.1	92.4	91.6	92.0	~480*
SafeHelmet Vision AI — YOLOv8n (Proposed)	97.8	95.2	94.9	95.0	~165

**** Commercial cloud API latency includes network round-trip; local inference only metrics would be comparable. Evaluated on identical 200-image test set.***

SafeHelmet Vision AI achieves mAP50 = 97.8%, representing a 4.6 percentage-point improvement over the YOLOv5n baseline trained on identical data and a 36.4 percentage-point improvement over the traditional Haar cascade baseline. The YOLOv8n architectural advances over YOLOv5n — particularly the anchor-free detection head and C2f backbone — contribute measurably to this improvement without increasing model size or inference latency, confirming the architectural superiority of YOLOv8 for this application domain. Notably, the proposed system exceeds the commercial cloud API in mAP50 by 2.7 percentage points while eliminating network latency, data transmission costs, and subscription fees.

Class-specific performance analysis (Table 2) reveals that helmet detection (mAP50 = 98.1%) marginally outperforms person detection (mAP50 = 97.4%), reflecting the more visually distinctive and consistent appearance of safety helmets compared to the diverse clothing and posture variation of construction workers. Recall for the person class (93.8%) is modestly lower than for helmets (96.0%), primarily attributable to partial occlusion of workers by equipment and scaffolding in dense site configurations — a finding that motivates the priority future enhancement of multi-angle camera fusion.

Table 2: Class-Specific Detection Metrics and Processing Latency Breakdown

Metric	Helmet Class	Person Class	Overall (mean)
Precision (%)	96.4	94.0	95.2
Recall (%)	96.0	93.8	94.9
F1-Score (%)	96.2	93.9	95.0
mAP50 (%)	98.1	97.4	97.8
False Positive Rate (%)	3.6	6.0	4.8
False Negative Rate (%)	4.0	6.2	5.1

Table 3: System Performance Under Varying Environmental and Image Conditions (n = 200 test images)

Condition Category	Test Images (n)	mAP50 (%)	Precision (%)	Recall (%)
Standard Daylight, Unobstructed View	60	98.6	96.8	96.2

Artificial Site Lighting (Night / Indoor)	40	95.3	92.7	91.8
High-Glare / Backlit Conditions	30	94.1	91.3	90.4
Dense Scenes (Partial Occlusion ≥ 30%)	40	93.7	90.2	88.9
Non-Standard Helmet Colours (Black / Silver)	30	96.4	94.1	93.6
Overall (All Conditions)	200	95.8	93.4	92.7

Environmental condition analysis (Table 3) demonstrates that the system maintains mAP50 ≥ 93.7% across all evaluated conditions, confirming robust generalisation beyond the training distribution. The largest performance degradation is observed under dense scene conditions with ≥30% object occlusion (mAP50 = 93.7%), where partial overlapping of workers and equipment reduces the completeness of bounding box predictions for occluded individuals. This finding directly motivates the highest-priority future enhancement of multi-angle camera integration, which would provide complementary viewpoints that reduce effective occlusion rates. Non-standard helmet colours exhibit minimal performance degradation (mAP50 = 96.4%) relative to the overall mean, confirming that the HSV augmentation strategy applied during training successfully prevented over-fitting to the colour distribution of training-set helmets.

Inference latency profiling on a reference Intel Core i5-10th generation CPU with 8 GB RAM (no GPU) yielded a mean end-to-end response time of 165 ms per image, well within the sub-200 ms target. The YOLOv8 inference step accounts for 82% of this total (135 ms), with PIL image loading (12 ms), NumPy conversion (4 ms), and Streamlit rendering (14 ms) contributing the remainder. User acceptance testing with five safety officers and site managers confirmed universal positive ratings for interface usability, result clarity, and practical utility, with participants successfully interpreting all detection outputs without instruction.

5. CONCLUSION

SafeHelmet Vision AI demonstrates that enterprise-grade, AI-powered construction site safety monitoring is achievable using freely available open-source tools, without the prohibitive subscription costs and infrastructure requirements of commercial platforms. The system achieves a validation mAP50 of 97.8%, a precision of 95.2%, and a recall of 94.9% — performance levels sufficient for production deployment in real construction safety monitoring workflows — while delivering sub-200 ms end-to-end response times on standard CPU hardware and requiring zero client-side installation or technical expertise from end users.

The transfer learning methodology, initialising from COCO-pretrained YOLOv8n weights before fine-tuning on 3,500 domain-specific annotated construction site images, proved highly effective in achieving high detection performance with a moderately sized training corpus. The comprehensive data augmentation strategy — encompassing mosaic composition, HSV colour jitter, geometric transformations, and scale variation — provided effective regularisation, as evidenced by mAP50 ≥ 93.7% across all evaluated environmental conditions including artificial lighting, high glare, partial occlusion, and non-standard helmet colours.

The Streamlit deployment architecture, combining @st.cache_resource model caching with a clean four-module application structure, validates Streamlit as a highly productive framework for delivering machine learning model capabilities to non-technical end users. The complete application, excluding training code, is implemented in fewer than 60 lines of Python — a demonstration of the productivity advantages achievable when modern ML deployment frameworks are applied to well-scoped industrial AI problems.Three primary limitations define the development roadmap. First, the current system operates on manually uploaded static images rather than continuous video streams, constraining its effectiveness as a real-time continuous monitoring tool; integration of OpenCV VideoCapture for live IP camera stream processing is the highest-priority enhancement. Second, the detection scope is limited to helmets and workers; expansion to cover safety vests, boots, gloves, and eye protection would transform the system into a comprehensive multi-class PPE compliance monitor. Third, the system cannot identify which specific worker has committed a violation; integration of face recognition or wearable ID detection would enable targeted enforcement.

Future work will address these limitations through: real-time video stream analysis with configurable frame sampling (2–5 fps); automated multi-channel alerting via email, SMS, and messaging platform webhooks; edge deployment on NVIDIA Jetson hardware using TensorRT-optimised model export; multi-camera synchronisation for comprehensive site coverage; and a fleet-scale safety analytics dashboard aggregating compliance data across multiple sites and time periods. SafeHelmet Vision AI establishes a validated, open-source foundation upon which this extended functionality can be systematically built, contributing to the broader objective of making AI-powered occupational safety monitoring accessible across the full diversity of the global construction industry.

6. REFERENCES

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, pp. 779–788, 2016.

[2] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLOv8," Version 8.0.0, Ultralytics Inc., 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[3] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, and C. L. Zitnick, "Microsoft COCO: Common Objects in Context," Proc. European Conference on Computer Vision (ECCV), Zurich, pp. 740–755, 2014.

[4] N. D. Nath, A. H. Behzadan, and S. G. Paal, "Deep Learning for Site Safety: Real-Time Detection of Personal Protective Equipment," Automation in Construction, vol. 112, Article 103085, 2020.

[5] J. Wu, N. Cai, W. Chen, H. Wang, and G. Wang, "Automatic Detection of Hardhats Worn by Construction Personnel: A Deep Learning Approach and Benchmark Dataset," Automation in Construction, vol. 106, Article 102894, 2019.

[6] L. Shen, H. Chen, A. Kirk, and Z. Yu, "Construction Safety Helmet Detection Based on Improved YOLOv5 Algorithm," Journal of Construction Engineering and Management, vol. 147, no. 11, pp. 1–14, 2021.

[7] A. Paszke, S. Gross, F. Massa, and others, "PyTorch: An Imperative Style, High-Performance Deep Learning Library," Advances in Neural Information Processing Systems 32 (NeurIPS), pp. 8024–8035, 2019.

[8] Streamlit Inc., "Streamlit Documentation," Version 1.28, 2024. [Online]. Available: https://docs.streamlit.io

[9] Roboflow Inc., "Roboflow Documentation: Computer Vision Tools and Dataset Management," 2024. [Online]. Available: https://docs.roboflow.com

[10] International Labour Organization, "Safety and Health at Work: Construction Sector," ILO Global Report, Geneva, Switzerland, 2023.

[11] Kaggle Inc., "Kaggle: Your Machine Learning and Data Science Community," 2024. [Online]. Available: https://www.kaggle.com

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, pp. 770–778, 2016.

[13] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," Proc. IEEE CVPR, pp. 4510–4520, 2018.

[14] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," Proc. IEEE CVPR, pp. 580–587, 2014.

[15] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going Deeper with Convolutions," Proc. IEEE CVPR, pp. 1–9, 2015.