International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 5
Sep-Oct 2026, PP 01-5

AI Factory Safety Monitor: A Real-Time Computer-Vision System for PPE Compliance, Danger-Zone Intrusion and Machine-Hazard Alerting on Industrial Shop Floors

Bhor Pranali Dinesh, Prof. Bangar A. P., Prof. Bhosale S.B, Prof. Karad V. A

Dept. of Computer Engineering,

Jaihind collage of Engineering, Pune,India

Abstract. Factory accidents continue to arise from lapses that are simple to describe but hard to police continuously: a worker without a helmet, a hand inside a press envelope, a person in the path of a moving vehicle. Manual supervision and periodic audits cannot watch every station at every moment. This paper presents the AI Factory Safety Monitor (AFSM), an edge-deployed computer-vision system that analyses ordinary industrial camera streams in real time to detect missing personal protective equipment (PPE), intrusions into configurable danger zones, and unsafe proximity to operating machinery. AFSM combines a single-stage object detector, a multi-object tracker, a pose-based posture check and a rule engine that converts raw detections into graded alerts using temporal filtering to suppress false alarms. Alerts are delivered locally (beacon, siren) and remotely (MQTT, SMS, supervisor dashboard), and a compliance-analytics layer records events for audit under occupational health and safety management practice. We describe the architecture, algorithms, deployment considerations and a proposed evaluation protocol, and discuss privacy, reliability and the limits of vision-based safety as a supplement to, not a replacement for, certified safety-rated controls.

Keywords: Industrial safety, computer vision, deep learning, PPE detection, edge AI, YOLO, object tracking, smart manufacturing,

I. Introduction

Manufacturing remains one of the most hazardous work environments. Machine entanglement, struck-by incidents, falls and exposure to hazardous substances account for a large share of workplace injuries, and many of them involve a known rule being broken for a short time: PPE removed "just for a minute", a guard bypassed, a pedestrian crossing a forklift lane. Safety management systems such as ISO 45001 [8] require employers to identify hazards and monitor compliance, but monitoring by supervisors and periodic audits is intermittent by nature.

Most factories already operate closed-circuit cameras, yet footage is usually reviewed only after an incident. Recent advances in deep learning allow the same cameras to be used as active sensors. Single-stage detectors such as YOLO [1], [3] run in real time on small embedded accelerators, and published studies have shown that helmets and high-visibility vests can be detected reliably in construction scenes [6], [7].

This paper proposes the *AI Factory Safety Monitor* (AFSM), a system that turns existing cameras into a continuous safety observer. The contributions are: (i) a modular edge architecture that keeps video on site and sends only events upstream; (ii) a rule engine with temporal filtering and graded severity, designed to reduce the alarm fatigue that causes workers to ignore

warnings; (iii) a configurable danger-zone model tied to live machine state, so that a zone is hazardous only when the machine is actually energised; and (iv) a practical evaluation and deployment protocol.

II. Related Work

Object detection. Two-stage detectors such as Faster R-CNN [2] give high accuracy but at a computational cost that is awkward at the edge. Single-stage detectors [1], [3] trade a small amount of accuracy for much higher frame rates and are the usual choice for on-site video analytics. Detectors are commonly trained and compared on large public datasets such as COCO [10] before being finetuned on domain data.

PPE and site-safety monitoring. Fang et al. [7] detected non-hardhat use from far-field surveillance video, and Nath et al. [6] proposed deep-learning models that verify several PPE items per worker. These works focus mostly on construction sites and on detection alone; they say little about alert policy, machine context or integration with plant systems, which AFSM addresses. *Tracking and pose.* Associating detections across frames with a deep appearance metric [4] and a Kalman filter [12] gives stable identities, which allows a violation to be counted once per person rather than once per frame. Multi-person pose estimation [5] provides body keypoints that can be used to test whether a hand or limb lies inside a restricted region.

III. System Architecture

AFSM follows the layered design shown in Fig. 1. Each layer has a single responsibility and communicates through simple message interfaces, so that components can be replaced or scaled independently.

1. Sensing Layer
Fixed IP cameras deliver H.264 streams over RTSP. Camera placement is part of the design: cameras are mounted at a height and angle that keep workers' heads and torsos visible and minimise occlusion by machinery. Where available, digital signals from the machine PLC (running, stopped, door open, e-stop) are read over a standard industrial protocol and attached to each camera as context.

2. Edge Inference Layer
Inference runs on an on-site edge node with a GPU or neural accelerator, so that raw video never leaves the plant and the system keeps working if the internet link fails. The node hosts four services:

- **PPE detector:** a single-stage detector fine-tuned to recognise person, helmet, high-visibility vest, safety glasses, gloves and safety footwear classes. Each person box is associated with the PPE boxes that fall inside it.
- **Tracker:** a deep-appearance tracker [4] with Kalman motion prediction [12] assigns a persistent identifier to each worker.
- **Zone and pose module:** danger zones are drawn as polygons on the camera image. The foot point of each tracked person, and optionally the wrist keypoints from a pose model [5], are tested against the polygons.
- **Machine-state reader:** converts PLC signals, or a vision cue such as a lit status tower, into a binary "hazardous" flag per zone.

3. Rule Engine and Alerting
The rule engine receives per-frame events and applies the rules in Table I. A rule fires only after its condition has held for a configurable number of consecutive frames. Notifications are published over MQTT [11] and mapped to local beacons, sirens and supervisor devices.

International Journal of Statistics and Mathematical Research Trends
E-ISSN: 3108-0871, Volume 2 Issue 5
Sep-Oct 2026, PP 01-5

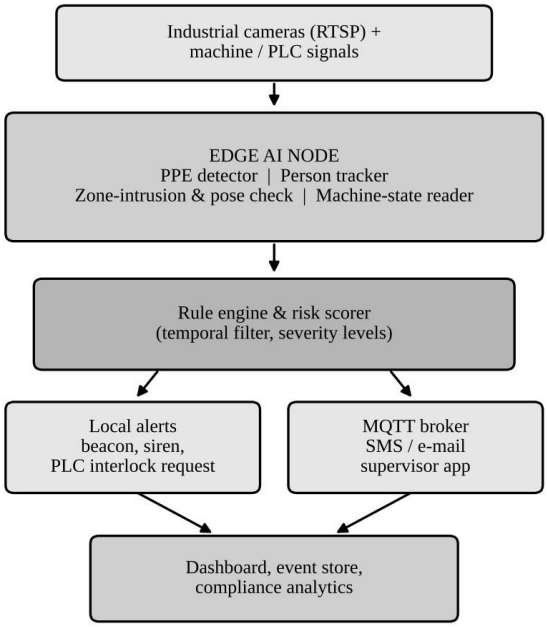


Fig.1. Layered architecture of the AFSM

4. Compliance Analytics

Every confirmed event is stored with a timestamp, camera, zone, rule, severity and a short annotated snapshot. Aggregated views (violations per shift, per zone, per rule, time-to-acknowledge) support toolbox talks, rootcause analysis and audit evidence.

IV. Methodology

1. Detection and Association

A person is judged compliant for item *k* if a detection of that item with confidence above a threshold lies in the expected body region of the person box (for example, a helmet in the upper third). Associating items to people by geometry avoids the common error of counting a helmet carried in the hand or lying on a bench.

2. Danger-Zone Test

For a tracked person with foot point *f* and a zone polygon *Z*, the person is inside the zone if *f* lies in *Z*. If pose is available, a wrist keypoint inside *Z* also counts, which captures a hand reaching into a press while the body stays outside. A zone is evaluated only while the machinestate flag is hazardous.

3. Temporal Filtering

Single-frame detections are noisy. A rule is confirmed when it is true in at least *m* of the last *n* frames for the same track identifier, and it is cleared only after a longer run of compliant frames. This hysteresis removes flicker caused by motion blur and brief occlusion, at the cost of a latency of roughly *n* divided by the frame rate.

4. Risk Score

To rank simultaneous events, each confirmed violation *i* is given a weight *w* reflecting its inherent severity and a duration factor *d*. The risk score of a zone is

- $R = \min(1, \Sigma\, w_i \cdot d_i)$ (1) and the alert level follows the thresholds in Table II. The compliance rate for a period is
- $C = 1 - (V / N)$ (2) where *V* is the number of person-observations with at least one violation and *N* is the total number of personobservations.

International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 5
Sep-Oct 2026, PP 01-5

Table 1: Rules Implemented in AFSM

Rule	Condition	Default level
PPE-1	Person without helmet	Warning
PPE-2	Person without vest in vehicle area	Warning
ZONE-1	Person in zone, machine running	Critical
ZONE-2	Wrist in press zone, cycle active	Critical
VEH-1	Person within lane of moving vehicle	Critical
EGR-1	Blocked exit or aisle	Advisory

Table 2: Alert Levels and Responses

Level	Risk R	Response
Advisory	< 0.3	Log, dashboard note
Warning	0.3 - 0.7	Beacon, voice message, notify supervisor
Critical	> 0.7	Siren, SMS, request machine stop via PLC

V. Implementation Considerations

1. Hardware and Software

A representative deployment uses an edge node with a mid-range GPU or an embedded module with a neural accelerator, serving several camera streams. Software is containerised, with separate services for ingest, inference, rule evaluation, MQTT messaging and a web dashboard, which allows restart of a failed component without stopping the others. Models are exported to an optimised inference runtime and run at reduced numerical precision to raise throughput.

2. Data Preparation

Public datasets rarely match a particular plant's lighting, uniforms and camera angles. The recommended approach is to start from a model pre-trained on a general dataset [10], collect a few thousand frames from the target cameras, annotate the PPE and person classes, and finetune. Hard cases such as backlit doorways, welding glare and workers carrying tools should be deliberately included. Periodic retraining on frames where supervisors marked an alert as false closes the loop.

3. Privacy and Workforce Trust

Continuous video of employees raises legitimate concern. AFSM limits the risk by processing on site, storing only event snapshots rather than continuous video, optionally blurring faces in stored images, restricting access by role, and publishing the monitoring policy to workers. Statistics are reported per zone and shift rather than per named individual. Involving workers and safety representatives before roll-out improves acceptance and the quality of zone definitions.

VI. Proposed Evaluation

Quantitative validation of AFSM should be c

arried out on labelled video from the target site, separated from the training frames by day and camera. Table III lists the proposed metrics. Detection quality is reported as precision, recall and mean average precision per class [10]; operational quality is reported as false alarms per camera-hour, which is more meaningful to operators than frame-level accuracy, and as end-to-end latency from event to alert.

Table 3: Proposed Evaluation Metrics

Metric	Purpose
Precision / recall per class	Detector quality
mAP at IoU 0.5	Overall detection
False alarms per camera-hour	Alarm fatigue

Event-to-alert latency	Timeliness
Frames per second per stream	Edge capacity
Missed-event rate in staged drills	Safety effectiveness

Staged drills, in which trained staff deliberately enter a zone or remove PPE under supervision, provide ground truth for missed events without exposing anyone to risk. Results should be reported separately for day, night, and glare conditions.

Limitations And Safety Integrity

AFSM is a monitoring and warning aid. It is not a safety-rated function in the sense of ISO 13849-1 [9], and a camera-based alert must not be the only means of preventing injury from a machine. Where AFSM requests a machine stop, the request should pass through an existing certified safety controller rather than act directly on the drive. Known limitations include occlusion, extreme lighting, unusual uniforms, and degraded accuracy when cameras are moved without recalibrating zones. Missed detections and false alarms are both unavoidable, so the system should be introduced alongside, and not instead of, guarding, training and supervision.

VII. Conclusion and Future Work

This paper presented the AI Factory Safety Monitor, an edge-based architecture that converts existing factory cameras into a continuous detector of PPE non-compliance and danger-zone intrusion, with temporal filtering, graded alerts and audit-ready analytics. The design emphasises low false-alarm rates, machine-aware zones, on-site processing and respect for worker privacy. Future work includes site trials following the proposed protocol, detection of additional hazards such as spills, fire and smoke, multi-camera fusion to handle occlusion, and anomaly detection that learns each plant's normal behaviour.

Acknowledgment

The authors thank the faculty and industry mentors who provided guidance on shop-floor safety practice.

References

1. J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779-788.
2. S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems (NIPS)*, 2015, pp. 91-99.
3. G. Jocher, A. Chaurasia and J. Qiu, "Ultralytics YOLOv8," 2023. [Online]. Available:
4. https://github.com/ultralytics/ultralytics
5. N. Wojke, A. Bewley and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *Proc.*
6. *IEEE Int. Conf. Image Processing (ICIP)*, 2017, pp. 36453649.
7. Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei and Y. Sheikh, "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 172-186, 2021.
8. 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, art. 103085, 2020.
9. Q. Fang, H. Li, X. Luo, L. Ding, H. Luo, T. M. Rose and W. An, "Detecting non-hardhat-use by a deep learning method from far-field surveillance videos," *Automation in Construction*, vol. 85, pp. 1-9, 2018.