iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** X **Month of publication:** October 2026

DOI: https://doi.org/10.22214/ijraset.2026.85522

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue X Oct 2026- Available at www.ijraset.com

An Integrated YOLOv9-e Framework for PPE Compliance and Sanitation Hazard Detection Using Worker Tracking and Temporal Evidence Filtering

Mohammed Saif Z.[1], Vinothini T.[2]
Department of Computer Science and Engineering, Adhiyamaan College of Engineering, Hosur, Tamil Nadu, India

Abstract: ***This paper presents an integrated computer-vision system for monitoring personal protective equipment (PPE) compliance and selected sanitation hazards in video streams. The system combines a fine-tuned YOLOv9-e detector with worker tracking, geometric PPE-to-person association, temporal evidence filtering, and persistent alert management. Two detection configurations are evaluated. The six-class configuration is derived from SH17 and contains person, helmet, vest, gloves, glasses, and shoes. The extended twelve-class configuration adds hairnet, apron, mask, rat, cockroach, and lizard using the Roboflow Universe Safety Food dataset. On the six-class validation set, YOLOv9-e records 75.94% precision, 64.25% recall, 69.34% mAP@0.5, and 47.04% mAP@0.5:0.95. A separate post-training evaluation reports 69.20% and 47.00% for the two mAP measures, respectively. The twelve-class validation run reports 84.89% precision, 76.45% recall, 81.15% mAP@0.5, and 54.36% mAP@0.5:0.95. These configurations use different class sets and data distributions and are therefore not treated as a controlled model comparison. The system further uses ByteTrack, anatomical spatial heuristics, a time-calibrated exponential moving average with hysteresis, and a three-second alert persistence rule. A detector-stage timing measurement of 29.67 ms, equivalent to approximately 33.7 frames per second, is reported; it excludes the complete video-to-dashboard pipeline. Evaluation was conducted on a local workstation using validation data and pre-recorded video, with no edge-device or multi-camera field validation.***

Keywords: Personal Protective Equipment, YOLOv9-e, Object Detection, SH17, ByteTrack, Spatial Association, Exponential Moving Average, Sanitation Hazard Detection.

I. INTRODUCTION

Workplace safety monitoring requires more than detecting protective equipment in isolated images. A useful monitoring system must determine whether equipment is associated with a particular worker, maintain that worker's identity across video frames, and avoid raising alerts for brief detection failures. Conventional manual inspections are periodic, while basic camera-based detection can produce unstable decisions under occlusion, motion blur, changing viewpoints, and small object sizes.

This work develops an integrated PPE monitoring platform around YOLOv9-e object detection. It extends frame-level detection with ByteTrack-based worker trajectories, geometric association between PPE and person detections, and temporal evidence filtering. The implementation also includes a broader twelve-class configuration that combines industrial PPE categories with food-workspace hygiene items and selected pest classes.

The contributions are implementation-oriented: (i) an integrated detection and monitoring pipeline; (ii) geometric PPE association using normalized worker-relative coordinates; (iii) worker-box recovery from headwear or vest detections when a person box is missing; (iv) a second inference pass on cranial and torso crops; and (v) time-calibrated EMA evidence filtering with hysteresis and alert persistence. The paper does not claim a new detector backbone, a new tracking algorithm, or a statistically established reduction in false alarms.

II. RELATED WORK

Deep learning has become a common approach for visual PPE monitoring. Nath et al. demonstrated deep-learning-based detection of PPE in construction-site imagery and highlighted the relevance of worker-level safety monitoring [1]. The SH17 dataset provides a broader benchmark for human safety and PPE detection in manufacturing scenes, with 8,099 images, 75,994 annotated instances, and 17 original classes [2].

©IJRASET: All Rights are Reserved

YOLOv9 introduces programmable gradient information and a generalized efficient layer aggregation network to improve object detection [3]. In this project, YOLOv9-e is used as the detection backbone. ByteTrack associates high- and low-confidence detections across frames to maintain object trajectories [4]. The present system uses the Ultralytics ByteTrack integration and supplements it with project-specific worker identity matching and compliance-state handling.

Unlike a detector-only evaluation, the proposed platform connects detection, worker tracking, PPE association, temporal state estimation, alert generation, and event persistence. The spatial association and temporal filtering stages are deterministic heuristics rather than learned pose-estimation or action-recognition models.

III. DATASET PREPARATION AND CLASS MAPPING

A. *SH17 Benchmark*

The SH17 dataset contains 8,099 annotated images and 75,994 object instances across 17 classes. Its supplied split contains 6,479 training images and 1,620 validation images. The dataset is distributed under the CC BY-NC-SA 4.0 license, which imposes non-commercial and share-alike conditions [2]. The six-class project taxonomy retains person, helmet, vest, gloves, glasses, and shoes. Eleven original classes are removed and the remaining labels are re-indexed to contiguous class identifiers. Images without any retained target annotation are also excluded, resulting in 6,207 training images and 1,544 validation images, for a total of 7,751 images.

B. *Extended Safety Food Dataset*

The extended configuration combines the remapped SH17 data with Safety Food, Roboflow Universe dataset version 3, exported on 4 September 2024 [5]. The source dataset contains labels for apron, cockroach, gloves, hairnet, lizard, mask variants, rat, and related positive/negative hygiene states. For the unified twelve-class taxonomy, the project retains six SH17 categories and maps hairnet, apron, positive mask detection, rat, cockroach, and lizard to six additional classes. Negative-state labels such as no_apron, no_gloves, no_hairnet, without_mask, and mask_weared_incorrect are not retained as separate detector classes.

The resulting extended dataset contains 11,251 training images and 2,680 validation images, totaling 13,931 images. The class mapping is summarized in Table I. The extended set contains data from two source datasets and therefore has a different class distribution and visual domain from the six-class SH17-only configuration.

TABLE I. DATASET CONFIGURATIONS AND CLASS TAXONOMY

Dataset / split	Images / classes
SH17 original	8,099 images; 75,994 instances; 17 classes
SH17 remapped	6,207 train + 1,544 validation = 7,751 images
Extended 12-class	11,251 train + 2,680 validation = 13,931 images
Six retained SH17 classes	person, helmet, vest, gloves, glasses, shoes
Six added classes	hairnet, apron, mask, rat, cockroach, lizard

The SH17 source is available through its published distribution and the Safety Food dataset is attributed to its Roboflow Universe project page. The SH17 license is non-commercial; any future commercial deployment must resolve dataset licensing and image-rights requirements before use.

IV. PROPOSED SYSTEM ARCHITECTURE

The system accepts RTSP streams, MP4 files, or webcam input. OpenCV performs video decoding and threaded frame ingestion. The implementation samples frames at a configured interval of three frames before forwarding them to the detection engine. YOLOv9-e performs primary inference, followed by class-specific confidence filtering. The pipeline then applies worker recovery and crop-based refinement, passes detections to ByteTrack, associates PPE objects with worker boxes, updates temporal compliance evidence, and generates alerts when a violation persists.

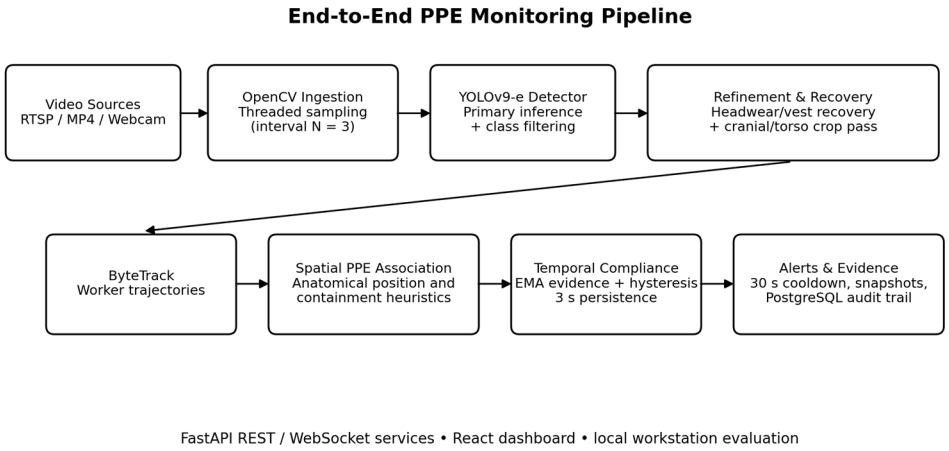


Fig. 1. Implemented processing stages and connected monitoring services.

A FastAPI backend exposes REST endpoints for inference, camera management, stream control, alert retrieval, and system health. WebSocket channels deliver video frames and alert events to the React dashboard. PostgreSQL stores camera metadata, detections, person records, alerts, and alert history. The dashboard provides Live Monitor, Testing Ground, Analytics, and Health views. The Analytics page includes a chart that may use placeholder values when backend history is sparse; it is therefore not treated as a fully validated analytics result in this paper.

V. YOLOV9-E TRAINING CONFIGURATION

A. Model and Training Setup

The detector is YOLOv9-e with approximately 58.1 million parameters, initialized from the pretrained yolo9e.pt weights. Training was performed using the Ultralytics framework with PyTorch 2.x and CUDA 12.1. The configured image size was 640 × 640 pixels, batch size was 4, and the optimizer was SGD/automatic selection. The initial learning rate was 0.001, final learning-rate factor 0.01, momentum 0.937, weight decay 0.0005, and warm-up duration three epochs. Automatic mixed precision was enabled.

Data augmentation included mosaic (disabled during the final ten epochs), mixup at 0.1, HSV augmentation, horizontal flipping at 0.5, and random erasing at 0.4. The six-class run was logged for 25 epochs. The extended twelve-class run was logged for 20 epochs and its recorded initialization references a previous last.pt checkpoint. Accordingly, the reported twelve-class metrics are attributed to that logged run; the available audit does not establish a clean, independent-from-scratch training comparison.

TABLE II. YOLOV9-E TRAINING CONFIGURATION

Parameter	Setting
Model / parameters	YOLOv9-e / approximately 58.1 million
Input size / batch	640 × 640 / 4
Optimizer / learning rate	SGD (automatic) / 0.001
Momentum / weight decay	0.937 / 0.0005
Warm-up / AMP	3 epochs / enabled
Mosaic / MixUp	1.0 (closed final 10 epochs) / 0.1
HSV H / S / V	0.015 / 0.7 / 0.4
Horizontal flip / erasing	0.5 / 0.4
Logged epochs, 6-class / 12-class	25 / 20

B. *Detection Thresholds and Refinement*

The primary detector pass uses a confidence threshold of 0.035 and an IoU threshold of 0.55. Class-specific post-filters are applied after inference. The implementation uses higher thresholds for selected categories, including sanitation hazards, and applies person-box size and confidence checks. These filters are operational rules in the deployed code and should not be interpreted as separately optimized thresholds.

When a person detection is missing, headwear or vest detections can trigger a worker-box recovery heuristic. The recovered box is synthesized from the PPE detection geometry. In addition, up to eight worker regions can receive a second inference pass. The cranial crop covers the upper 38% of the worker region with a 5% margin, while the torso crop covers approximately 10% to 88% of worker height with an 8% horizontal margin. These crops are resized for inference at 384 pixels. This process is intended to recover small or partially occluded PPE detections; no ablation experiment is available to quantify its independent contribution.

VI. WORKER TRACKING AND SPATIAL PPE ASSOCIATION

A. *Worker Tracking*

Detections are passed to the Ultralytics ByteTrack implementation using persistent tracking across frames. The standard tracker configuration uses high- and low-confidence thresholds of 0.50 and 0.10, a matching threshold of 0.80, and a track buffer of 30 frames. The project also applies a custom identity allocator and spatial matching logic using intersection-over-union and centroid distance. The custom matching layer is distinct from native ByteTrack association and is used to maintain compliance records when detections are temporarily inconsistent.

B. *Anatomical Spatial Association*

For a worker bounding box with top coordinate y_1 and height h_w, the vertical position of a PPE detection center c_y is normalized as:

$$\mathbf{y_rel = (c_y - y_1) / h_w}$$

The normalized vertical coordinate is checked against class-specific anatomical ranges, while horizontal containment tolerances account for imperfect detector boxes. Helmet and hairnet detections are associated with upper-body regions, vests and aprons with torso regions, and shoes with the lower region. When direct crop-to-worker identity is unavailable, an expanded worker box and containment ratio are used to select the most plausible person association. These are geometric heuristics and do not constitute learned human-pose estimation.

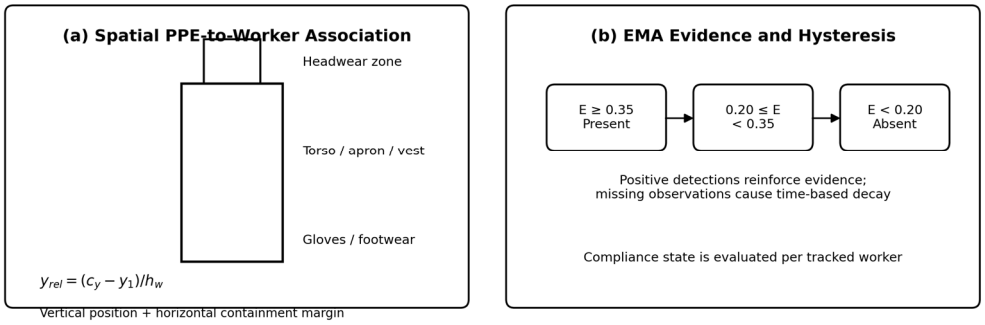


Fig. 2. Conceptual representation of worker-relative association and temporal evidence thresholds.

VII. TEMPORAL COMPLIANCE EVALUATION AND ALERT MANAGEMENT

A. *EMA Evidence Update*

The implementation maintains a smoothed evidence score E for each PPE class associated with a tracked worker. The update is time-calibrated rather than using one fixed smoothing coefficient. When a detection is present with raw confidence $C \geq 0.04$, the evidence is initialized as $E = \max(C, 0.65)$. For subsequent positive observations, the adaptive coefficient is $\alpha = \min(0.85, \max(0.40, 10\Delta t))$, and the evidence is updated as:

$$\mathbf{E_t = (1 - \alpha)\, E_(t-1) + \alpha \max(C, 0.60)}$$

When the item is not detected, the evidence decays exponentially. The decay time constant is $\tau = \max(1.0, T_persistence / 1.1786)$. With a persistence setting of 3.0 seconds, τ is approximately 2.545 seconds. The absence update is:

$$E_t = E_{(t-1)} \exp(-\Delta t / \tau)$$

B. Hysteresis and Compliance State

The filter uses two thresholds. Evidence at or above 0.35 confirms the PPE item as present; evidence below 0.20 marks it absent. Values in the interval $0.20 \leq E < 0.35$ retain the previous state, providing hysteresis. A worker is classified as COMPLIANT when all required items are present, PARTIAL when at least one required item is present but one or more are missing, and NON_COMPLIANT when all required items are absent. These states are evaluated against the active camera or zone policy.

C. Alert Persistence and Evidence

A non-compliant state must persist for 3.0 seconds before an alert is emitted. A 30-second cooldown is enforced per camera and worker track to reduce repeated notifications. The alert manager captures a JPEG evidence snapshot and records the event in PostgreSQL. Alert delivery and webhook dispatch are handled asynchronously, while WebSocket messages update connected monitoring clients. The filter and alert mechanisms are covered by unit and regression tests; however, no large-scale frame-by-frame ground-truth experiment was conducted to quantify false-alarm reduction.

VIII. EXPERIMENTAL RESULTS AND DISCUSSION

A. Detection Metrics

The reported detection metrics are validation results from the respective training runs. The six-class run reached 75.94% precision, 64.25% recall, 69.34% mAP@0.5, and 47.04% mAP@0.5:0.95 at its final logged epoch. The highest recorded mAP@0.5 during training was 69.39%. A separate post-training evaluation reported 69.20% mAP@0.5 and 47.00% mAP@0.5:0.95. These values are close but are retained as separate measurements because they come from different evaluation records.

The twelve-class run reported 84.89% precision, 76.45% recall, 81.15% mAP@0.5, and 54.36% mAP@0.5:0.95. The twelve-class score must not be interpreted as proof that the extended model is more accurate than the six-class model: the class taxonomy, dataset composition, and training setup differ.

TABLE III. YOLOV9-E VALIDATION PERFORMANCE

Configuration	**Validation metrics (P / R / mAP@0.5 / mAP@0.5:0.95)**
6-class, final logged epoch	75.94% / 64.25% / 69.34% / 47.04%
6-class, separate post-training eval	Not recorded / Not recorded / 69.20% / 47.00%
12-class, logged run	84.89% / 76.45% / 81.15% / 54.36%

For the six-class model, the per-class AP@0.5 values recorded in the post-training evaluation are 91.25% for person, 74.30% for helmet, 75.73% for glasses, 64.59% for shoes, 58.62% for gloves, and 50.71% for vest. The lower AP for gloves and vest indicates that small-object scale, occlusion, and visual variation remain challenging.

B. Detector Timing

The available timing report records 1.15 ms preprocessing, 26.70 ms inference, and 1.82 ms postprocessing, totaling 29.67 ms per image. This corresponds to approximately 33.7 frames per second for the detector stage. The measurement does not include full RTSP decoding, tracking, spatial association, database writes, WebSocket transmission, or dashboard rendering. It must therefore not be described as end-to-end system throughput.

TABLE IV. DETECTOR-STAGE LATENCY

Stage	**Latency / throughput**
Preprocessing	1.15 ms
YOLOv9-e inference	26.70 ms
Postprocessing	1.82 ms
Total detector stage	29.67 ms
Equivalent detector-stage rate	Approximately 33.7 FPS

C. Implementation Verification

The technical audit records successful unit and regression tests for class mapping, compliance evaluation, PPE association, alert management, and webhook integration. Video runs were conducted locally using pre-recorded MP4 input and a local GPU workstation. The project includes FastAPI endpoints and WebSocket support for edge-agent communication, but the audit does not establish deployment or validation on an edge device. It also does not establish simultaneous multi-camera performance.

D. Interpretation

The results demonstrate that the implemented detector can identify the selected PPE and sanitation classes on the prepared validation sets. The worker association and temporal stages provide a structured route from object detections to worker-level compliance events. Nevertheless, detector mAP is not equivalent to worker-level compliance accuracy. A dedicated annotated video benchmark is required to measure track identity stability, PPE association precision/recall, alert delay, false alarms per hour, and complete end-to-end latency. No 87.7% mAP@0.5 or 91.4% compliance recall value is reported because the available audit did not identify reproducible logs or a defined ground-truth evaluation protocol supporting those figures.

IX. LIMITATIONS AND FUTURE WORK

Several limitations affect the scope of the findings. First, the evaluation is limited to a single local workstation and does not establish edge-device performance, deployment scalability, or multi-camera throughput. Second, the twelve-class model combines datasets from different visual domains, and the current evaluation does not include a separate external test set. Third, the spatial PPE association relies on bounding-box geometry rather than explicit pose estimation. Fourth, small objects such as gloves and glasses can be difficult to detect at long camera distances. Fifth, the system detects general eyewear and footwear but does not verify industrial certification standards or protective ratings.

The SH17 dataset license is CC BY-NC-SA 4.0, so commercial use requires careful licensing review and potentially replacement or relicensing of the training data. Future work should include a held-out test set, class-balanced evaluation, ablation studies for worker recovery and crop refinement, annotated multi-worker video evaluation, complete end-to-end latency measurement, and testing on edge hardware. Additional security work is also required before deployment, including API authentication and authorization.

X. CONCLUSION

This paper described an integrated YOLOv9-e-based platform for PPE compliance and selected sanitation-hazard monitoring. The system combines object detection with worker tracking, anatomical spatial association, worker-box recovery, crop-based refinement, temporal EMA evidence filtering, hysteresis, and persistent alert management. The six-class SH17 configuration achieved 69.34% mAP@0.5 in its final logged validation epoch, while the extended twelve-class configuration achieved 81.15% mAP@0.5 on its own validation set. These figures characterize different configurations and are not a controlled comparison. The measured 29.67 ms latency applies to the detector stage only. The implementation provides a foundation for worker-level safety monitoring, but further video-level, multi-camera, edge-device, and licensing validation is necessary before claims of field readiness or commercial deployment can be made.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

XI. ACKNOWLEDGMENT

The author gratefully acknowledges Mrs. T. Vinothini, M.E., Assistant Professor, Department of Computer Science and Engineering, Adhiyamaan College of Engineering, for academic guidance and supervision.

REFERENCES

[1] 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. no. 103085, 2020, doi: 10.1016/j.autcon.2020.103085.

[2] H. M. Ahmad and A. Rahimi, “SH17: A dataset for human safety and personal protective equipment detection in manufacturing industry,” Journal of Safety Science and Resilience, vol. 6, no. 2, pp. 175–185, Jun. 2025, doi: 10.1016/j.jnlssr.2024.09.002.

[3] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, “YOLOv9: Learning what you want to learn using programmable gradient information,” in Computer Vision – ECCV 2024, Lecture Notes in Computer Science, vol. 15089, 2024.

[4] Y. Zhang et al., “ByteTrack: Multi-object tracking by associating every detection box,” in Computer Vision – ECCV 2022, Lecture Notes in Computer Science, vol. 13682, 2022, doi: 10.1007/978-3-031-20047-2_1.

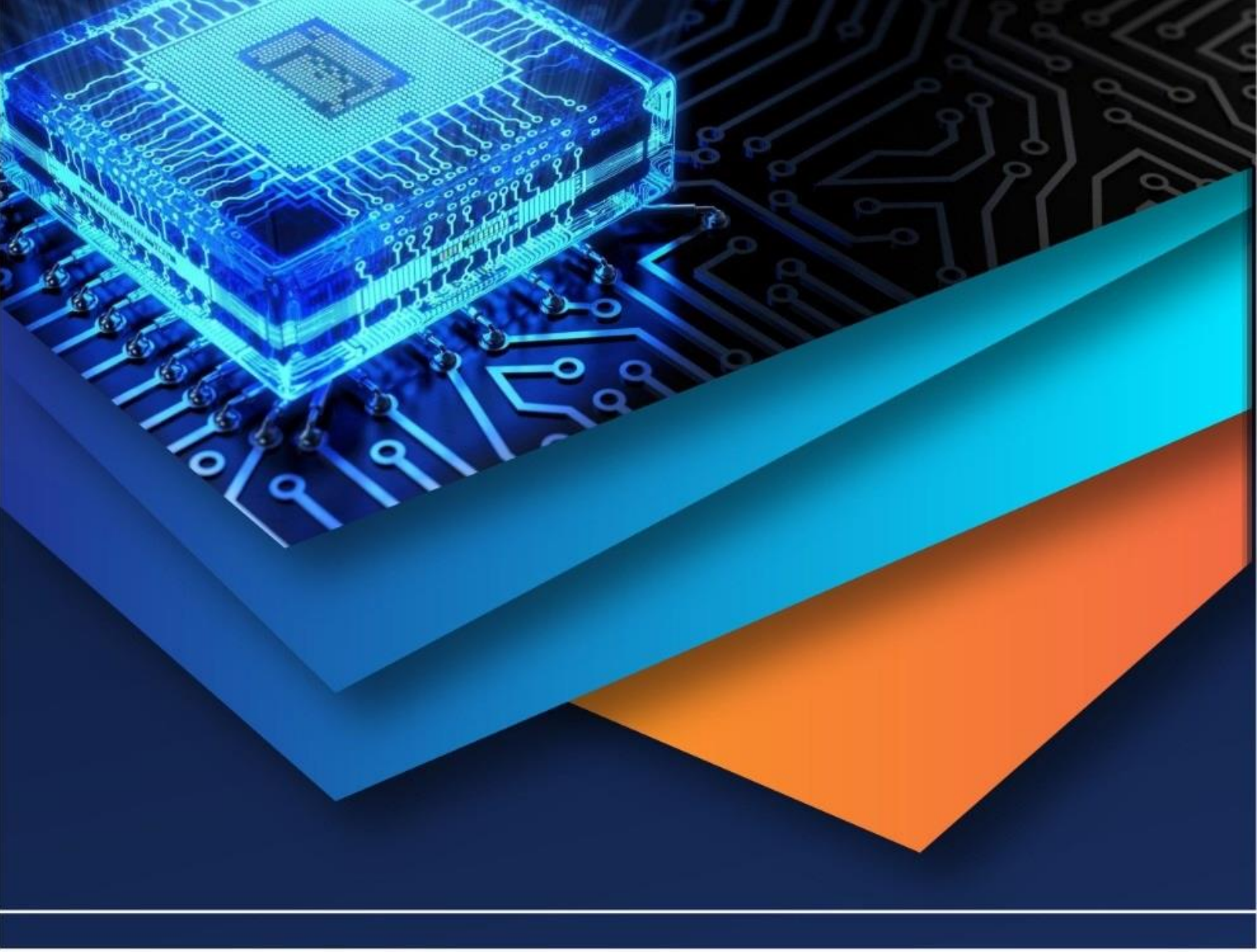
[5] [5] Safety Food System, “Safety Food,” Roboflow Universe, Dataset Version 3, Sep. 4, 2024. [Online]. Available: https://universe.roboflow.com/safety-food-system/safety-food/dataset/3. License: CC BY 4.0.

[6] [6] Occupational Safety and Health Administration, “Personal protective equipment,” 29 CFR § 1910.132, U.S. Department of Labor.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue X Oct 2026- Available at www.ijraset.com

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)