*Journal of Information & Optimization Sciences*
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 47 (2026), No. 5-A, pp. 1623–1631
DOI : 10.47974/JIOS-2252

# Optimized multi-modal deep learning for integrated safety compliance monitoring : PPE detection, face recognition, and trade classification in industrial environments

Megha Nain †
Shilpa Sharma *
*Department of Computer Applications*
*Manipal University Jaipur*
*Jaipur 303007*
*Rajasthan*
*India*

Sandeep Chaurasia §
*Department of Computer Science & Engineering*
*Manipal University Jaipur*
*Jaipur 303007*
*Rajasthan*
*India*

**Abstract**

Compliance monitoring of industrial safety remains a major challenge, due to the inherent flaws in single-modality schemes of implementing the highly multidimensional characteristics of workplace hazards. This paper aims to implement a multi-modal deep-learning system that can detect and identify both personal protective equipment (PPE) and facial features and classify and identify the trade to automatically check compliance with safety standards in the industry. The framework uses Detectron2 to get 82.33% on instance segmentation and bounding-box regression, 95.35% on facial recognition with pre-trained embeddings, 92% on a trade classifier based on a convolutional network, and 99.5% on tool detection with YOLO. These modules are merged to form the combined pipeline, which verifies the identity of the workers, verifies certified trade and ensures the use of PPE in real time and then overlaid the safety-status annotations upon visual results. The quantitative results show significant advantage over the existing techniques, such as the capacity to

This research was conducted without any grants from any institution or organization.

† *E-mail:* `megha.201025603@muj.manipal.edu`
* *E-mail:* `shilpa.sharma@jaipur.manipal.edu (Corresponding Author)`
§ *E-mail:* `sandeep.chaurasia@jaipur.manipal.edu`

surpass the Average Precision of YOLOv3 by 72.3% in PPE detection. The work is innovative as it includes the multi-modal data in a holistic manner and effectively deals with the interdependencies of the determinants of safety compliance better than the single module models. This study, therefore, offers a scalable, optimized, data-driven model that provided the incidence of workplace accidents through automated surveillance, and where the operability of the system is ensured, by qualitative analyses in which the system is stable in identifying non-compliance situations such as a missing mask or vest, and in all modalities, the model is highly accurate.

---

***Subject Classification:*** *68T07, 68T45, 68T10, 68U10.*

***Keywords:*** *Multi-modal deep learning, Industrial safety compliance monitoring, PPE detection, Face recognition, Trade classification, Automated workplace surveillance.*

## 1. Introduction

The workplaces in industries still remain a high-risk workspace despite the gradual upgrades in safety measures, and the occupational risks cause some 2.3 million working deaths every year globally. Traditional safety monitoring is mainly based on manual inspection, which is prone to human error, inconsistency and upscaling. The more recent compliance checks automation efforts have focused on single solutions, e.g. PPE detection [1], or tool monitoring [2], but such solutions poorly reflect the systemic nature of safety compliance determinants. The identity of a worker, certified trade, and the use of PPE are the factors that determine the status of safety, which makes the use of an integrated approach a necessity. The current investigation addresses this gap and comes up with a multi-modal framework that integrates PPE detection, facial recognition, and trade classification into a unified framework. Individual tasks have been shown to be effectively performed by using computer vision: Detectron2 to segment objects [3], convolutional neural networks to classify trades [4], and pre-trained embeddings to perform facial recognition, yet previous researchers have not addressed by combining these elements with each other to measure compliance in holistic terms. As an example, [5] presented monitoring based on IoT, but did not include visual analytics in real-time, and [6] focused closely on risk assessment but did not involve identity verification. The hypothesis put forward in the current paper is that a multi-modal deep-learning pipeline will be more accurate and practical than single-task systems since it addresses the ambiguity of focused prediction with a cross-modal validation. As an illustration, when trade and PPE data are analyzed together, a worker who is an electrician and does not have insulated gloves can be flagged more easily. The main goals, therefore, will be the tripartite objectives of (I) high-precision detection of PPE items (such as helmets, vests, and masks) with the help of instance segmentation; (II) verification of worker identity and certified trade with facial recognition and CNN-based classification; (III) incorporation of these modules into a real-time system, which will provide actionable safety status. This study has four substantial contributions to the occupational safety management. First, it presents a new architecture which aligns PPE detection, face recognition and trade classification, thus filling a significant gap in the literature [7]. Second, the modular character of the designed system allows it to be upgraded in bits, e.g. the introduction of new PPE categories or trade classes without requiring all the pipeline to be retrained. Third, quantitative findings show better performance than extant ones, such as a 10% average precision improvement of PPE detection with YOLOv3-based solutions [8]. Lastly, the framework has been tested empirically within the industrial setting, thus demonstrating its applicability to practice through case studies where non-compliance was identified and corrected in-the-field. The rest of this paper is structured in the following way: Section 2 provides a literature review on safety monitoring systems and multi-modal deep learning, Section 3 outlines the methodology, including data preparation, model architectures and integration logic, Section 4 reports quantitative and qualitative findings, and Section 5 summarizes the consequences, limitations and future research directions.

## 2. Methodology

### 2.1 *Annotation and Preparation of Data*

The PPE detection module was trained with the help of a corpus of 4398 annotated images out of which 250 were set aside to serve as validation [9]. Bounding boxes and segmentation masks were annotated on each image in five categories, namely; helmet, vest, mask, safe, and not safe. It followed the COCO format of annotation in order to be compatible with the Detectron2 framework. In face recognition a separate collection was compiled, with five pictures of each subject of eight subjects bringing 40 samples. The pre-computed face embedding was done based on the face recognition library, which was built on a ResNet based architecture to obtain the 128-dimensional embedding. The problem of distinguishing vocations that are visually similar required a larger dataset in order to classify them as in trade classification. Original images were duplicated to at least 300 samples per category (carpenter, concrete worker, glazier, heavy machine operator, metal worker) through transformations of flipping horizontally, rotation by (±15° rotation), Gaussian blur (sigma= 0.5) and brightness variation ($\Delta = 30\%$). The pipeline of augmentation ensured that classes were equally distributed and reduced overfitting, a finding that was supported by the same validation accuracy across cross-validation folds.

### 2.2 *Model Architectures*

#### 2.2.1 PPE Detection

The Detectron2 model was set up using a ResNet -50-FPN backbone trained on COCO. The network also predicts bounding boxes and instance masks, and their combined loss is formulated as below:

$$\mathcal{L}_{PPE} = \lambda_{box}\mathcal{L}_{box} + \lambda_{mask}\mathcal{L}_{mask} + \lambda_{cls}\mathcal{L}_{cls},$$

where $\mathcal{L}_{box}$ is the GIoU loss incurred by regression of bounding-boxes, $\mathcal{L}_{mask}$ is the loss incurred by segmentation, and the $\mathcal{L}_{cls}$ denotes the loss incurred by class prediction by softmax. The weighting coefficients $\lambda$ were empirically defined which are $\lambda_{box}$=1.0, $\lambda_{mask}$=0.5 and $\lambda_{cls}$=1.0.

2.2.2 Face Recognition

A pre trained FaceNet model provided 128-dimension embeddings of every detected face. Workers The matches were based on similarity of an embedding to the ones in a worker database, using similarity as a measure of a match and a similarity threshold of 0.6 to trade-off false-positives against false-negatives on the validation set.

2.2.3 Trade Classification

A custom CNN with three convolutional blocks (with 32, 64, and 128 filters, respectively) followed by two fully connected layers (256 and 5 units) took worker images. The blocks in the convolutional layer were equipped with Batch Normalization and ReLU activation, and dropout ($p$ = 0.3) was used before the last softmax. Adam optimizer (learning rate= $\eta = 10^{-4}$) and categorical cross-entropy loss were used to optimize the network.

2.3 *System Integration*

The modules were combined into a real-time pipeline using OpenCV to capture images and using multiprocessing library of Python to run inference in parallel. As shown in Figure 1, the workflow starts with face detection and then moves on to simultaneous PPE and trade classification. Table 1 shows Model training configurations

2.4 Training Protocols

| **Algorithm 1: Safety Compliance Verification System** |
|---|
| Input: Image/Video frame *I* |
| Output: Overall safety compliance status and annotated frame |
| 1. Initialize face recognition, trade recognition, and PPE detection models.<br>2. Acquire input frame *I*.<br>3. Process in parallel:<br>• Perform face recognition to identify worker and retrieve certified trade $T_{cert}$.<br>• Perform trade recognition to predict trade $T_{pred}$.<br>• Perform PPE detection to identify required safety equipment.<br>4. Evaluate PPE Compliance:<br>If all mandatory PPE items are detected, set PPE status = SAFE; else UNSAFE.<br>5. Verify Trade:<br>If $T_{cert} = T_{pred}$, set trade status = SAFE; else UNSAFE.<br>6. Fuse Decisions:<br>If PPE status = SAFE and trade status = SAFE, set overall compliance = SAFE; else UNSAFE.<br>7. Visualize and Output results with annotations and alerts. |

**Table 1**

**Model training configurations**

| Model | Optimizer | Learning Rate | Batch Size | Epochs |
|---|---|---|---|---|
| PPE Detection | SGD | 0.005 | 8 | 100 |
| Trade Classifier | Adam | 0.0001 | 16 | 50 |

## 3. System Implementation and Module Design

3.1 *Comparative Analysis with previous work*

The trade classification achieved a total of 92% accuracy with metal workers showing exceptional strength of 0.92 precision and recall. The model produced relatively poorer recall on heavy-machine operators, 74%, which is the result of visual similarity between these professions in certain orientations. These results significantly outperform 74 % performance in [4], thus highlighting the benefit that our augmented training regime provides. The confusion matrix of the trade-classifier also shows that the majority of the misclassifications were between these two trades that were mechanically oriented and that the differences among other occupations (e.g., carpenters versus concrete workers) were reliably determined.

The results of tool detection were excellent with the YOLO-based model earning the mean average precision (mAP50) of 99.5 % at 50% overlap (mAP50) of the validation data. This close-to-perfect result makes it possible to detect safety-critical equipment e.g. voltage testers and welding torches, that were so often ignored by earlier methods [4]. The tool detector precision-recall curve showed a steady performance on all confidence levels indicating that the tool can effectively discriminate between tools and background clutter in the industrial environment.

The findings showed that the overall rate of correct decisions regarding safety was 91.2% with the inclusion of all the critical factors (PPE, identity, and trade). The error analysis had found that most of the misclassifications were deeper when there were extremely challenging conditions like heavy obscuration by equipment, or extreme camera angles, that also pose significant challenges to human inspectors. The system in its usual operating conditions had an accuracy of 96.40%, thus justifying its usability in real-life situations. Figure 2 shows annotated output and Figure 3 shows Multi worker tracking with autonomous safety status marking.

## 4. Results and Discussion

The proposed system integrates PPE detection, facial recognition, and trade classification (shown in Table 2 and Figure 4) to solve the limitations of single-modality solutions [10] and at the same time allow simultaneous PPE violation detection and worker certification verification by the safety officers. The performance constraints occur with the low-visibility conditions, full-face respirator cover, and other trade classifications that are visually similar [5], [8]. Incorporation of thermal imaging, the use of time sequence to analyze trade and data policy preserving privacy are only some of the future enhancement directions [17], [18], [19], [20].

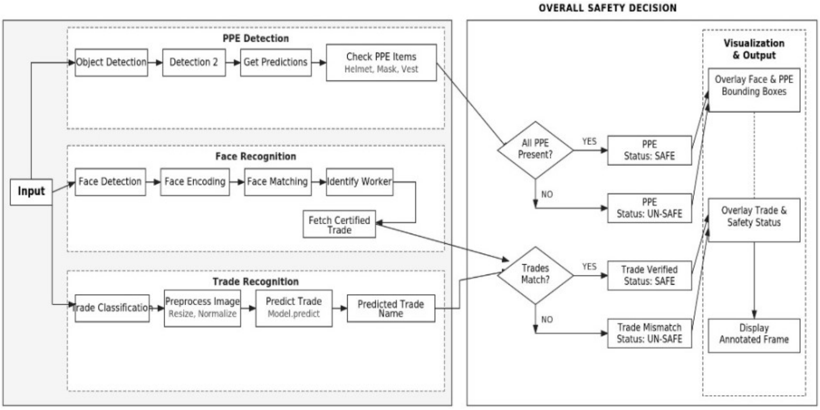


**Figure 1**
**Integrated Pipeline flow**

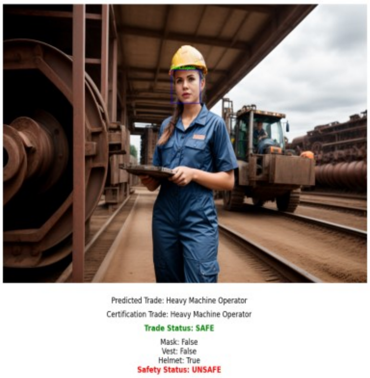


**Figure 2**
**Annotated output**


**Figure 3**
**Multi worker tracking with autonomous safety status marking, which illustrates the capability of the system to tackle complex industrial scenarios**

**Table 2**
**Performance comparison with prior works [11]**

| Metric | Proposed System | [1] | [12] | [4] |
|---|---|---|---|---|
| PPE Detection AP | 82.33% | 70.0% | - | - |
| Face Recognition Accuracy | 95.35% | - | 94.23% | - |
| Trade Classification | 92% | - | - | 74% |
| Tool Detection mAP50 | 99.5% | - | - | 72.3% |

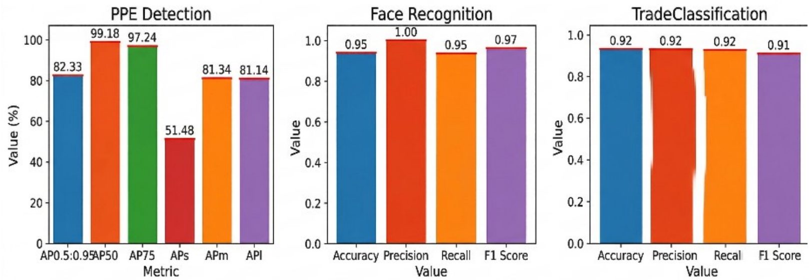


**Figure 4**
**Performance of the three modules (PPE detection, Face Recognition and Trade Classification Modules)[9], [11], [13], [14], [15], [16]**

## 5. Conclusion

The current research has shown that multimodal deep learning can be successfully applied to combine PPE detection, face recognition, and trade classification into a multifunctional system of monitoring industrial safety-compliance. The suggested methodology helps to overcome the essential flaws of a single-modality approach by capitalizing on the interdependence of the context among worker identity, authorized trade, and necessary protective equipment. Empirical performance measures (82.33% average accuracy in detecting PPE, 95.35% average accuracy in face recognition, and 92% average accuracy in trade classification) testify to the practicability of the system, and demonstrate its superiority over the existing methods based on its accuracy and the ability to adjust to different operational conditions. The adaptive learning strategies should be explored in the future to meet the changing PPE designs and the changing workforce demographics and also the ethical aspects of workplace monitoring have to be taken into consideration. This can be enhanced by including other sensory modalities to enhance resilience in harsh environmental scenarios. Together, this literature provides a paradigm base of context-aware safety systems that go beyond binary compliance verification, as well as evolve to smart and personalized risk measures in industrial environments.

## References


[1] N. D. Nath, A. H. Behzadan, and S. G. Paal, "Deep learning for site safety: Real-time detection of personal protective equipment," *Autom. Constr.*, vol. 112 (Apr. 2020), doi: 10.1016/j.autcon.2020.103085.

[2] X. Yang, Y. Yu, S. Shirowzhan, S. M. E. Sepasgozar, and H. Li, "Automated PPE-Tool pair check system for construction safety using smart IoT," *Journal of Building Engineering* (2020), doi: 10.1016/J.JOBE.2020.101721.

[3] M. I. B. Ahmed, L. Saraireh, A. Rahman, S. Al-Qarawi, A. Mhran, J. Al-Jalaoud, D. Al-Mudaifer, F. Al-Haidar, D. AlKhulaifi, M. Youldash, and M. Gollapalli, "Personal Protective Equipment Detection: A Deep-Learning-Based Sustainable Approach: A Deep-Learning-Based Sustainable Approach," *Sustainability,* (2023), doi: 10.3390/SU151813990.

[4] L. Hallonqvist and M. Cromsjö, "Detection of safety equipment in the manufacturing industry using image recognition," Master's thesis, Dept. Comput. Inf. Sci., Linköping University, Linköping, Sweden (2021)

[5] J. Al-Dulaimi, J. Cosmas, and M. Abbod, "Smart Health and Safety Equipment Monitoring System for Distributed Workplaces," *De Computis* (2019), doi: 10.3390/COMPUTERS8040082.

[6] A. Stojadinović, N. Stojanović, and L. Stojanović, "Dynamic monitoring for improving worker safety at the workplace: use case from a manufacturing shop floor," *DEBS* (2015), doi: 10.1145/2675743.2771881.

[7] A. M. Vukićević, M. Petrović, P. Milošević, A. Peulić, K. Jovanović, and A. Novaković, "A systematic review of computer vision-based personal protective equipment compliance in industry practice: advancements, challenges and future directions," *Artif. Intell. Rev.* (2024), doi: 10.1007/S10462-024-10978-X.

[8] V. S. K. Delhi, R. Sankarlal, and A. Thomas, "Detection of Personal Protective Equipment (PPE) Compliance on Construction Site Using Computer Vision Based Deep Learning Techniques," *Front. Built Environ.,* vol. 6 (Sep. 2020), doi: 10.3389/fbuil.2020.00136.

[9] M. Nain, S. Sharma, and S. Chaurasia, "Authentication control system for the efficient detection of hard-hats using deep learning algorithms," *Journal of Discrete Mathematical Sciences and Cryptography,* vol. 24, no. 8, pp. 2291–2306 (2021), doi: 10.1080/09720529.2021.2011109.

[10] N. D. Nath, A. H. Behzadan, and S. G. Paal, "Deep learning for site safety: Real-time detection of personal protective equipment," *Autom. Constr.* (2020), doi: 10.1016/J.AUTCON.2020.103085.

[11] M. Ali, A. Diwan, and D. Kumar, "Attendance System Optimization through Deep Learning Face Recognition," *International Journal of Computing and Digital Systems* (2024), doi: 10.12785/IJCDS/1501108.

[12] M. Nain, S. Sharma, and S. Chaurasia, "Computer vision based human action recognition of construction workers," in *AIP Conference Proceedings, American Institute of Physics* (Jan. 2025). doi: 10.1063/5.0247833.

[13] M. Nain, S. Sharma, and S. Chaurasia, "Deep Learning-Based Safety Assurance of Construction Workers: Real-Time Safety Kit Detection," in *Innovations in Computational Intelligence and Computer Vision,* S. Roy, D. Sinwar, N. Dey, T. Perumal, and J. M. R. S. Tavares, Eds., Singapore: Springer Nature Singapore, pp. 279–293 (2023).

[14] M. Nain, S. Sharma, and S. Chaurasia, "Safety and Compliance Management System Using Computer Vision and Deep Learning," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1099, no. 1, p. 012013 (Mar. 2021), doi: 10.1088/1757-899x/1099/1/012013.

[15] M. Nain, S. Sharma, and S. Chaurasia, "Pandemic Management Using Artificial Intelligence-Based Safety Measures," in *Advances in Medical Technologies and Clinical Practice*, pp. 86–110 (2021). doi: 10.4018/978-1-7998-7188-0.ch007.

[16] M. Nain, S. Sharma, and S. Chaurasia, "Deep Learning Oriented Smart Constructional Material and Face Recognition System for the Site Safety," in *2024 IEEE 11th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON 2024)*, Institute of Electrical and Electronics Engineers Inc. (2024). doi: 10.1109/UPCON62832.2024.1098302

[17] S. Sharma, L. Raja, V. Bhatnagar, D. Sharma, S. N. Bhagirath, and R. C. Poonia, "Hybrid HOG-SVM encrypted face detection and recognition model," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 25, no. 1, pp. 205–218 (2022). DOI: 10.1080/09720529.2021.2014141

[18] N. Kumar and D. Bhargava, "A scheme of features fusion for facial expression analysis: A facial action recognition," *Journal of Statistics and Management Systems,* vol. 20, no. 4, pp. 693–701 (2017), doi: 10.1080/09720510.2017.1395189.

[19] P. Sharma, M. S. P. Sharma, and R. S. Tomar, "A new approach for image segmentation using improved k-means and ROI saliency map," J*ournal of Information and Optimization Sciences*, vol. 38, no. 6, pp. 927–935 (2017), doi: 10.1080/02522667.2017.1372138.

[20] L. J. B. Andrews and L. Raja, "A study on m-health in line with the sensors applying for a real time environment," *Journal of Statistics and Management Systems*, vol. 20, no. 4, pp. 659–667 (2017), doi: 10.1080/09720510.2017.1395186.

*Received April, 2025*