iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: IX Month of publication: September 2025

DOI: https://doi.org/10.22214/ijraset.2025.74380

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 13 Issue IX Sep 2025- Available at www.ijraset.com

YOLOv7-Based Deep Learning Model for Detecting Full Safety Gear on Construction Workers

Vyshnavi M[1], Abhijeeth S[2], Charan H D[3], Hemanth Gowda J[4], Arun Kumar M[5]
[1]Assistant Professor Dept. of CSE, Sapthagiri College of Engineering
[2, 3, 4, 5]Dept. of CSE, Sapthagiri College of Engineering

Abstract: The safety of construction workers is critical aspect of the industry, as a lack of compliance with protective measures often leads to serious accidents. This research presents a computer vision-based solution that employs deep learning to automatically recognize personal protective equipment (PPE), such as helmets, gloves, safety jackets, goggles, and protective footwear. The system is implemented using the YOLOv7 object detection framework, which has been trained on a carefully prepared custom dataset. Each image in the dataset was annotated with bounding boxes to indicate the position and category of safety gear. After multiple training cycles, the model demonstrated strong recognition ability across different PPE types. Evaluation metrics, including precision, recall, F1-score, and mean Average Precision (mAP@0.5), confirm the effectiveness of the approach, with the best performance achieving an mAP of 87.7%. These outcomes highlight the potential of the proposed system to support real-time monitoring of safety compliance on construction sites

I. INTRODUCTION

Construction sites are among the most hazardous work environments, with a significant number of injuries and fatalities resulting from the absence or improper use of safety gear. Reports from government and safety boards reveal that in many incidents, workers either failed to wear helmets correctly or neglected other essential protective equipment, accounting for a large proportion of workplace accidents. Recent developments in artificial intelligence, particularly deep learning, have enabled the automation of visual tasks that were previously dependent on human supervision. Leveraging these advancements, this study focuses on developing a reliable framework to detect whether construction workers are wearing complete safety equipment. The core objective is to identify critical PPE items—helmets, goggles, jackets, gloves, and boots—with high accuracy and in real time.

The YOLOv7 object detection algorithm serves as the backbone of this system, offering both efficiency and speed in identifying multiple objects within an image. Making it suitable for supervised learning. One of the challenges faced was dealing with small or blurred objects in noisy construction environments, which often reduced detection accuracy. Nevertheless, the systematic training process, combined with performance evaluation using precision, recall, F1-score, and mAP, ensured the robustness of the proposed approach. Novelty of this work lies in its ability to integrate computer vision with occupational safety requirements, thereby supporting real-time monitoring on construction sites.Not only enhances detection accuracy but also contributes to enforcing safety regulations and reducing accident risks in the construction industry.

II. LITERATURE SURVEY

Several researchers have explored deep learning methods to improve safety equipment detection in industrial and construction environments. Their contributions form the foundation for this study.

Yang Li et al. [1] developed a helmet detection framework for real-time use on construction sites. Their model was built using SSD-MobileNet, which is based on convolutional neural networks. The system achieved a mean Average Precision (mAP) of 36.82% with a precision rate of 95%, demonstrating reasonable effectiveness but with limitations in recall.

In mining-related applications, another study [3] introduced FM-YOLOv7, which incorporated a fused-MBCA module to enhance feature extraction. The system also adopted an efficient intersection-over-union loss function, allowing for faster convergence. This method achieved an mAP@0.5 of 85.7%, showing significant promise for industrial environments.

Chen et al. [4] improved the Tiny YOLOv3 framework and achieved 97.24% along with mAP of 95.56%. In another study, Hanetal. [5] applied YOLOv5 for helmet detection, reporting a mAP of 92.2%. To make the system easier to use, they further developed a graphical user interface.

Similarly, Santosh et al. [6] introduced a YOLOv3-based method for real-time detection of personal protective equipment (PPE).

Their approach reached an accuracy of 96.51%, with precision, recall, and F1-scores all close to 0.97. In related research, Ieamsaard et al. [7] designed a YOLOv5 model for mask detection and achieved 96.5% accuracy using a publicly available dataset of 853 images.

Beyond PPE, YOLO has also been successfully adapted for other application areas. For example, Wangetal. [8] suggested a YOLOv7-tiny–based vehicle identification method, achieving 80.8% with a lighter model suitable for real-time monitoring. SaiShilpaetal. [9] worked with YOLOv3 on the COCO dataset, which contains 91 classes, though their study used 80. They reported nearly perfect detection results with 100% accuracy at most thresholds.

Dr. S.V. Viraktanath et al. [10] analyzed YOLO's architecture and demonstrated its capability for classification tasks when trained on domain-specific datasets.

Improvements over YOLOv5 were proposed by Hyun-Ki Jung et al. [12], who used drone-based datasets to train an enhanced version of YOLOv5. Their model achieved 90.7% precision, 87.4% recall, and a mAP of 95.5%, outperforming the original YOLOv5. AhatshamHayat [13] also implemented YOLOv5x on a dataset of 5,000 helmet images, recording an mAP of 92.44% with accuracy values above 89%.

A study by Jye-HwangLo [14] trained a model with 11,000 images and 88,725 PPE labels, testing YOLOv3, YOLOv4, and YOLOv7. Their results showed YOLOv7 achieving 97.95%, precision of 92.25%, recall of 98.59%, and F1-score of 95.31%. Similarly, Omer Kaya et al. [15] compared Faster R-CNN with YOLOv7 for pedestrian crosswalk detection.

Jing Hu et al. [16] applied an optimized YOLOv3 for helmet detection, obtaining a mAP of 93.5% at 35 fps on a dataset of 20,554 images. Their dataset was divided into positive and negative classes to identify workers wearing or missing helmets. In a different application, Steven Kolawole et al. [17] worked with NSL datasets, training YOLO models that reached 80.57%, recall of 95%, & map of 95%.

Further improvements were seen in Fangbo Zhou et al. [18], who trained several YOLO models on 6,045 images, with YOLOv5s achieving 110 fps and YOLOv5x reaching an mAP of 94.7%. WendongGaietal. [19] enhanced YOLOv7, achieving higher accuracy with a processing speed of 112.4 fps, using a large-scale helmet dataset.

Finally, WeiFangetal. [20] and XiangLongetal. [21] proposed variations of YOLO models for helmet detection, reporting mAP values of 65.7% and 45.20%, respectively. While these results showed limited performance.

III. METHODOLOGY

In this research, we plan to apply a deep learning approach using the YOLO algorithm to detect the different types of safety gear that construction workers are required to wear. To achieve reliable results, an appropriate dataset will be collected, carefully labelled, and then split into training, validation, and testing sets for the model.

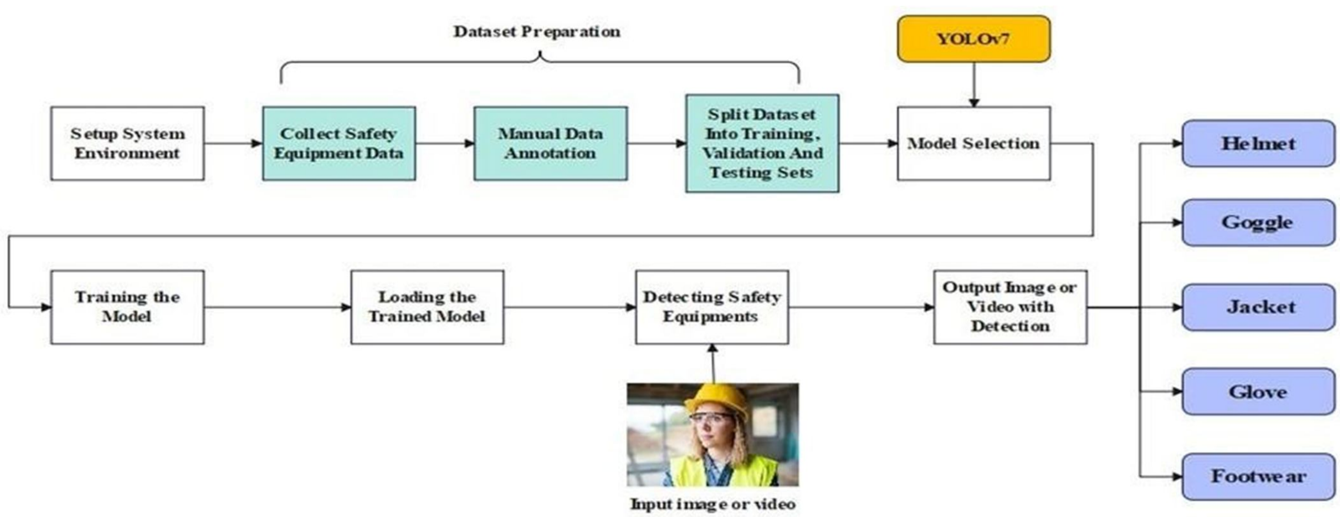


Fig. 1. Safety Equipment Detection Workflow Diagram

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IX Sep 2025- Available at www.ijraset.com

A. Research Focus and Instruments

The primary goal of this study is to design a computer vision system capable of identifying whether construction workers are wearing all required safety equipment. The target equipment includes helmets, jackets, gloves, goggles, and protective footwear. For object detection, we adopted the YOLOv7 (You Only Look Once) architecture, which is well known for its speed and accuracy in real-time applications. The model was set up to run in a virtual machine, which required two types of computing resources—one on the local system for training and another within the virtual machine itself. The local computer used for training was equipped with an Intel Core i5 (8th generation) processor clocked at 4.0 GHz and 16 GB of RAM. For the virtual machine, a Tesla V100 GPU was provided, offering 15 GB of memory, 12.64 GB of RAM, and 78 GB of storage space. The development environment was built around Python 3.10 and PyTorch, while Google Colab was used as the cloud platform for carrying out both training and testing tasks.

B. Data Collection

Due to limited access to real construction sites, dataset preparation relied mainly on online sources. Images and video frames were collected from YouTube construction-related footage, Google Images, and specialized construction websites. These sources provided diverse examples of workers wearing different types of PPE. Overall approximately 1,000 images were gathered. Three subsets—70% for training, 15% for validation, and 15% for testing.

C. Training the Dataset

The dataset was divided into training, testing, and validation sets, ensuring each subset contained proportional distributions of PPE categories. The training was run for multiple epochs, allowing the model to gradually minimize its loss function and improve prediction accuracy. Similar training experiments were also performed on YOLOv5s, YOLOv5m, and YOLOv7-x for comparative analysis. Indicators helped evaluate the learning progress and fine-tune the parameters. Performance was assessed at different confidence thresholds to measure how well the system detected PPE across multiple scenarios.

Fig. 3. Manual Data Annotation using Labeling Annotator tool

```
0 0.234375 0.120863 0.175781 0.135252
0 0.497559 0.276978 0.114258 0.096403
0 0.798340 0.217986 0.137695 0.119424
1 0.260742 0.161151 0.095703 0.066187
1 0.503418 0.323741 0.081055 0.040288
1 0.797363 0.284892 0.106445 0.048921
2 0.200684 0.561871 0.245117 0.671942
2 0.474609 0.668345 0.222656 0.505036
2 0.822754 0.634532 0.264648 0.546763
3 0.323242 0.736691 0.113281 0.141007
3 0.467285 0.695683 0.088867 0.076259
3 0.512695 0.648921 0.083984 0.120863
3 0.753906 0.551079 0.111328 0.109353
```

Fig. 4. text file (.txt) generated after saving the above

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IX Sep 2025- Available at www.ijraset.com

D. Training the Dataset

The annotated dataset was separated into three parts: training, testing, and validation. This division helped ensure that the model could be trained effectively and then ion of each safety gear category.

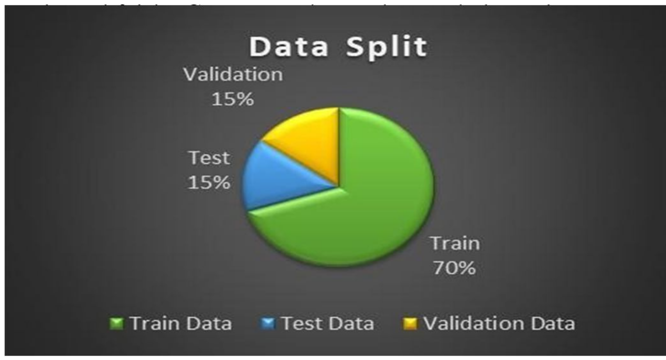


Fig. 5. Data Distribution

E. Network Architecture of YOLOv7

YOLO models are designed as single-stage object detectors, combining speed with accuracy. The architecture includes four main components:

Backbone: Based on Darknet-53, responsible for extracting feature maps from input images.

Neck: Combines multi-scale features using a feature pyramid to improve detection of objects of varying sizes.

Head: Generates bounding boxes and class probabilities for detected objects.improved accuracy in detecting tiny objects. Input Module: Handles preprocessing and resolution scaling.

One of the key advantages of YOLOv7 is its ability to process images at higher resolution (608×608 pixels compared to 416×416 in YOLOv3) while maintaining high speed. The use of anchor boxes with varied aspect ratios enables better detection of objects with different shapes and sizes. Additic and partially obscured objects

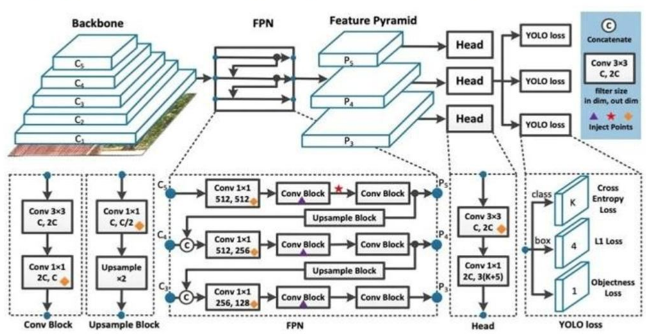
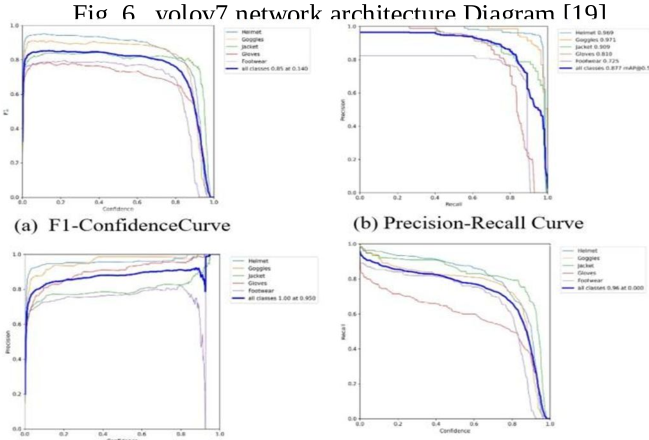


Fig. 6. yolov7 network architecture Diagram [19]



(a) F1-ConfidenceCurve
(b) Precision-Recall Curve
(c) Precision Curve
(d) Recall Curve

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IX Sep 2025- Available at www.ijraset.com

IV. RESULT AND DISCUSSION

This section present the performance evaluation of the proposed model, highlighting its strengths, limitaions, and areas for improvement. After completing the data collection and pre processing stages, the YOLOv7 framework was applied to detect saftety equipments. The dataset was divided into three subsets: 701 images for training and 151 images each for testing and validation resulting in a total of 1003 labelled images. The Model's effectiveness was assessed using common performance metrics such as precision. Recall F1- score and mean Average precision (mAP). These Metrics allowed for a balanced evaluation of detection accuracy and robustness. The Results demonstrate that the system is capable of reliable identifying safety gear across different image samples. While the overall outcomes are encouraging, the analysis also reveals areas where detection could be further optimized, particularly in handling variations in lighting, occlusion, and worker posture.

A. *Comparison of Trained Models*

To assess the effectiveness of the YOLOv7, YOLOv7-X, YOLOv5s, and YOLOv5m models. Among these, YOLOv7 demonstrated the strongest performance, achieving a precision of 84.1%, recall of 87.1%, and an F1-score of 85.0%. Its mean Average Precision at IoU threshold 0.5 (mAP@0.5) was recorded as 87.7%, which was higher than other tested variants.The YOLOv7-X model followed closely, with an mAP@0.5 of 86.0%, precision of 87.3%, and recall of 86.1%. While this model showed competitive results, its performance remained slightly below the baseline YOLOv7. In contrast, YOLOv5s and YOLOv5m recorded noticeably lower metrics, with F1-scores of 78.0% and 74.9% respectively, and mAP@0.5 values of 81.1% and 75.5%.These findings confirm that YOLOv7 offers a balanced combination of precision, recall, and accuracy, making it the most suitable model for detecting safety equipment on construction sites.

Fig. 8. Output of YOLOv7

B. *Performance Evaluation Using Precision-Recall Curves*

The precision-recall analysis revealed the relationship between detection confidence and accuracy. The goggles category achieved the highest detection accuracy at 97.1%, whereas footwear performed the worst with 72.5%. When aggregated across all categories, the system achieved an average precision-recall accuracy of 87.7% and recall accuracy of 96.0%.

C. *Video-Based Evaluation*

In addition to static image testing, the model's performance was examined using video data. A constructionrelated video containing 3,715 frames was processed in 68.689 seconds, corresponding to an average of nearly 70 frames per second (FPS). This high FPS rate demonstrates the model's ability to function in near real-time scenarios, making it suitable for live monitoring applications on construction sites.

D. *Challenges in Detection*

Despite strong overall performance, the system faced limitations when detecting PPE items in low-quality images or cluttered environments. Small and partially obscured objects, such as gloves and shoes in complex backgrounds, were often missed or misclassified. These issues highlight the need for dataset expansion and improved feature representation techniques.

E. Comparison with Existing Studies

The performance of the proposed YOLOv7-based model was compared with results reported in earlier studies.

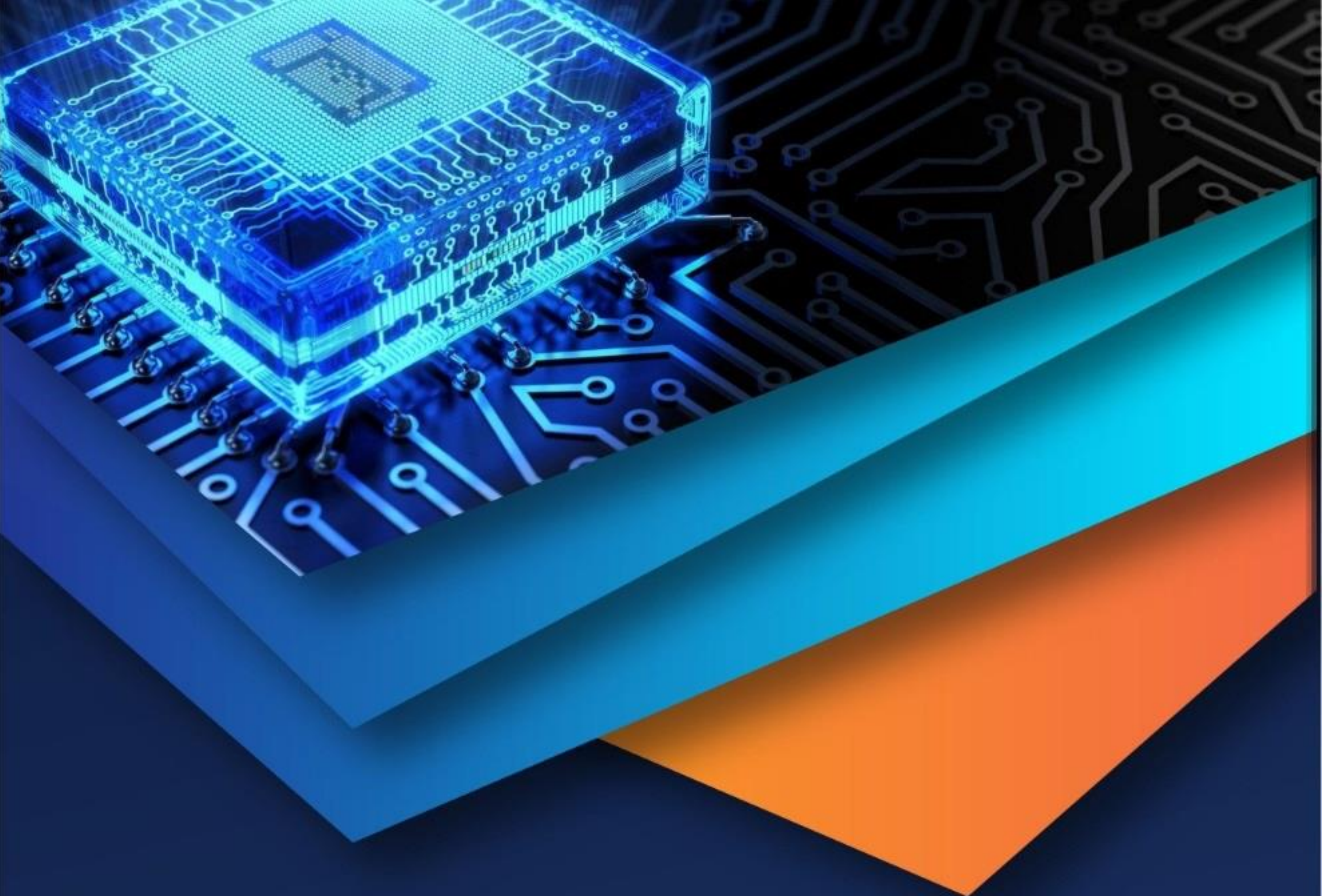
These results indicate that the YOLOv7-based framework outperforms several existing approaches, offering more reliable and accurate detection of PPE.

V. FUTURE WORK

Looking ahead, there are several directions in which this project can be extended. A key improvement would be to expand the. This would make the model more adaptable and robust in handling different real-world scenarios. Another promising direction is to explore newer deep learning architectures, such as YOLOv8 or other state-of-the-art models, which may provide higher accuracy and faster detection. Finally, integrating the detection system into realtime monitoring platforms or wearable devices could significantly strengthen on-site safety management and provide immediate feedback to workers and supervisors.

VI. CONCLUSION

This project presented a YOLOv7-based model designed to detect essential safety equipment. The system achieved an average mAP@0.5 of **0.877**, demonstrating its effectiveness in accurately identifying and classifying different types of protective gear. These results highlight the potential of computer vision in promoting safety compliance on construction sites. By providing a reliable and efficient method for monitoring personal protective equipment, the system can help reduce the risk of accidents, support safety enforcement, and contribute to safer working conditions in the construction industry.



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY