Enhancing Construction Site Safety: Personal Protective Equipment Detection Using Yolov11 and OpenCV

Ruchitraa Rajagopal, Harshitha Pulluru, Aaradhya Joshi and Shanmuganathan C.
Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India

Keywords: Construction Site Safety (CSS), Personal Protective Equipment (PPE), You Only Look once (YOLO), Real-Time Object Detection (RTOD), Site-Risk Assessment (SRA), Worker Safety Compliance (WSC), Machine Learning (ML), Computer Vision (CV).

Abstract: One of the most emerging industries in the world is construction. Even after installing a lot of safety rules and regulations, construction site accidents are a major threat. This is because of non-compliance with the assigned safety protocols and the lack of a system assigned for supervision. The paper provides a model that can automate the safety system using Computer Vision (CV) and Machine Learning (ML) techniques. The model analyzes the surveillance footage to detect workers and their personal protective equipment (PPE), such as safety vests, masks, and hard hats. It also calculates the distance between workers and machinery and sends alerts in case of unsafe distances. The system also provides a graphical user interface where the site managers or supervisors can easily monitor the surveillance footage with the model results in real time. The warnings are stored in a database along with a video clip and other essential information. This way, the supervisors can view the database and get an idea about the type of warnings and when and where they occurred. This way, they can guide the workers regarding site safety. The model used by the system is YOLOv11. It is trained and tested to give better results in various weather conditions. The model achieves a mAP of 81% at 0.5 IoU and a mAP of 60.3% at 0.5 to 0.95 IoU, with an overall accuracy, precision, and recall of 97%, 87%, and 76%, respectively. It is computationally efficient, with 14.7GB FLOPs, 6.4MB parameters, and an inference speed of 2.4ms, making the model applicable for real-time analysis.

1 INTRODUCTION

Construction sites are naturally as well as inherently dangerous environments where the worker's safety is a constant challenge. The presence of heavy machinery, hazardous materials, and complex workflows increases the likelihood of accidents, thus making compliance with safety regulations critical. While PPE such as safety vests, masks, and hard hats play a key role in injury prevention, monitoring them manually takes up much time, and is often ineffective in large-scale sites Chen et al., (2021).

Traditional safety methods, which were quite laborious, such as manual checking and surveillance cameras, are often slow and difficult to scale, making it difficult to detect and prevent accidents. Advancements in the field of CV and ML have opened a lot of new techniques to automate safety monitoring. AI-powered detection models, such as YOLO (You Only Look Once), have shown promising results in identifying safety gear and worker behavior from video feeds. However, challenges such as detecting small objects, handling occlusions, and maintaining accuracy in varying lighting and environmental conditions remain. This research introduces a safety monitoring system that leverages AI and integrates YOLOv11 to identify PPE and monitor worker proximity to heavy machinery in real-world systems Chen et al., (2021).

Included in the system is a graphical user interface (GUI), where the site managers can monitor the workers and their compliance with safety protocols. The proposed system improves detection accuracy by incorporating multi-scale feature extraction, attention mechanisms, and bounding box regression Feng et al., (2024). Mosaic data augmentation is applied to the training part of the set, while the testing part is modified to simulate various weather conditions, such as high brightness, dust, and motion blur. To further improve site safety, the system includes a proximity

Rajagopal, R., Pulluru, H., Joshi, A. and C., S.
Enhancing Construction Site Safety: Personal Protective Equipment Detection Using Yolov11 and OpenCV.
DOI: 10.5220/0013882800004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 2*, pages 344-353
ISBN: 978-989-758-777-1

detection algorithm that helps prevent worker-machinery collisions.

This paper discusses the methodology, implementation, performance evaluation, and also potential future improvements of the proposed system. The findings highlight how AI-driven monitoring can transform workplace safety, minimize accidents, and enhance regulatory compliance in high-risk industries.

2 RELATED WORKS

One of the most emerging industries in the world construction. Although these industries contribute a lot to the nation's GDP, even after installing a lot of safety rules and regulations, construction site accidents are a major threat. Therefore, a lot of methods are being implemented to reduce the risk of accidents. Some of these methods include the Safety Detection method based on improved YOLOv8, integrating an AI model for construction site safety, and using Fast R-CNN and CNN to detect bounding box coordinates. A summary of these findings is presented in Table 1.

Table 1: Summary of Findings.

S.No.	Paper's Name	Targeted Object	Findings
1	"Safety Detection Method based on improved YOLOv8"	Safety Helmet	Proposed study includes an improved algorithm YOLOv8n-SLIM-CA. It uses mosaic data augmentation and coordinate attention mechanism.
2	"Deep Learning Based Workers Safety Helmet Wearing Detection on Construction Sites Using Multi-Scale Features"	Safety Helmet	This study focuses on the addition of multi-scale features and attention mechanism along with the baseline YOLOv5 model.
3	"A Novel Implementation of an AI-Based Smart Construction Safety Inspection Protocol in the UAE"	Safety Harness	This study focuses on developing a CNN-based model to detect safety harness. The deep learning network uses YOLOv3.

One of the articles, "A Novel Implementation of an AI-Based Smart Construction Safety Inspection Protocol in the UAE", focuses on integrating AI, precisely a deep learning approach which is used to supervise and detect safety violations Shanti et al., (2021). This paper focuses on the development of a CNN-based technique for safety that provides supervising workers working at construction sites using a real-time detection and monitoring algorithm, YOLOv3. It trains a CNN that is used for detecting equipment such as safety vests and safety helmets. The main challenge that this method faces is the difficulty of obtaining surveillance video and training all the data sets required by the CNN models.

Another research paper that proposes using CV and ML for the detection of safety helmets is "Safety Helmet Detection Based on Improved YOLOv8" Lin, B, (2024). Safety helmets are essential for protecting workers from head injuries on construction sites, but relying on manual supervision to ensure compliance can be inefficient and prone to mistakes. Deep learning models like YOLO have made helmet detection possible in real-world implementation, but they often struggle with spotting small or partially hidden helmets in busy environments. To overcome it, the paper uses YOLOv8n-SLIM-CA. This improved detection model enhances accuracy using Mosaic data augmentation, a Slim-Neck structure, and Coordinate Attention. These upgrades help the model focus better on safety helmets, reduce complexity, and improve detection in challenging conditions.

Compared to the standard YOLOv8n, the model boosts accuracy by 2.151% (mAP@0.5), reduces model size by 6.98%, and lowers computational load by 9.76%, making it faster and more efficient Lin, B, (2024). Tested on the Safety Helmet Wearing Dataset (SHWD), it outperforms other detection models by identifying helmets more accurately, even in crowded, distant, or cluttered backgrounds. This makes YOLOv8n-SLIM-CA a powerful tool for real-world safety monitoring Lin, B, (2024). Looking ahead, integrating this model into edge devices could make real-time helmet detection even more practical and accessible for industrial safety.

Another research paper that focuses on a CNN based detection of safety helmets is "Deep Learning Based Workers Safety Helmet Wearing Detection on Construction Sites Using Multi-Scale Features" Han et al., (2022). Ensuring people wear safety helmets on-site is crucial for preventing injuries from falling objects, but traditional monitoring methods can take much time and are inclined to a lot of mistakes. This study presents an improved deep learning approach using YOLOv5, enhanced with a fourth detection scale to identify small objects better and an attention mechanism to improve feature extraction.

To address the challenge of limited training data, targeted data is augmented, followed by the usage of transfer learning resulting in a 92.2% mean average precision (mAP) and an improvement of 6.4% in accuracy, with object detection of just 3.0 ms speed at 640×640 resolution Han et al., (2022). This makes the model precise and applicable for real-time analysis. Thus, automating safety compliance checks minimizes the need for constant manual supervision, allowing site managers to identify and respond to potential risks more efficiently.

3 METHODOLOGY

3.1 Proposed System Architecture

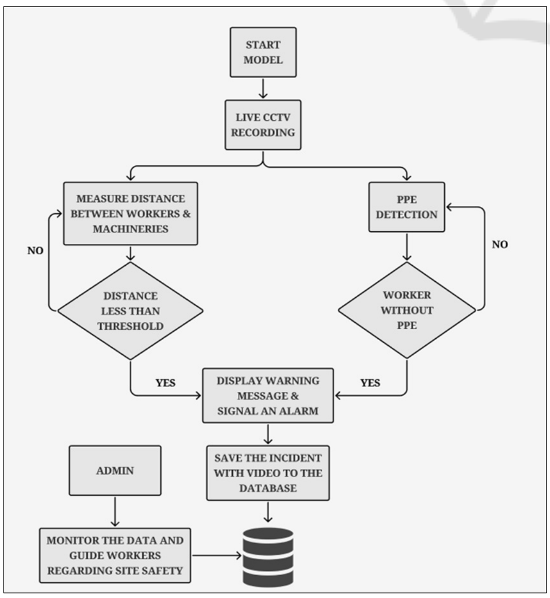

Figure 1: Architecture Diagram.

The model analyzes live video feeds. This ensures real-time monitoring of worker compliance. The model training is done on diverse construction site datasets. So, PPE can be detected in varying lighting and diverse environmental conditions Sridhar et al., (2024). YOLO gives a bounding box as an output for the detected object, which streamlines safety enforcement. YOLOv11 significantly improves workplace safety as well as adherence to various safety rules. Figure 1 shows the proposed system's architecture diagram.

3.2 YOLOv11 Architecture

YOLOv11 is an object detection model applied in a single shot. It is built for real-time applications, which offers both speed and accuracy. Traditional models process images in multiple stages. While YOLOv11 analyzes the entire image in just a single pass, which makes it highly efficient. Its working involves the division of the image into a grid and then the prediction of their object locations and their classifications simultaneously. It comes with improved feature extraction and attention mechanisms. Thus, it excels in identifying small or overlapping objects with greater precision. To enhance its accuracy, it utilizes optimized loss functions like Complete IoU (CIoU) Mahmud et al., (2023). This helps fine-tune object localization as well as minimize incorrect detections. YOLOv11 comes with a lightweight design, which ensures quick processing. Thus, YOLOv11 is a perfect choice for tasks that demand real-time object recognition without compromising performance. Figure 2 shows the architecture of the YOLOv11 model.

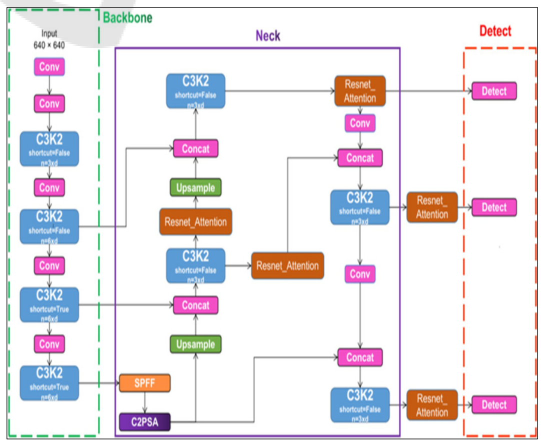

Figure 2: YOLOv11 Architecture Diagram.

3.3 Single Shot Detector

SSD uses the technique of Non-Max Suppression, which is used to eliminate any duplicate detections that would occur. So, by using non-max suppression, only the most relevant bounding boxes will be retained. In dynamic construction sites, speed and accuracy are of utmost importance when we need to employ it for real-time detection of PPE. The architecture of the SSD is also lightweight. This is essential as it can ensure easy deployment on edge devices Jankovic et al., (2024). This also enables the system to work without relying on high-end hardware. Such deployment in construction sites allows the workers and supervisors to be provided with instant alerts (2024). Thus, SSD plays a huge role in enhancing worker safety. So, the integration of SSD and OpenCV with deep learning techniques enhances the model and allows it to perform real-time safety monitoring with minimum latency Shetty et al., (2024). The reliability of the model has also increased because of its ability to work in different lighting conditions and various environmental variations. Through proper training, SSD can improve compliance enforcement and significantly reduce workplace hazards.

3.4 Bounding Box Regression for PPE Detection and Machinery Proximity

Bounding box regression is a fundamental technique used in object detection models like YOLO and SSD to localize objects within an image precisely. In the context of construction site safety, bounding box regression is used to detect PPE compliance by identifying helmets, vests, gloves, and other safety gear. Merely enclosing the detected objects in bounding boxes is not sufficient. Thus, Bounding Box Regression calculates x, y, width and height. Bounding Box Regression is important in construction site monitoring. By calculation of Euclidean distance between the workers and machinery, proximity alerts are designed to trigger when the computed distance between the two bounding boxes falls below a certain threshold. Such trigger alerts help minimize the chances of accidents by sending proactive notifications to the workers as well as the supervisors Al-Azani et al., (2024). The method also benefits from non-maximum suppression (NMS). Bounding box regression accuracy depends on proper dataset labelling and model training with diverse construction site images. The attention mechanism is one of the advancements of deep learning, and it further refines the accuracy of bounding boxes. This ensures precise detection even in complex environments. Thus, the BBR technique can be efficiently used to optimize the real-time monitoring of safety compliances Gautam et al., (2024). When combined with an alert system, workplace safety is enhanced as the hat workers are instantly made aware of potential hazards.

3.5 Attention Mechanism-Based CNN

Attention mechanisms in deep learning have revolutionized object detection by improving feature extraction and focusing on the most relevant image regions. In PPE detection for construction site safety, attention mechanisms help CNN models prioritize critical areas, such as worker faces, helmets, and vests, while ignoring irrelevant background details. So, the approach of ignoring irrelevant parts and focusing on significant features improves detection accuracy in cluttered environments as well as other complex environments. This is done by an attention mechanism based on CNN, which assigns higher weights to the important regions of the image Guan et al., (2024). This ensures that the model captures the local as well as global dependencies efficiently without any compromise. This means that the model is enhanced to detect PPE even in challenging conditions like poor lighting, occlusions or even low-lighting video feeds. Thus, the integration of Attention Mechanism based CNN with YOLO and OpenCV for detection of PPE in real-time achieves greater precision. Construction sites have different environments, which the model can quickly adapt to Ponika et al., (2023). This is a major advantage to ensure safety compliance across such diverse environments. Another advantage is the visual highlight of the area influencing the prediction. Such transparency helps the safety officers to detect errors more easily.

The usage of attention-based mechanisms helps in processing large-scale datasets efficiently. Real-time applications require such machinations to handle large data. Thus, the model training is proceeded with a diverse dataset consisting of various combinations of colors, textures, and placements using this methodology Han et al., (2024). This methodology ensures a reliable and efficient safety monitoring system that proactively mitigates construction site hazards.

3.6 Image Pre-Processing

Image preprocessing also is used to ensure that the model focuses on relevant objects. This is done before enhancing visual clarity. Normalization is applied to scale pixel values between 0 and 1. This is to prevent intensity variations from affecting detection performance. Sobel and Canny filters are edge detection techniques that are used to sharpen image boundaries. This helps in the easy detection of PPE equipment like vests and helmets by the model. The Grayscale Conversion also simplifies the image complexity by reducing the amount of unnecessary data which is processed. This method is done while retaining key object structures. This method is also quite useful in cases of background distractions. Segmentation is a technique that can be used to separate or isolate the workers and their PPE from the cluttered environments. Thus, when segmentation is used, it leads to more precise detections Krishna et al., (2021). Thus, before inputting the input image to modal for training, preprocessing them leads to the deep learning algorithms to work more efficiently. As a result, a robust and adaptable PPE detection system can be achieved.

Image preprocessing is used to enhance visual clarity. Not only this but it is also used for the extraction of features by focusing only on relevant objects. Normalization is such a technique. This is used to scale the pixel values between the range of 0 and 1 inclusive. Thus, even with varying intensities, the detection performance is not affected. There are many edge detection techniques like Sobel and Canny filters. They are used to sharpen the boundary of images. This sharpening ensures easier detection of PPE equipment like safety helmets and vests. In grayscale conversion, the image is retained only of its key structures. This simplification reduces the unnecessary processing of data. When there are background distractions, segmentation techniques can be used Gautamet al., (2024). This is a very powerful deep-learning technique which isolates the key image parts from unnecessary key structures. Thus, refining images in preprocessing is vital to ensure that model training works efficiently.

3.7 Image Augmentation

This is a technique when we need to artificially expand our datasets. This helps models better generalize to real-world scenarios. PPE detection should be able to work across various lighting conditions, angles and environments for real-time deployment. Thus, augmentation helps the model to be trained with a diverse set of images. Flipping or rotating the images provides the model with various data points. Scaling is also used to adjust the size of the images. Thus, the model trained using various augmentation techniques can recognize PPE from multiple orientations. Thus, the model can reliably identify PPE of different sizes. Random cropping of the images is also added to the dataset Han et al., (2022). This prepares the model to identify PPE even if it is partially visible in cases where vests or helmets may be partially obscured. Thus, the application of augmentation techniques prevents overfitting. It also allows the model to perform reliably even in unpredictable conditions.

Colour-based augmentations are also applied to enhance the detection further. This includes brightness, contrast and hue adjustments. Construction sites often have varying lighting conditions. The adjustment of brightness levels in images ensures that the model is exposed to diverse image lighting. This ensures effectiveness in both well-lit and dim conditions. A contrast enhancement application is used so that clear differentiation of PPE from the background is achieved. Noise is added to the training dataset to simulate real-world scenarios where camera imperfections may occur. This makes the model resilient to real-world scenarios Shanti et al., (2021). Finally, techniques like image overlays and motion blur are applied to the dataset. At the same time, the data recorded by the camera would have images or video live feeds of workers who are partially covered. They might also be moving. Such real-world scenarios need to be considered while training the model Azatbekuly et al., (2024). Using augmentation, the model becomes highly adaptable to such scenarios. As a result, a detection system across challenging and diverse environments is achieved.

4 DATA AND EXPERIMENTAL SETUP

4.1 Computer Configurations

The model was trained on a laptop equipped with a NVIDIA GeForce GTX 1650 Ti GPU. The detection model was implemented using CUDA, Ultralytics, OpenCV, and PyTorch. Table 2 provides the computer configurations.

Table 2: Computer Configuration.

CONFIGURATION	TYPE
System	Windows 11
CPU	AMD Ryzen 7 4800HS (8 cores, 16 threads)
Memory	16GB
GPU	NVIDIA GeForce GTX 1650 Ti
OS type	64-bit
Python	version 3.12.7
NVIDIA Driver	version 555.97
CUDA	version 12.6
Ultralytics	version 8.3.78
PyTorch	version 2.6.0+cu126
OpenCV	version 4.11.0

4.2 Dataset

The dataset consists of 717 images, divided into the following parts: the train part (73%) contains 521 images, the validation part (16%) contains 114 images, and the test part (11%) contains 82 images. Each set includes images and corresponding label files (.txt). The dataset contains 10 labels, numbered 0 to 9: Hardhat, Machinery, NO-Hardhat, Mask, Safety Cone, NO-Mask, Safety Vest, Vehicle, NO-Safety Vest, and Person. The train set is modified using mosaic data augmentation, and each model is trained on that set (2025). The model is then evaluated on a modified set under different conditions, including RGB, grayscale, blur, dust, and maximum brightness, ensuring evaluation across all scenarios. Figure 3 presents the graphical results of the train set.

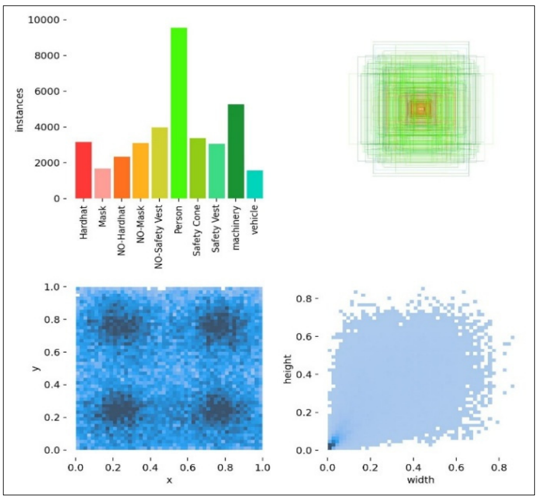

Figure 3: Graphical Results of the Train Set.

4.3 Evaluation Metrics

4.3.1 Mean Average Precision (mAP)

mAP determines the overall detection efficiency by averaging the precision scores across multiple recall thresholds. It is widely used in detection to assess the accuracy of bounding box predictions (2025). For PPE detection, a greater mAP indicates correctly identifying helmets, vests, gloves, and other safety gear with high recall and precision. This is useful in ensuring compliance monitoring in real-time environments.

$$\mathrm{mAP} = \frac{1}{N} \sum_{i=1}^{N} \mathrm{AP}_i \qquad (1)$$

4.3.2 Intersection over Union

IoU evaluates the intersection of the ground truth with the predicted bounding box. A high IoU is better for object localization, which is critical for PPE detection in construction sites. Poor IoU may result in misclassification or failure to detect safety gear, leading to compliance issues. It is commonly used to filter out incorrect detections in object detection tasks.

$$\mathrm{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \qquad (2)$$

where the numerator is the intersection of actual and predicted bounding boxes, and the denominator is their combined area.

4.3.3 Precision and Recall

Precision measures how many detected PPE items are correct, while recall measures the actual PPE items were successfully identified. High precision results in lower false alarms, while high recall results in lower missed detections. In PPE detection, striking a balance is crucial for reliable compliance monitoring in workplaces.

$$\mathrm{Precision} = \frac{TP}{TP+FP} \qquad (3)$$

$$\mathrm{Recall} = \frac{TP}{TP+FN} \qquad (4)$$

4.3.4 Accuracy

Accuracy is a fundamental metric used to determine the overall efficiency of a PPE detection system. It measures the correctly classified instances (PPE and non-PPE) from the total number of predictions (2025). A greater accuracy indicates correctly identifying workers wearing PPE and those without it, minimizing misclassifications. However, accuracy alone may not always reflect model reliability, especially if there is an imbalance in the dataset (e.g., more PPE-wearing workers than non-compliant ones). To get a clearer picture, accuracy is often analyzed alongside precision and recall to ensure a balanced performance.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{5}$$

4.3.5 F1-Score

It is the harmonic mean of recall and precision. It provides a metric to determine the overall efficiency of the PPE detection model. It is particularly useful when there is an imbalance between detected and actual PPE items, ensuring that false negatives and false positives are minimized.

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision+Recall}} \tag{6}$$

A higher F1-score indicates a well-balanced model that effectively detects safety gear while minimizing incorrect detections.

5 RESULTS AND DISCUSSION

5.1 Model Training

We trained the model using the same training set as the default YOLOv11 detection models. We conducted the training on 640 × 640 pixels images for 100 iterations. The accuracy improved as the no. of epochs increased. On the other hand, the box_loss, dfl_loss, and cls_loss gradually decreased to 0.75, 1.04, and 0.56, respectively. Table 4 shows the Comparison of existing object detection models with the proposed model. Figure 4 depicts a decline in loss for both training and validation. Figure 6 (b) shows the precision-recall curve used to analyze mAP. The mAP@0.5 stabilized at approximately 0.771 after 58 iterations and reached 0.810 after 100 iterations, while mAP@0.5-0.95 progressively increased to 0.603. Moreover, Figure 6 (a), (c), and (d) illustrate the F1-score, precision, and recall curves in relation to confidence levels for each class. Accuracy, Precision, Recall, and F1-Score of each class detected by the proposed model Shown in Table 3.

Table 3: Accuracy, Precision, Recall, and F1-Score of Each Class Detected by the Proposed Model.

Class	Accuracy	Precision	Recall	F1-Score
Hardhat	0.970883	0.904762	0.760000	0.826087
Mask	0.989081	0.978261	0.900000	0.937500
NO-Hardhat	0.960874	0.911765	0.626263	0.742515
NO-Mask	0.957234	0.835443	0.660000	0.737430
NO-Safety Vest	0.955414	0.786517	0.700000	0.740741
Person	0.953594	0.720721	0.800000	0.758294
Safety Cone	0.985441	0.928571	0.910000	0.919192
Safety Vest	0.975432	0.939759	0.780000	0.852459
Machinery	0.985441	0.911765	0.930000	0.920792
Vehicle	0.947225	0.791667	0.570000	0.662791

Table 4: Comparison of Existing Object Detection Models with the Proposed Model.

Model	size (MB)	mAP @0.5(%)	mAP @0.5:0.95(%)	Parameters (MB)	FLOPs (GB)	Speed (ms)
YOLO11n	5.4	58.0	39.5	2.6	6.5	1.5
YOLO11s	18.4	70.1	47.0	9.4	21.6	2.5
YOLO11m	38.8	73.3	51.5	20.1	68.0	4.7
YOLO11l	49.0	77.3	53.4	25.3	86.9	6.2
Our Model	10.5	81.0	60.3	6.4	14.7	2.4

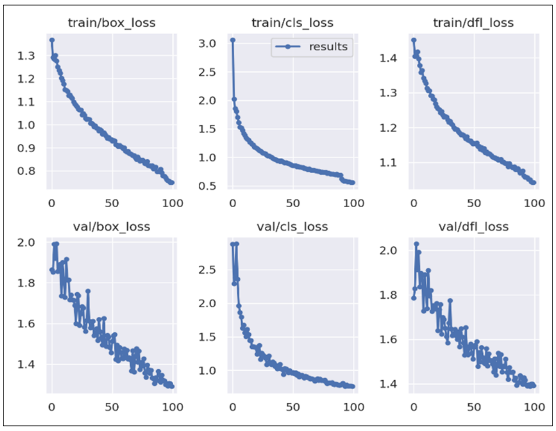

Figure 4: Training and Validation Loss Curves.

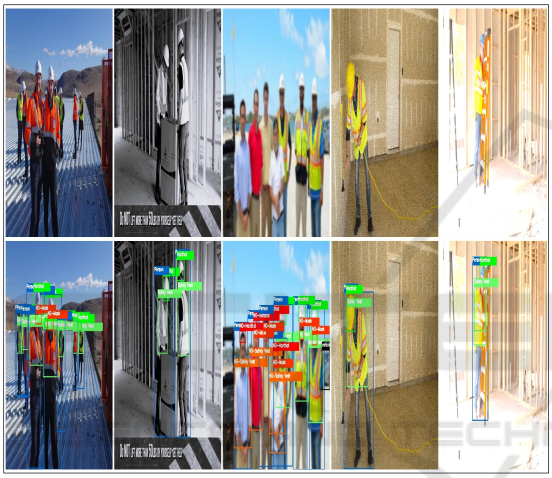

Figure 5: Detection Results Before and After Model Analysis.

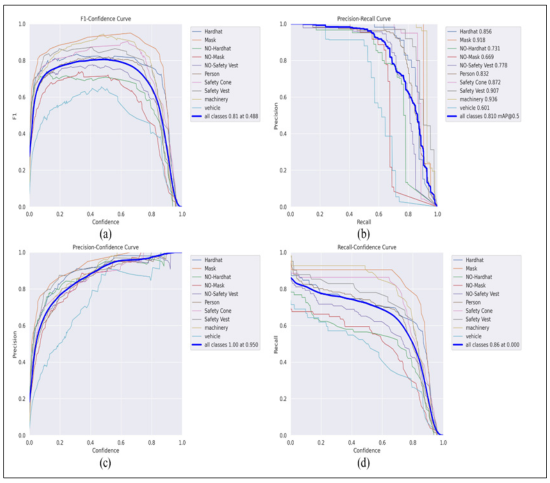

Figure 6: Evaluation Curves: (A) F1-Score Vs. Confidence; (B) Precision Vs. Recall; (C) Precision Vs. Confidence; (D) Recall Vs. Confidence.

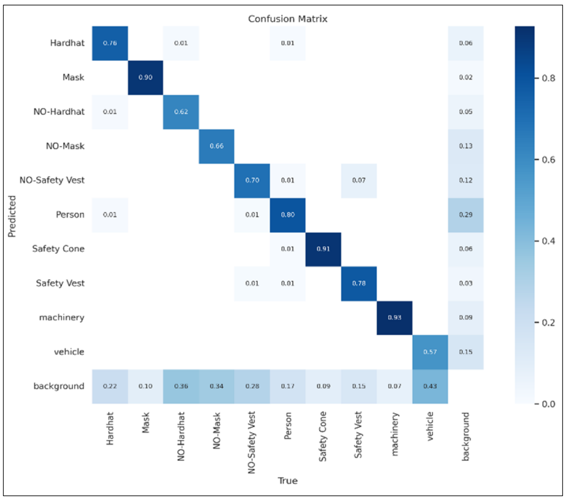

Figure 7: Confusion Matrix.

5.2 Testing and Validation

After training, the testing and validation part is where we evaluate the model's efficiency. Fig. 7 shows the confusion matrix generated by comparing the model's results with the image labels from the test set. The accuracy, precision, recall, and F1-Score were then computed for each class, as shown in Table 3.

We further validated the model by testing it on the following scenarios: (1) RGB, (2) gray-tone, (3) Blur effect, (4) Dust effect, (5) High brightness, and (6) Real-time Videos. Fig. 5 depicts the before and after for the first five conditions. For RGB and grayscale testing, we first evaluated the images in their original RGB format and then again in grayscale. For maximum brightness, dusting, and blur effects, the same set of images was processed under extreme brightness and with added noise to simulate real-world conditions such as strong sunlight or dusty environments. In the case of the blur effect, images were blurred up to 30% to mimic humid weather conditions.

Apart from image analysis, the model is also capable of processing videos by capturing and analyzing frames in real-time. It supports multiple video formats, including MP4, AVI, 3GP, WMV, MOV, FLV, MKV, WEBM, HTML5, AVCHD, MPEG-2, and MPEG-1. We can analyze CCTV footage for PPE detection and worker-machinery proximity monitoring in real time. The model can process videos ranging from 144p to 2160p without noticeable delays and supports frame rates of 30 FPS, 60 FPS, and beyond Alvarez et al., (2023).

5.3 Comparison of Detection Algorithms

In this study, we selected various sizes of YOLOv11 detection models for comparative experiments. Training the model using the same training set as the default YOLOv11 detection models rather than using transfer learning, because pre-trained models used different datasets for training. The primary comparison parameters include model size, processing speed, mAP values at 0.5 and 0.5-0.95 IoU, FLOPs, and the number of parameters. Table 4 presents the comparison results. We can see in Table 4 that our model performs better than other YOLOv11 models. Achieving a 3.7% increase in mAP at 0.5 IoU as well as 6.9% increase in mAP going from 0.5 to 0.95 IoU than YOLOv11l while maintaining size and speed between YOLOv11n and YOLOv11s. With a higher mAP and lower size and speed, it is computationally efficient compared to other models, making the model applicable for real-time analysis Azatbekuly et al., (2024).

5.4 Research Novelty

As mentioned earlier, multiple researchers have tried to automate construction site safety protocols using AI. Most studies focus on ML and CV techniques for safety helmet detection. Some research extends to detecting masks and vehicles alongside safety helmets using computer vision. However, none of these studies have performed PPE detection along with worker-machinery proximity monitoring. The proximity detection algorithm is a crucial component that, when integrated with PPE detection, could enhance workplace safety and prevent accidents.

The research achieved an overall accuracy of 97% and an overall precision of 87% in detecting PPEs. The results achieved are better than those of the previous studies and show slight improvements over existing models. Furthermore, the model is trained to detect objects under various weather conditions, including sunny, humid, and dusty environments. In contrast, most previous studies used only RGB images for training and testing. Our model, however, accurately detects all necessary PPEs, machinery, vehicles, and worker-machinery proximity with high precision across different weather conditions Sridhar et al., (2024).

6 CONCLUSIONS

This paper presents a fast and accurate model for detecting PPEs worn by workers and issuing alerts in case of non-compliance. The model detects objects in images with an inference speed of 2.4 seconds and outperforms YOLOv11l with a 3.7% increase in mAP at 0.5 IoU as well as 6.9% increase in mAP going from 0.5 to 0.95 IoU. To improve detection accuracy, mosaic data augmentation was included during training, allowing the model to detect small-scale objects effectively. Furthermore, the model can also track worker movements near machinery using bounding box regression and issue alerts if unsafe proximity is detected. We also trained the model to function under different weather conditions. The model achieves a mAP of 81% at 0.5 IoU and a mAP of 60.3% at 0.5 to 0.95 IoU, with an overall accuracy, precision, and recall of 97%, 87%, and 76%, respectively. It is computationally efficient, with 14.7GB FLOPs, 6.4MB parameters, and an inference speed of 2.4ms, making the model applicable for real-time analysis.

REFERENCES

Al-Azani, S., Luqman, H., Alfarraj, M., Sidig, A. A. I., Khan, A. H., Al-Hamed, D.: Real-Time Monitoring of Personal Protective Equipment Compliance in Surveillance Cameras. IEEE Access, 12, 121882–121895 (2024). https://doi.org/10.1109/ACCESS.2024.3451117

Alvarez, M. R., Vega, C. Q., Wong, L.: Model for Recognition of Personal Protective Equipment in Construction Applying YOLO-v5 and YOLO-v7. In: 2023 International Conference on Electrical, Computer and Energy Technologies - ICECET, pp. 1–6. IEEE, Cape Town (2023). https://doi.org/10.1109/ ICECET 58911.2023.10389215

Azatbekuly, N., Mukhanbet, A., Bekele, S. D.: Development of an Intelligent Video Surveillance System Based on YOLO Algorithm. In: 2024 IEEE 4th International Conference on Smart Information Systems and Technologies - SIST, pp. 498–503. IEEE, Astana (2024). https://doi.org/10.1109/ .2024. 10629617

Biswas, M., Hoque, R.: Construction Site Risk Reduction via YOLOv8: Detection of PPE, Masks, and Heavy Vehicles. In: 2024 IEEE International Conference on Computing, Applications and Systems - COMPAS, pp. 1–6. IEEE, Cox's Bazar (2024). https://doi.org/10.1109/COMPAS60761.2024.107962 53

Chen, B., Wang, X., Huang, G., Li, G.: Detection of Violations in Construction Site Based on YOLO

Algorithm. In: 2021 2nd International Conference on Artificial Intelligence and Computer Engineering - ICAICE, pp. 251–255. IEEE, Hangzhou (2021). https://doi.org/10.1109/ICAICE54393.2021.00057

Dermatology 2.0: Deploying YOLOv11 for Accurate and Accessible Skin Disease Detection: A Web-Based Approach. Scientific Figure on ResearchGate (2024). https://www.researchgate.net/figure/Shows-the-architecture-of-YoloV11_fig2_389021414

Feng, R., Miao, Y., Zheng, J.: A YOLO-Based Intelligent Detection Algorithm for Risk Assessment of Construction Sites. J. Intell. Constr., 2(4), 1–18 (2024). https://doi.org/10.26599/Jic.2024.9180037

Gautam, V., Maheshwari, H., Tiwari, R. G., Agarwal, A. K., Trivedi, N. K., Garg, N.: Feature Fusion-Based Deep Learning Model to Ensure Worker Safety at Construction Sites. In: 2024 1st International Conference on Advanced Computing and Emerging Technologies - ACET, pp. 1–6. IEEE, Ghaziabad (2024).https://doi.org/10.1109/ACET61898.2024.1073 0698

Han, C.; Zhang, J.; Wu, H.: Fall Detection System Based on YOLO Algorithm and MobileNetV2 Model. In: 10th International Conference on Systems and Informatics (ICSAI), pp. 1–5. IEEE, Shanghai (2024).https://doi.org/10.1109/ICSAI65059.2024.1089 3853

Han, K., Zeng, X.: Deep Learning-Based Workers Safety Helmet Wearing Detection on Construction Sites Using Multi-Scale Features. IEEE Access, 10, 718–729 (2022). https://doi.org/10.1109/ACCESS.2021.3138407

Jankovic, P., Protić, M., Jovanovic, L., Bacanin, N., Zivkovic, M., Kaljevic, J.: YOLOv8 Utilization in Occupational Health and Safety. In: 2024 Zooming Innovation in Consumer Technologies Conference - ZINC, pp. 182–187. IEEE, Novi Sad (2024).https://doi.org/10.1109/ZINC61849.2024.1057 9310

Krishna, N. M., Reddy, R. Y., Reddy, M. S. C., Madhav, K. P., Sudham, G.: Object Detection and Tracking Using YOLO. In: 2021 Third International Conference on Inventive Research in Computing Applications - ICIRCA, pp. 1–7. IEEE, Coimbatore (2021).https://doi.org/10.1109/ICIRCA51532.2021.95 44598

Li, Z., Guan, S.: Efficient-YOLO: A Research on Lightweight Safety Equipment Detection Based on Improved YOLOv8. In: 2024 2nd International Conference on Artificial Intelligence and Automation Control - AIAC, pp. 246–249. IEEE, Guangzhou (2024).https://doi.org/10.1109/AIAC63745.2024.1089 9732

Lin, B.: Safety Helmet Detection Based on Improved YOLOv8. IEEE Access, 12, 28260–28272 (2024). https://doi.org/10.1109/ACCESS.2024.3368161

M, L., J, R., M, A.: Real-time Hazard Detection System for Construction Safety Using Hybrid YOLO-ViTs. In: 2024 International Conference on Emerging Research in Computational Science - ICERCS, pp. 1–5. IEEE, Coimbatore (2024). https://doi.org/10.1109/ ICERCS 63125.2024.10895680

Mahmud, S. S., Islam, M. A., Ritu, K. J., Hasan, M., Kobayashi, Y., Mohibullah, M.: Safety Helmet Detection of Workers in Construction Site Using YOLOv8. In: 2023 26th International Conference on Computer and Information Technology - ICCIT, pp. 1–6. IEEE, Cox's Bazar (2023). https://doi.org/10.1109/ICCIT60459.2023.10441212

Menon, S. M., George, A., N, A., James, J.: Custom Face Recognition Using YOLO.V3. In: 2021 3rd International Conference on Signal Processing and Communication - ICSPC, pp. 454–458. IEEE, Coimbatore (2021). https://doi.org/10.1109/ICSPC51351.2021.9451684

N, S., Sridhar, S., Sudhir, S., Adithan, V. M., Vignesh, G.: Safefaceyolo: Advanced Workplace Security Through Helmet Detection and Facial Authorization. In: 2024 2nd International Conference on Artificial Intelligence and Machine Learning Applications - AIMLA, pp. 1–5. IEEE, Namakkal (2024). https://doi.org/ 10.1109/ AIMLA59606.2024.10531484

Ponika, M., Jahnavi, K., Sridhar, P. S. V. S., Veena, K.: Developing a YOLO-Based Object Detection Application Using OpenCV. In: 2023 7th International Conference on Computing Methodologies and Communication - ICCMC, pp. 662–668. IEEE, Erode (2023). https://doi.org/10.1109/ ICCMC56507. 2023.10084075

Roboflow Universe Projects: Construction Site Safety Dataset. Roboflow Universe (2024). Accessed on: Feb. 26, 2025. https://universe.roboflow.com/roboflow-universe-projects/construction-site-safety

Shanti, M. Z., Cho, C.-S., Byon, Y.-J., Yeun, C. Y., Kim, T.-Y., Kim, S.-K.: A Novel Implementation of an AI-Based Smart Construction Safety Inspection Protocol in the UAE. IEEE Access, 9, 166603–166616 (2021). https://doi.org/10.1109/ACCESS.2021.3135662

Shetty, N. P., Himakar, J., Gnanchandan, P., Prajwal, V., Jamadagni, S. S.: Enhancing Construction Site Safety: A Tripartite Analysis of Safety Violations. In: 2024 3rd International Conference for Innovation in Technology - INOCON, pp. 1–6. IEEE, Bangalore (2024).https://doi.org/10.1109/INOCON60754.2024.1 0511598