Advanced AI-Driven Safety Compliance Monitoring in Dynamic Construction Environment

Aisha Hassan[1], Ali H. Hassan, Ph.D[2], Yasmin Christensen[3], Hussain Alsadiq Ph.D[4]
Undergraduate Student, Department of Architectural Engineering, University of Prince Mugrin, Medinah, Alaqool, Saudi Arabia[1]
Assistant Professor, Department of Computer Science, University of Prince Mugrin, Medinah, Alaqool, Saudi Arabia[2]
Independent Researcher, Medinah, Saudi Arabia[3]
Assistant Professor, Department of Mechanical Engineering, University of Prince Mugrin, Medinah, Alaqool, Saudi Arabia[4]

***Abstract*—Construction safety is a critical global concern due to the high-risk environment faced by workers, with accidents often leading to serious injuries and fatalities. To enhance construction management, this study proposes a scalable deep-learning model for real-time compliance monitoring of safety regulations. The research gap addressed is the lack of real-time, scalable AI solutions for safety compliance monitoring in dynamic construction environments. The YOLOv11n model was trained and evaluated to identify and track the use of safety helmets and vests in extreme dynamic environments, ensuring timely detection of non-compliance. It is hypothesized that the YOLOv11n model will outperform baseline models in accuracy and real-time monitoring speed. The YOLOv11n model outperformed other baseline models, with precision, recall, and mean average precision scores of 89.5%, 85%, and 91.6%, respectively, and a real-time processing speed of 71.68 FPS. Its lightweight size and performance make it suitable for deployment. Integrated with a person-detection framework, the system provides real-time desktop alerts for safety violations, enhancing safety compliance. These findings contribute to construction automation by advancing scalable AI-driven solutions for proactive safety compliance, reducing accidents, and improving operational efficiency on construction sites.**

Keywords—YOLOv11n; personal protection equipment (PPE); construction safety; real-time object detection; deep learning; AI-driving compliance systems

I. Introduction

The construction industry is one of the most hazardous sectors globally, accounting for approximately 20% of workplace fatalities annually, according to [1]. Among these incidents, 38.4% result from trips, slips, and falls, highlighting the critical need for improved safety measures. Other more critical cases include being struck by objects and electrocutions, often due to a lack of compliance with personal protective equipment (PPE) protocols [2]**.** To address these risks, measures such as personal protective equipment (PPE) -including helmets, vests, gloves, and boots- are implemented to ensure workers' safety and reduce casualties. However, despite widespread adoption of safety measures, challenges remain in monitoring workers' compliance with PPE on construction sites due to several factors. First, the dynamic and complex conditions encountered in real-world construction sites with constantly changing weather and conditions can affect visual monitoring systems [3]. Second, manual monitoring is costly and difficult to scale for large-scale construction projects [4]. Third is environmental factors, as extreme dust, weather environments, and fluctuating lighting conditions may hinder accurate PPE detection. Moreover, studies [5] and [6] have highlighted that standard object detection models lack an accurate ability to detect small or overlapping objects and cluttered scenes, all of which are common on construction sites.

Recent advancements in artificial intelligence (AI) and deep learning have great potential for improving safety monitoring in construction sites. Similarly, in high-stakes domains like healthcare, the need for explainable AI models has been highlighted to encourage trust and aid in decision making [7]. Reinforcing the advances of applying AI models to dynamic construction settings. These technologies enable real-time data processing, object categorization, and predictive decision-making, allowing for the proactive identification of safety violations and potential risks [8]. It has been demonstrated that methods that combine deep learning with preprocessing techniques can improve recognition accuracy in areas like emotion recognition [9], [10]. Among these technologies are the You Only Look Once (YOLO) family of models, which are recognized for being fast and efficient deep learning algorithms for real-time object detection, identifying and classifying objects in a single step by predicting bounding boxes and class probabilities across an image [11].

Several studies have utilized YOLO in real-time detection of personal protection equipment (PPE) such as helmets, vests, and other safety gear, on construction sites. Studies by [6], [12], [13] emphasize the use of advanced YOLO versions, including YOLOv5, YOLOv8, and modified iterations, for real-time detection of PPE violations. The latter works highlight the models' strength in accurately identifying safety helmets, vests, and other protective equipment. In [6], the researchers focused on examining the feasibility of the addition of computer vision technology in construction sites through consistent training processes of YOLO models, to improve overall prediction accuracy. Likewise, the study conducted by [13] proposes an advanced PPE detection system that ensures safety compliance and enhanced productivity in high-risk environments.

In [12], the authors additionally suggests an improved YOLOv5 algorithm based on a Reverse-Unet framework (RUYOLO) to enhance small object detection and accuracy performance. Similarly, studies [14] and [15] have a shared emphasis on integrating advanced technologies (e.g. UAVs (Unmanned Aerial Vehicles) and IoTs (Internet of Things)) with deep learning algorithms to enhance safety monitoring devices. Both studies applied object detection models, such as YOLOv7 and MR-Net (which is a combination of MobileNet and Deep Residual Network components), to identify safety equipment gears. Bian et al. focus on the detection of safety helmets in specifically challenging areas such as construction power line energy fields. They proposed a YOLOv7 model integrated with an extended efficient layer aggregation network (E-ELAN), achieving high results in detection accuracy.

Despite significant advancements in object detection technologies, research on the latest YOLOv11 model for PPE detection remains limited. Most of the existing studies focus primarily on static environments and curated datasets, overlooking the dynamic and complex conditions encountered on real-world construction sites, such as low visibility and cluttered backgrounds. Furthermore, few studies compare YOLO versions under consistent training conditions, making model growth and trade-offs difficult to measure.

This study aims to bridge the gap by proposing and evaluating an optimized YOLOv11n, specifically focusing on helmet detection and PPE compliance to safety standards in construction work sites in dynamic construction settings. The model is trained using a domain-specific dataset augmented to improve detection accuracy under dynamic conditions, such as extreme weather conditions and background clutter. The performance is benchmarked against previous YOLO models, YOLOv3, YOLOv5, and YOLOv8, highlighting significant improvements in detection accuracy, inference speed, and deployment efficiency. Overall, this research aims to demonstrate how AI can be used to minimize safety violations and improve safety compliance in construction sites, with the specific use of improved deep learning models.

The key contributions of this study are as follows:

- Present a comprehensive analysis of multiple YOLO models to examine their potential in detecting PPE compliance under dynamic construction site conditions.
- Implementation of an optimized YOLOv11n model to achieve high object detection accuracy and real-time processing efficiency while reducing computational load and model size, making it suitable for real-time deployment.
- Integrated dual-model framework by merging two YOLO models to detect PPE compliance in construction sites, with real-time notification code that triggers desktop alerts when failure to comply with PPE standards is detected, demonstrating its practical use in live monitoring scenarios.

The rest of the study is structured as follows:

Section II explores previous studies of safety monitoring models. Section III details the training and optimization processes of the YOLOv11n model, including the data collection, preprocessing, model training, system integration, and performance evaluation. Section IV presents the results of the study. While Section V includes a discussion of the results highlighting the overall efficiency in real-time object detection. Section VI concludes the study, outlines the challenges of the study and suggests future research.

II. Literature Review

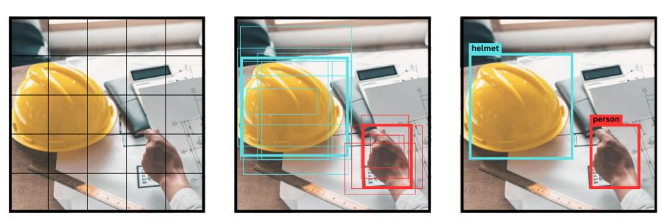
With the ever-growing construction industry, safety in construction is a great challenge faced worldwide [2]. For this reason, object detection for PPE equipment has become a priority to prevent accidents caused by safety violations [4]. Several studies have used deep learning models to combat this challenge, with the most popular choice being YOLO-based models, because of their ability for real-time detection and high effectiveness. A study made in the United Arab Emirates (UAE) developed a Convolutional Neural Network (CNN) model by using a model algorithm based on YOLOv3 for safety monitoring [16]. Their goal was to show that AI can be used to minimize accidents on construction sites by detecting workers' safety helmets, harnesses and lifelines. They were able to achieve a 94% accuracy rate in object detection, indicating the high success rate of the YOLO models.

This literature review investigates the effectiveness of deep learning models, particularly YOLO, in monitoring PPE compliance on construction sites. It compares different YOLO versions to identify which model offers the optimal safety compliance capabilities. The main research questions guiding this review are: How can AI deep learning models be used to enhance PPE compliance monitoring in construction settings, and what are the strengths and limitations of each of the YOLO models? Additionally, this review focuses on the gap by comparing these models' performances in detecting PPE violations.

A. *YOLO Based Detection for Safety Compliance*

In recent years, deep learning models, specifically YOLO models, have emerged as a popular choice for object detection because of their high effectiveness in the real-time detection of PPE equipment regarding speed and accuracy. Because of this, multiple studies have used this model, showing significant improvement in identifying PPE gear.

YOLO operates by dividing images into an S × S grid, where each grid cell detects objects within its boundaries as shown in Fig. 1. For each cell, the model predicts bounding boxes to determine the object's location and a confidence score indicating the likelihood of the object's presence [16]. These predictions include the object's coordinates, such as center, width, and height and its class probabilities, allowing YOLO to identify and localize multiple objects simultaneously [17]. This information is processed in a single forward pass which makes it efficient for real-time detection. Non-Maximum Suppression (NMS) is then applied to eliminate the overlapping prediction to ensure that most accurate bounding box is retained for each object.



S × S grid (image input) | Predictions + Confidence | Output

Fig. 1. YOLO object detection workflow.

Since its first release in 2016, the YOLO models have undergone significant enhancements and improvements. YOLO was established as a single-stage detector, identifying and classifying objects through a single network. They are smaller in size and faster, which makes them faster to train and deploy [17]. With the development of YOLO versions such as YOLOv3, YOLOv5, and YOLOv8, the models are enhanced, and performance is faster and more accurate. The latest release of the YOLO model, YOLOv11, in September 2024, introduces advanced features further optimizing the model for object detection accuracy and precision.

B. Related Works on PPE Detection

The studies examined highlight the significance of YOLO-based object detection models in enhancing real-time PPE compliance monitoring systems. Notably, versions such as YOLOv5 and YOLOv8 have demonstrated high accuracy and speed, even in complex environments. As seen in a study conducted by [6], 4844 images were gathered by the authors and trained on both YOLO versions, YOLOv5 and YOLOv8, with the same dataset. The results showed that the validation accuracies achieved by both models were 94.7% and 95.4%, respectively, making the YOLOv8 model the superior of the two. Similarly, in a study by [18], the YOLOv8 was optimized to improve the overall accuracy and effectiveness in small object detection. The model used mosaic data augmentation, including scaling images and cropping, which improved its object detection performance.

The YOLOv3 and RCNN models to identify the highest performance in detecting safety helmets in a construction setting were compared in [19]. An overall of 5000 images were gathered and then processed using two techniques, resizing and data augmentation. The datasets were then split into a ratio of 8:2 based on training and testing, respectively. The study utilized mAP, precision and recall, comparing the two models. After training, the YOLOv3 and RCNN models achieved 97.12% and 96.05% mAPs, respectively. Subsequently, the YOLOv3 model acquired a precision of 96.53% and a recall of 97.87%, while the RCNN model showed a precision of 94.1% and a recall of 96.75%. Despite challenges the models faced, such as detecting small objects and image lighting effects, the YOLOv3 model outperformed the RCNN with greater precision in recognizing safety helmets. Researchers in [20] focused on detecting safety gear using the YOLOv3 model. In their study, they stressed the significance of not only focusing on safety helmets but also on other safety gear, such as safety vests. 300 images were gathered and divided into two classes: compliant and non-compliant workers. 80% of the image data was set for training, while the remaining was used for data verification. The data was then analyzed and labelled by using LabelImg, which is used to annotate images with a bounding box. The YOLOv3 model was then trained using the data sets, achieving 92.99% mAP during the second iteration of the study. The model was also tested on live capture and video data, likewise, presenting high accuracy results. The authors concluded that even with a limited amount of data, the model received high precision in detecting safety gear. They also suggest further development in detecting other safety equipment through additional recent versions, expanding datasets and training models.

The use of a Mobile Residual Network (MR-Net) for object detection in safety monitoring was explored in [15]. They proposed three consecutive systems for enhanced object detection, the first being a sensor unit connected to an IoT network. A cloud server was then deployed to issue a smart alert, followed by the processing of gathered image data for detection using an NLM filter. YOLOv3 was then used to detect objects and categorize them based on four core classes (person, machines, vehicle, and safety cones) and further classify objects using MR-Net. The study then compared previous models based on accuracy, precision, recognition, positive outcomes, and negative outcomes. The results achieved optimal values of 91.2%, 90.5%, 92.3%, 90.9%, and 91.9%, respectively.

Detecting safety helmets in construction settings is difficult due to two main challenges: the small size of helmets at a further distance and the sheer size of the background image in contrast to the helmet's proportion. For this reason, authors in [3] analyzed video data to improve helmet detection and prevent false detections from occurring. The test datasets were collected from construction sites and utilized with multi-scale feature detection and training to enhance the YOLOv3 model. The original YOLOv3 model was then compared with the new optimized model, resulting in 93.5% mAP for the optimized model and 91.3% mAP for the YOLOv3 model. The enhanced model shows greater accuracy in detecting safety helmet violations than the original YOLO model by 2.2%.

To address the challenges in detection models resulting from small object detection, high rate of overlapping, and frequent occlusions, the study in [21] introduced an improved safety helmet detection model based on YOLOv5, integrating EIOU-loss to enhance model improvement. 4,400 images were gathered, ranging from different environments and distances, then labelled into three categories: helmet, head, and uncertainty. The improved models, retitled as YOLO-ESC and YOLO-ESCA, were compared with YOLOv3 and standard versions of YOLOv5. While the previous YOLO versions achieved greater precision and recall than the improved models, the model sizes of the YOLO-ESC and YOLO-ESCA are smaller, making them better candidates for deployment. Despite the improved enhancements, the study recognizes the limitations of the model in small object detection, making it a goal for further research development.

The performance of advanced YOLOv5 model variants (n, s, m, l, x) for safety helmet detection has been examined in [22]. The study conducted a comparative analysis based on key performance indicators, including mean Average Precision (mAP), precision, and recall, across two target classes: “head” and “helmet”. The findings indicated that YOLOv5n demonstrated superior speed, achieving 62.5 FPS for image data and 70.4 FPS for video data. However, in terms of detection accuracy, the YOLOvX variant achieved the highest mAP score of 95.8%, making it the most precise among the evaluated models. Similarly, in [5], the authors proposed an improved lightweight YOLOv5s model to enhance safety helmet detection through image augmentation techniques, aiming to optimize model robustness in dynamic environments. A comparative analysis was conducted between the improved YOLOv5s model and other advanced models such as YOLOv3, YOLOv4, YOLOv7 and YOLOv8. The improved model attained a mAP of 94.1%, exceeding that of the YOLOv7-Tiny and similar to that of the YOLOv8s model. The results indicate a notable improvement in the model while maintaining its lightweight characteristics. The study concludes with future research recommendations for further exploring systems, to diminish parametric defects and improve accurate detection of objects.

The lack of focus on other PPE elements such as safety vests, gloves, boots and glasses has been highlighted in previous research. To address this, a safety monitoring model was proposed that compares YOLOv5 and YOLOv7 [23]. A dataset of a total of 743 images was collected and categorized into 6 classes: boots, glasses, gloves, helmets, person, and safety vest. The data was then annotated using LabelImg and directed for training using both models individually. The metrics used for evaluation are mAP@0.5 of precision and recall. The final values of the models showed that the superior system in precision and recovery of image detection was YOLOv7, with a value of 87.3%. The value of the YOLOv5 was marginally lower at 79.6%, indicating a slight insufficiency of the process metric.

Due to the shortcomings of traditional safety compliance monitoring methods in construction sites, the need for integrating object detection systems has become increasingly important. Recent research has demonstrated that improved object detection algorithms, specifically YOLOv8, are efficient in detecting PPE items [24]. In this study, a total of 3,750 images were annotated and split into five classes: Vests, Safety shoes, No vests, No helmets, and No safety shoes. The dataset was then trained with YOLOv8 using a maximum of 100 epochs, where performance is evaluated periodically. The model achieved significant results in accuracy, scoring 100% and 98% in the detection of helmets and vests, respectively. The model faced a slight regression in detecting safety shoes with a value of 84%, despite this, the model's performance and effectiveness in PPE detection were prominent.

The limited focus on computer vision software and work safety and wellbeing has contributed to the study in [25], which explores two model frameworks based on YOLOv8 pretrained nano and small architectures. These models were trained for hard hat and head detection using datasets provided by Roboflow. The YOLOv8 nano and small models were evaluated based on precision, recall, mAP@50, and mAP@50-95. The YOLOv8 small model outperformed the YOLOv8 nano model, with a high precision value of 92%. The study discusses the restraints of exploring more enhanced models due to the computational constraints limiting the number of test models, suggesting further research on YOLOv8 structures with longer test phases.

C. Synthesis of Current Studies

The studies reviewed, demonstrate significant potential in enhancing PPE compliance detection using YOLO deep learning models. YOLOv5 and YOLOv8, in particular, presented substantial results in safety helmet and vest detection, reaching an accuracy value of 95% [6]. Nevertheless, the application of these models in real-world construction environments continues to face significant challenges, as highlighted by recent studies. These challenges include ensuring environmental robustness under various conditions, limitations in multi-PPE detection, and issues related to computational efficiency.

1) Environmental robustness. A major challenge in current detection models is their ability to provide accurate and comprehensive detection under various environmental conditions. As emphasized by [26], we need scalable energy-efficient solutions to balance model performance and resource constraints in real-world environments. The advancement of the YOLO model for safety equipment detection was explored in [3] and [14]. Their studies reported a significant drop when these models were exposed to complex site conditions, such as dust, low visibility and cluttered backgrounds- factors that are common in construction work sites. Despite the success in object detection in static backgrounds, the YOLO models still faced limitations in small object detection and compliance monitoring in dynamic environments- as current models often utilize clean, high-quality data, which limits their ability to comply with the ever-changing work environment.

2) Multi-PPE detection. Additionally, most existing research focuses primarily on helmet detection, overlooking other PPE items such as safety vests, gloves and shoes, which are equally essential for the safety of construction workers (Table I). According to studies by [23] and [20], most studies have focused on helmet detection while little attention has been on other critical safety gear such as gloves, safety vests, masks and goggles. The authors have further highlighted the importance of focusing on various safety gear to ensure safety compliance and minimize workplace accidents. Future research should focus on improving YOLO frameworks to further enhance small object detection and recognition of multiple types of PPE in dynamic environments. Additionally, further studies should evaluate these models using real-time data from live construction sites to assess their practical effectiveness.

3) Computational efficiency. The YOLO family of object detection models has evolved significantly over time, with each version showing improved capabilities in terms of speed, accuracy and computational efficiency as shown in Table II. Although the earlier models, such as YOLOv3 and YOLOv5s, demonstrated notable performance, they often required higher computational resources as they have relatively large model sizes, which are 117 MB and 7.06 MB, respectively and higher FLOPs of 39.2G and 7.70G, respectively. YOLOv7 improved upon previous versions, attaining a speed of 14.1ms despite its medium model size of 80.91 MB, indicating its efficiency in fast optimization. The YOLOv8n achieved a significantly smaller model size of 3.2 MB, reducing latency to 12.4 ms, resulting in an impressive FPS of 80.4.

While these high-preforming variants provide advanced accuracy, their practical deployment is limited by their computational demands. As [27] and [28] note, larger models require high computational resources such as large memory resources, which limit their scalability on real-life work sites. This challenge highlights the balance between deployability and model complexity in resource-constrained environments.

TABLE I. QUANTITATIVE COMPARISON OF OBJECT DETECTION MODELS FOR PPE COMPLIANCE MONITORING

Reference	Algorithm	Environment	Datasets	Class	Precision%	Recall%	mAP50%
[19]	YOLOv3	Static	5000	Helmet	96.53%	97.87%	97.12%
[21]	YOLO-ESC	Dynamic	44,000	All	96.5%	98.2%	98.7%
				Helmet	98.1%	99.6%	99.1%
				Head	97.5%	99.1%	99.1%
				Uncertainty	93.9%	95.8%	97.8
[22]	YOLOv5n	Static	7063	Helmet	89.3%	93.5%	94.2%
				Head	85.9%	93.5%	
[5]	Imp.YOLOv5s	Dynamic	36,918	Helmet	92.9%	91.6%	94.1%
[23]	YOLOv7	Static	743	All	95.7%	87.3%	92%
[25]	YOLOv8n	Dynamic	7035	Helmet	89.5%	44.9%	51.7%
				Head			
[24]	YOLOv8m	Static	2934	All	89%	92%	95.6%

TABLE II. PERFORMANCE COMPARISON BETWEEN PREVIOUS MODELS AND THE CURRENT IMPROVED YOLOV11N MODEL

Algorithm	Study	Model Size	Parameters (M)	Latency (ms)	FLOPs (G)	Frames per Second (FPS)
YOLOv3	[21]	medium	117	39.2	65.0	25.5
YOLOv5s	[27]	small	7.06	18	7.70	55.56
YOLOv7	[14]	medium	80.91	14.1	101.8	70.92
YOLOv8n	[29]	nano	3.2	12.4	8.7	80.4

III. METHODOLOGY

This research develops and evaluates the YOLOv11n model for PPE compliance monitoring in construction sites. High-quality datasets of 19,076 images were collected from various sources, annotated for safety helmets and vests, and preprocessed with resizing and data augmentation to improve detection accuracy under dynamic conditions such as extreme weather conditions and fluctuating lighting. The YOLOv11n model, selected for its computational efficiency, was trained on a T4 GPU using Google Colab, with key metrics such as precision, recall, and mAP used for performance evaluation. The model was then tested on prerecorded and live video feeds to validate its effectiveness in real-time safety monitoring.

A. Image Collection

High-quality images and videos of construction sites and PPE items were collected from multiple sources to ensure diverse and balanced datasets. Publicly available resources from sites like Unsplash and Pexels were gathered, increasing safety gear and construction site image variety. Additionally, two distinct datasets were forked from Roboflow [30], [31]. Each dataset specifically focused on safety helmets and safety vests, respectively, further enhancing the model with a wide range of PPE items and variations. The addition of these datasets was essential to develop a well-trained model with high accuracy in object detection. In total, 19,076 images were utilized for dataset preparation, providing an exemplary dataset for effective model training. The images collected represent a diverse collection of PPE imagery, considering factors such as lighting, dynamic environments, and movement, which ensured adaptation to real-time conditions.

B. Dataset Preparation

The datasets were merged, creating a single diverse PPE dataset, allowing for enhanced model training. The dataset was then annotated using Roboflow by implementing bounding boxes on PPE items based on two classes (Safety Helmet and Safety Vest) as shown in Fig. 2. This annotation process ensures precise labelling required for high-accuracy training. An additional dataset obtained from Roboflow was also labelled using a single class (Person) [32] to be integrated into a separate training for enhanced PPE compliance monitoring.

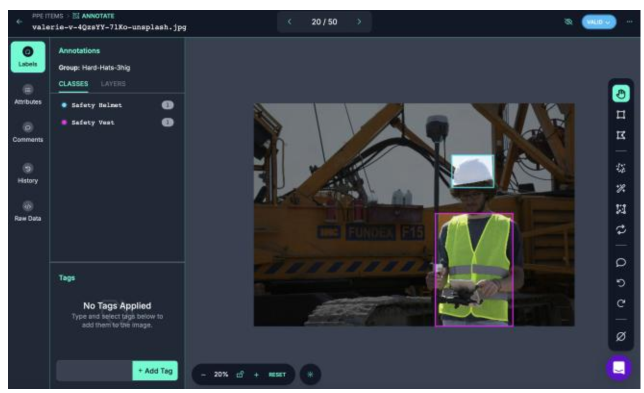


Fig. 2. Screenshot of Roboflow annotation interface based on two class labels: safety helmets and safety vests.

Each class was annotated with a count ratio of approximately 2:1 for safety helmets and safety vests, respectively. The ratio difference is due to the challenges AI models face in small object detection. Thus in order to enhance the detection of safety helmets in dynamic backgrounds, the annotation count was increased (Table III). For the PPE dataset, the images were split based on training, validation and testing. Each set was divided into 16,650, 1625, and 801 images, respectively, as shown in Table IV.

TABLE III. ANNOTATION COUNT FOR EACH CLASS

Class	Count
Safety Helmet	15,287
Safety Vest	6,590

TABLE IV. DATASET DISTRIBUTION FOR TRAINING, VALIDATION, AND TESTING

Training	Validation	Testing
16,650	1625	801

1) Preprocessing. To maintain a consistent processing size and improve model training, all images in the dataset were resized to a resolution of 640 × 640 pixels. This uniform size allows for standardized processing across the dataset and equally complies with the standard image size of YOLO models. Additionally, Auto-Orient was applied to correct the image orientation and ensure proper alignment for optimal object detection. This step ensures standardized preprocessing for consistent results across the dataset. This preprocessing stage is crucial for the success of data augmentation and model training processes while reducing computational overload.

2) Data augmentation. To further enhance model training, data augmentation is implemented in the dataset. Data augmentation is a critical process to enhance model recognition and generalization capabilities, as it utilizes the original dataset to generate a diverse dataset based on multiple variations. This process simulated dynamic conditions such as extreme weather and variable lighting, enhancing the model's robustness. The augmentation strategy, which included applying 25% grayscale alterations to images, assigning a blur with a 4.8-pixel radius, and addition of noise with a deviation of 1.8 pixels, were chosen for their ability to simulate real-world conditions (Fig. 3). For instance, the grayscale modifications imitate limited visibility due to stormy weather, while noise and blur alterations simulate dust, rain, or light fluctuations. By expanding the dataset diversity, this process plays a pivotal role in increasing model training accuracy.

Fig. 3. Example of data augmentation added to construction site image. Original image (Left) and augmented image (Right) with 25% Grayscale, blur 4.8px, and noise 1.8px.

C. Model Training

This research employs the YOLOv11n model obtained by Ultralytics and trains it for PPE detection using Google Colab. The YOLOv11n model is selected because of its lightweight architecture, optimized inference speed and low memory usage compared to the other YOLOv11 models (YOLOv11s, YOLOv11m, YOLOv11l, YOLOv11x), making it a significant choice where computational efficiency is a priority, such as construction sites [33]. The model is trained using a carefully curated dataset for PPE detection, ensuring high-speed training and performance. The training process was conducted on a T4 GPU via Google Colab, allowing for faster training time and model optimization. The coding environment was set up in both Google Colab and Spyder, Python 3.12. Table V further details the training environment.

TABLE V. MODEL TRAINING ENVIRONMENT

Component	Specifications
Hardware	GPU: T4 (Google Colab), CPU (MacOS CPU model)
Program	Google Colab, Anaconda (Spyder, Python 3.12)
Framework	PyTorch
Libraries	Ultralytics, OpenCV, NumPy, Roboflow, Matplotlib
Algorithm	YOLO11n

Hyperparameters are set in the training environment to ensure optimal accuracy and recall values in PPE detection (Table VI). The dataset used is defined as data.yaml, which contains data for classes and paths to images used for training. The model was trained for 20 epochs to establish sufficient exposure to the dataset and achieve optimal learning. A batch size of 16 and an image size of 640 pixels were selected for efficient model training and processing. The model weights were further optimized for real-time detection performance by implementing fine-tuning settings, including reducing precision to FP16, using a batch size of 8, and resizing input images to 320 × 320 pixels. These optimizations were crucial for enhanced speed and reduced computational load.

TABLE VI. MODEL TRAINING HYPERPARAMETERS

Hyperparameter	Value
Data	data.yaml
Epoch	20
Batch	16
Imgsz	640

1) Metric monitoring. After model training is complete, the model is evaluated using the following metrics:

Precision%: Percentage of accurately detected PPE as shown in Eq. (1):

$$Precision = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}} \quad (1)$$

Recall%: Percentage of actual PPE identified as computed according to Eq. (2):

$$Recall = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}} \quad (2)$$

Mean average precision (mAP50): The average precision values are calculated at an intersection over union (IoU) threshold of 50%.

Mean average precision (mAP50-95): The average precision values are calculated at an IoU threshold of 50% to 95%.

Confusion matrix: The model's ability to identify objects correctly. Table VII defines each confusion matrix.

TABLE VII. CONFUSION MATRIX DEFINITIONS

Metric	Definition
True Positive (TP)	The number of instances predicted correctly as a positive class
True Negative (TN)	The number of instances predicted incorrectly as a positive class
False Positive (FP)	The number of instances predicted incorrectly as a negative class
False Negative (FN)	The number of instances predicted correctly as a negative class

Class Loss: Indicates how the model performs in classifying each PPE item.

Latency (L): As shown in Eq. (3), it is the time required to generate a prediction during the inference process.

$$L = \frac{\text{Total Inference Processing Time (s)}}{\text{Number of Images Processed}} \tag{3}$$

Frames per Second *(FPS*): The number of frames processed per second is calculated as in Eq. (4):

$$FPS = \frac{\text{Batch Size}}{\text{Latency (ms)}} \times 1000 \tag{4}$$

Floating Point Operations (FLOPs): Indicates the computational cost and efficiency of the model as measured in Eq. (5):

$$FLOPs = 2 \cdot K^2 \cdot C_{\text{in}} \cdot C_{out} \cdot H_{\text{out}} \cdot W_{\text{out}} \tag{5}$$

K: Kernel size

C_{in}: Number of input channels

C_{out}: Number of output channels

H_{out}: Output height

W_{out}: Output width

Complex O Notation (O): The computational complexity during inference can be estimated as shown in Eq. (6) for a single layer and Eq. (7) for the total complexity:

$$Single\ \text{Layer} = O(K^2 \cdot C_{\text{in}} \cdot C_{\text{out}} \cdot H_{\text{out}} \cdot W_{\text{out}}) \tag{6}$$

$$Total\ Complexity = O(\sum_{i=1}^{L} K_i^2 \cdot C_{\text{in},i} \cdot C_{\text{out},i} \cdot H_{\text{out},i} \cdot W_{\text{out},i}) \tag{7}$$

Monitoring these metrics is an important process in developing an efficient model to avoid overfitting or underfitting, and ensuring that the model performs well with real-time data.

D. System Integration

The PPE-trained YOLOv11n model is integrated with a pretrained YOLO model improved for person detection. This integration allows for enhanced PPE compliance detection and system notification, making the model effective for real-time safety compliance monitoring on construction sites. The two models are merged using Anaconda (Spyder 3.12), including a code that executes a desktop alert for failure to comply with PPE requirements, further enhancing the model for optimal construction site safety.

E. Performance Evaluation

To further evaluate the performance of the enhanced model, it is tested on prerecorded video data of construction sites and live video feed using an IP Webcam application. The evaluation focuses on the accuracy of the bounding boxes in detecting PPE objects, noting the model's ability to detect safety helmets, safety vests, and safety violations. Key performance metrics such as precision, recall, and mAP are recorded, assessing the model's capabilities in dynamic environments and its effectiveness in real-time compliance monitoring on construction sites.

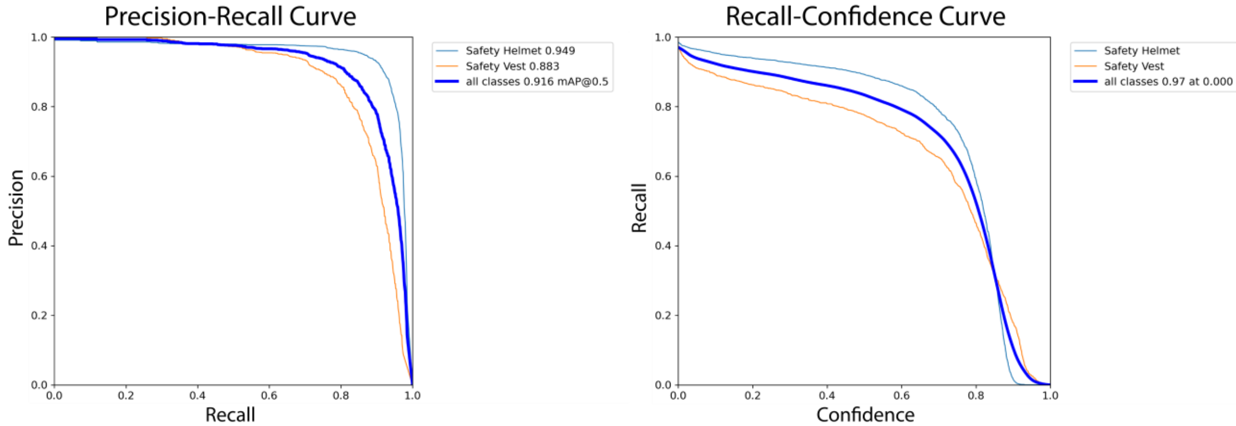
IV. RESULTS

The YOLOv11n model received a training progression over 20 epochs. At the 20th epoch, the model achieved a precision of 89.48%, 85.01% recall, 91.62% mAP50, and 64% mAP50-95, indicating significant performance outcomes (Table VIII).

TABLE VIII. PERFORMANCE METRICS OF YOLOV11N BASED ON PRECISION, RECALL, AND MAP50

Algorithm	Precision%	Recall%	mAP50%
YOLOv11n	89.5%	85%	91.6%

The Precision-Recall curve shown in Fig. 4(a) illustrates the performance of each class and overall accuracy. Safety helmets received the highest confidence at 94.9%, revealing the model's efficiency in helmet detection. Likewise, safety vests achieved an optimal confidence of 88.3% resulting in an overall mAP50 of 91.6%. The model maintains exceptional results at an optimal threshold, achieving a high value of 97% as shown in Fig. 4(b) in the Recall-Confidence curve. This indicates the model's strong ability to detect minimal false negatives. Notably, as the confidence threshold increases, the recall value might slightly decrease to prevent false positives from showing, signifying the model's ability to balance between recall and precision.



(a) (b)

Fig. 4. YOLOv11n : (a) Precision-Recall Curve; (b) Recall-Confidence Curve.

The confusion matrix, which represents the model's capability in identifying objects correctly based on three classes (Safety Helmet, Safety Vest, and Background), indicates remarkable results. As shown in Fig. 5, the TP rate for safety helmets reached an optimal 94%, revealing the model's strength in detection accuracy and background separation. The model experienced false negatives primarily in poor lighting conditions and during object overlaps, where smaller or obscured PPE items were not detected. Conversely, false positives were observed when background elements, such as construction equipment, resembled PPE items. These findings indicate that environmental factors significantly impact detection accuracy. Despite the model experiencing slight confusion in distinguishing safety vests from background images due to similar features or overlap in color, the class nevertheless obtained a notable TP value of 86%. These results signify the robustness of the model's performance and effectiveness in identifying target classes.

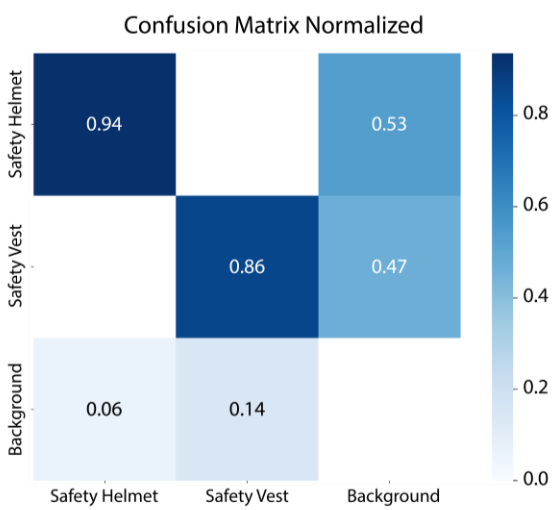


Fig. 5. Confusion matrix across three classes: safety helmets, safety vests, and background.

Instance distribution and bounding box visualization are equally analyzed, as shown in Fig. 6. The instance chart (Left) indicates a significant class count of safety helmets in contrast to safety vests, suggesting increased trained data contributing to higher accuracy in helmet detection. The bounding box visualization shown in Fig.6 (Right) illustrates the predefined shapes the model applies to determine object location and size. The anchor boxes are well distributed and vary in size, indicating the model's capability in detecting objects of various dimensions.

Following the optimization process, which included reducing computational load by implementing FP16 precision for faster inference times, resizing images to 320 pixels, and minimizing batch size from 16 to 8, the model is evaluated for latency, FPS and FLOPs using NVIDIA GPU. The model achieved a significant processing speed of 71.68 FPS, a paramount increase in contrast to the video playback rate of 25 FPS (Table IX). The average latency equally was 13.91ms indicating near real-time detection of objects. The computational complexity of the model is equally calculated achieving a value of 801.68 MMac and a compact model size of 2.59 MB, these values demonstrate the efficiency for deployment on edge devices with limited resources in dynamic construction environments.

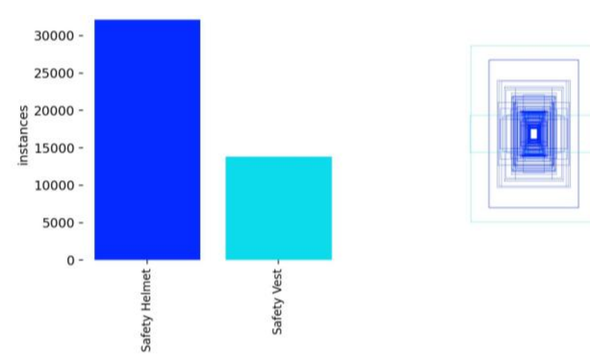


Fig. 6. Instance distribution and bounding box visualization analysis.

TABLE IX. PERFORMANCE METRICS OF OPTIMIZED YOLOV11N BASED ON FPS, LATENCY, MODEL SIZE, AND FLOPS

Algorithm	Video FPS	Processing FPS	Latency (ms)	Model Size (M)	FLOPs (G)
YOLO11n	25	71.68	13.951	2.59	0.801

The enhanced model is evaluated on the distributed dataset set for testing, prerecorded video data and live webcam feed, to further validate its effectiveness in real-time dynamic construction environments as shown in Fig. 7. The results displayed exceptional performance in both static and dynamic settings, with consistent detection accuracy despite fluctuations in lighting, worker movement and environmental conditions. These findings demonstrate the efficiency of the model's application in safety compliance monitoring on construction sites.

To ensure PPE compliance in construction sites, the improved YOLOv11n model is integrated with a pretrained YOLO model aimed specifically for person detection. The supplementary model is enhanced with additional training at 30 epochs receiving an overall precision value of 99.8% and recall of 100%. The newly merged dual model framework is additionally integrated with a desktop alert code to further implement PPE compliance on construction work sites and notify supervisors of non-compliance breaches. The model achieved optimal results displaying efficient performance in construction compliance monitoring. Fig. 8 shows a real-time PPE violation desktop alert sent on MacOS.

Fig. 7. Image evaluation: Training set dataset (left) and inference ran on prerecorded video data (right).

Fig. 8. Real-time detection of PPE compliance and safety violations.

V. DISCUSSION

The optimized Yolov11n model obtained higher accuracy in detecting PPE compliance, exceeding the precision value of both YOLOv8n and YOLOv5n in dynamic conditions [22], [25]. It demonstrated superior recall and precision, outperforming previous models and validating its suitability for real-time monitoring in construction sites. This aligns with research by [18] who emphasize the importance of real-time detection in complex environments. While safety helmets obtained a significant TP value of 94%, the model experienced a slight regression in the TP value for safety vests, which could be due to the model's struggle in differentiating the class from complex backgrounds with similar color overlaps and feature distinction. This lower accuracy has equally been noted in previous research studies, such as that by [24]. Future iterations of the model should incorporate advanced background segmentation techniques to address this challenge.

Additionally, the model demonstrated significant ability in detection of objects in real-time environments, as evidenced by its increased FPS and reduced latency, making it highly efficient for fast paced construction sites. The model outperformed the inference speed of the earlier YOLOv3 model by 78.5% and the YOLOv5s model by 22.5% [21], [27]. The YOLOv11n model equally attained a lighter model size and reduced computational cost, in contrast to the heavy weight and high cost of the YOLOv7 architecture, signifying the ability to perform on edge devices in dynamic environments without compromising speed and accuracy [14]. While the YOLOv8n model demonstrated a slightly higher inference speed of 80.4 FPS, the improved YOLOv11n model displays a significant balance with lower FLOPs and model size, making it a superior choice for deployment on complex construction sites where computational efficiency and real-time performance are critical [29].

Despite the model's strengths, the research faced several challenges and limitations in achieving optimal accuracy results. This could be due to an imbalance in class counts between safety helmets and safety vests, which could have led to biased detection outcomes. Likewise, while the nano architecture of the YOLOv11 model enhanced detection speed and training efficiency, a larger model architecture might have potentially increased performance significantly. This limitation arises from the limited computational load of the MacOS CPU, and the restricted time limit of Google Colab used during training. Addressing these challenges, by using a balanced dataset and increasing GPU for longer training time, could enhance the model further for PPE detection.

Furthermore, due to the conversion to FP16 and resizing of the images to 320 pixels during the inference process in order to improve latency and reduce computational cost, the model could potentially face a slightly lowered detection performance. Finetuning the model further, with a larger batch size during training and exploring enhanced precision techniques, including merging FP32 and FP16 where necessary, could enhance overall performance in real-time applications. Expanding the dataset to include additional PPE items such as gloves, glasses, and boots is recommended for future research endeavors to develop a more comprehensive safety object detection model. Increasing the number of epochs and fine-tuning hyperparameters, while ensuring class balance, could further enhance the model's performance and detection accuracy. Moreover, deploying the model on edge devices, such as sensors and IoTs, could additionally advance the model in object detection making it crucial for real-time applications. Implementing these steps would not only potentially enhance detection capabilities but equally allow for widespread implementation in challenging conditions and high-risk environments ensuring worker safety and construction requirements compliance.

Overall, the results of the YOLOv11n model demonstrate exceptional performance in real-time detection within dynamic environments, with further optimization suggestions paving way for a more comprehensive model and its deployment in real-time critical settings.

VI. Conclusion

This research proposes a deep learning model using an optimized YOLOv11n algorithm to address safety challenges on construction sites. Crucial performance metrics such as precision, mAP50, Latency, and FLOPS, are evaluated to ensure real-time PPE compliance monitoring in resource-constrained dynamic environments. The YOLOv11n demonstrated significant results, achieving 89.5%, 85%, and 91.6% for precision, recall, and mAP50, respectively while maintaining a notable balance between speed and computational load. With a latency of 13.95 ms and computational cost of 0.801 GFLOPs, the model exhibits a lightweight and computationally efficient architecture capable of object detection on edge devices. Additionally, the study integrated a PPE compliance system with desktop notifications to alert supervisors of safety violations, demonstrating the model's robustness and practical application for real-time monitoring.

The research, however, encountered several challenges. The class imbalance led to a potential bias in detection results, and the limited computational resources prevented the exploration of larger, more complex architectures. Additionally, the need for increased speed and reduced model size in complex environments may slightly impact detection performance, practically for small object detection. These challenges underscore the importance for further enhancements and evaluation.

Future research recommendations include improving detection performance by further optimizing the model with additional hyperparameters, extended training epochs and balanced datasets. Expanding the model's scope to include additional PPE items, such as safety glasses, boots, and gloves, would further enhance the model's applicability. Furthermore, exploring the integration of the model on edge devices could significantly advance its capabilities for real-time detection in complex construction sites. These efforts aim to address current limitations and pave the way for safer, more efficient construction sites through advanced, scalable deep learning models and real-time compliance monitoring systems.

References

[1] U.S. Bureau of Labor Statistics, "A look at falls, slips, and trips in the construction industry : The Economics Daily: U.S. Bureau of Labor Statistics." Accessed: Nov. 09, 2024. [Online]. Available: https://www.bls.gov/opub/ted/2024/a-look-at-falls-slips-and-trips-in-the-construction-industry.htm

[2] A. Mohammadi, M. Tavakolan, and Y. Khosravi, "Factors influencing safety performance on construction projects: A review," Saf Sci, vol. 109, pp. 382–397, Nov. 2018, doi: 10.1016/J.SSCI.2018.06.017.

[3] J. Hu, X. Gao, H. Wu, and S. Gao, "Detection of Workers Without the Helments in Videos Based on YOLO V3," in 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), IEEE, Oct. 2019, pp. 1–4. doi: 10.1109/CISP-BMEI48845.2019.8966045.

[4] S. Li and A. Zhu, "Recent Advancements with Human Behavior Recognition and AI in Construction Automation: A literature Review," in

ICARM 2024 - 2024 9th IEEE International Conference on Advanced Robotics and Mechatronics, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 759–764. doi: 10.1109/ICARM62033.2024.10715843.

[5] C. Du, X. Zhang, and C. Wu, "Method for Safety Helmet Detection Based on Improved YOLOv5," in Proceedings - 2023 International Conference on Computer, Internet of Things and Smart City, CIOTSC 2023, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 43–48. doi: 10.1109/CIoTSC60428.2023.00013.

[6] K. Kim, K. Kim, and S. Jeong, "Application of YOLO v5 and v8 for Recognition of Safety Risk Factors at Construction Sites," Sustainability, vol. 15, no. 20, p. 15179, Oct. 2023, doi: 10.3390/su152015179.

[7] A. H. Hassan, R. bin Sulaiman, M. Abdulhak, H. Kahtan, and U. Kebangsaan Malaysia, "Bridging Data and Clinical Insight: Explainable AI for ICU Mortality Risk Prediction Institute of IR 4.0 (IIR4.0)," 2025. [Online]. Available: www.ijacsa.thesai.org

[8] M. Dol and A. Geetha, "A Learning Transition from Machine Learning to Deep Learning: A Survey," in Proceedings of the 2021 International Conference on Emerging Techniques in Computational Intelligence, ICETCI 2021, Institute of Electrical and Electronics Engineers Inc., Aug. 2021, pp. 89–94. doi: 10.1109/ICETCI51973.2021.9574066.

[9] Z. ullah, L. Qi, D. Binu, B. R. Rajakumar, and B. Mohammed Ismail, "2-D canonical correlation analysis based image super-resolution scheme for facial emotion recognition," Multimed Tools Appl, vol. 81, no. 10, 2022, doi: 10.1007/s11042-022-11922-3.

[10] M. A. A. M. Divyashree, "Machine Learning Technique for Image Classification System using Assembly," in International Journal of Engineering Research & Technology (IJERT), 2020. [Online]. Available: www.ijert.org

[11] K. Alshammari, "Enhancing Vision-Based Religious Tourism Systems in Makkah Using Fine-Tuned YOLOv11 for Landmark Detection," 2025. [Online]. Available: www.ijacsa.thesai.org

[12] L. Sun and L. Wang, "An improved YOLO V5-based algorithm of safety helmet wearing detection," in Proceedings of the 34th Chinese Control and Decision Conference, CCDC 2022, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 2030–2035. doi: 10.1109/CCDC55256.2022.10033448.

[13] D. Thakur, P. Pal, A. Jadhav, N. Kable, V. Bhagyalakshmi, and S. Deshpande, "YOLOv8- Based Helmet and Vest Detection System for Safety Assessment," in 2023 International Conference on Network, Multimedia and Information Technology, NMITCON 2023, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/NMITCON58196.2023.10275958.

[14] H. Bian et al., "Detection Method of Helmet Wearing Based on UAV Images and Yolov7," in ITNEC 2023 - IEEE 6th Information Technology, Networking, Electronic and Automation Control Conference, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 1633–1640. doi: 10.1109/ITNEC56291.2023.10082536.

[15] L. R. Jadhav and M. Hudnurkar, "Mobilenet and Deep Residual Network for Object Detection and Classification of Objects in IoT Enabled Construction Safety Monitoring," in Proceedings of the 2023 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems, ICSES 2023, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ICSES60034.2023.10465566.

[16] M. Z. Shanti et al., "A Novel Implementation of an AI-Based Smart Construction Safety Inspection Protocol in the UAE," IEEE Access, vol. 9, pp. 166603–166616, 2021, doi: 10.1109/ACCESS.2021.3135662.

[17] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, Dec. 2016, pp. 779–788. doi: 10.1109/CVPR.2016.91.

[18] B. Lin, "Safety Helmet Detection Based on Improved YOLOv8," IEEE Access, vol. 12, pp. 28260–28272, 2024, doi: 10.1109/ACCESS.2024.3368161.

[19] N. K. Anushkannan, V. R. Kumbhar, S. K. Maddila, C. S. Kolli, B. Vidhya, and R. G. Vidhya, "YOLO Algorithm for Helmet Detection in Industries for Safety Purpose," in 3rd International Conference on Smart Electronics and Communication, ICOSEC 2022 - Proceedings, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 225–230. doi: 10.1109/ICOSEC54921.2022.9952154.

[20] R. C. Reyes, R. V. Sevilla, G. S. Zapanta, J. V. Merin, R. R. Maaliw, and A. Ferrer Santiago, "Safety Gear Compliance Detection Using Data Augmentation-Assisted Transfer Learning in Construction Work Environment," in 2022 IEEE International Conference on Electronics, Computing and Communication Technologies, CONECCT 2022, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/CONECCT55679.2022.9865757.

[21] P. Jin, H. Li, W. Yan, and J. Xu, "YOLO-ESCA: A High-Performance Safety Helmet Standard Wearing Behavior Detection Model Based on Improved YOLOv5," IEEE Access, vol. 12, pp. 23854–23868, 2024, doi: 10.1109/ACCESS.2024.3365530.

[22] Kisaezehra, M. U. Farooq, M. A. Bhutto, and A. K. Kazi, "Real-Time Safety Helmet Detection Using Yolov5 at Construction Sites," Intelligent Automation and Soft Computing, vol. 36, no. 1, pp. 911–927, 2023, doi: 10.32604/iasc.2023.031359.

[23] M. R. Alvarez, C. Q. Vega, and L. Wong, "Model for Recognition of Personal Protective Equipment in Construction Applying YOLO-v5 and YOLO-v7," in International Conference on Electrical, Computer and Energy Technologies, ICECET 2023, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ICECET58911.2023.10389215.

[24] M. I. Al-Khiami and M. M. Elhadad, "ENHANCING CONSTRUCTION SITE SAFETY USING AI: THE DEVELOPMENT OF A CUSTOM YOLOV8 MODEL FOR PPE COMPLIANCE DETECTION," in Proceedings of the European Conference on Computing in Construction, European Council on Computing in Construction (EC3), 2024, pp. 577–584. doi: 10.35490/EC3.2024.307.

[25] P. Jankovic, M. Protic, L. Jovanovic, N. Bacanin, M. Zivkovic, and J. Kaljevic, "YOLOv8 Utilization in Occupational Health and Safety," in 2024 IEEE Zooming Innovation in Consumer Technologies Conference, ZINC 2024, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 182–187. doi: 10.1109/ZINC61849.2024.10579310.

[26] A. H. Hassan, E. M. Ahmed, J. M. Hussien, R. bin Sulaiman, M. Abdulhak, and H. Kahtan, "A cyber physical sustainable smart city framework toward society 5.0: Explainable AI for enhanced SDGs monitoring," Research in Globalization, vol. 10, Jun. 2025, doi: 10.1016/j.resglo.2025.100275.

[27] L. Wang, X. Zhang, and H. Yang, "Safety Helmet Wearing Detection Model Based on Improved YOLO-M," IEEE Access, vol. 11, pp. 26247–26257, 2023, doi: 10.1109/ACCESS.2023.3257183.

[28] R. Ayachi, Y. Said, M. Afif, A. Alshammari, M. Hleili, and A. Ben Abdelali, "Assessing YOLO models for real-time object detection in urban environments for advanced driver-assistance systems (ADAS)," Jun. 01, 2025, Elsevier B.V. doi: 10.1016/j.aej.2025.03.077.

[29] Glenn Jocher, Ayush Chaurasia, and Jing Qiu, "YOLOv8 - Ultralytics YOLO Docs." Accessed: Dec. 12, 2024. [Online]. Available: https://docs.ultralytics.com/models/yolov8/#key-features

[30] Universe Datasets, "Hard Hat Universe Dataset > Overview." Accessed: Dec. 02, 2024. [Online]. Available: https://universe.roboflow.com/universe-datasets/hard-hat-universe-0dy7t

[31] Roboflow Universe Projects, "Safety Vests Dataset > Overview." Accessed: Dec. 02, 2024. [Online]. Available: https://universe.roboflow.com/roboflow-universe-projects/safety-vests

[32] Yolov8, "Person Dataset." Accessed: Dec. 02, 2024. [Online]. Available: https://universe.roboflow.com/yolov8-mn29d/person-hecxr

[33] Ultralytics YOLO11, "YOLO11 🚀 NEW - Ultralytics YOLO Docs." Accessed: Dec. 04, 2024. [Online]. Available: https://docs.ultralytics.com/models/yolo11/#citations-and-acknowledgements