Sivanraj, S., Uduwage, D.N.L.S. and Tripathi, M., 2025. Comparison of YOLO algorithms for PPE compliance monitoring at construction sites. In: Waidyasekara, K.G.A.S., Jayasena, H.S., Wimalaratne, P.L.I. and Tennakoon, G.A. (eds). *Proceedings of the 13th World Construction Symposium*, 15-16 August 2025, Sri Lanka. pp. 438-450. DOI: https://doi.org/10.31705/WCS.2025.33. Available from: https://ciobwcs.com/papers/

COMPARISON OF YOLO ALGORITHMS FOR PPE COMPLIANCE MONITORING AT CONSTRUCTION SITES

S. Sivanraj[1], D.N.L.S. Uduwage[2] and M. Tripathi[3]

ABSTRACT

Safety is a critical concern in the construction industry, where workers are exposed to various hazards that can lead to serious injuries or fatalities. Personal Protective Equipment (PPE) is vital in protecting the workers and increasing their safety. However, ensuring consistent PPE compliance among construction workers remains a challenge. To overcome this challenge, this study developed automated PPE compliance monitoring models through the You Only Look Once (YOLO) object detection algorithm. The variants of YOLO algorithms such as YOLOv8-s, YOLOv9-s, and YOLOv11-s were trained to identify the best performance model to detect and classify the presence of humans and four major PPE items: helmets, high-visibility vests, gloves, and boots. To prevent the overfitting of the models, early stopping with a patience level of 20 epochs was set to the models. The system was externally validated using the construction sites' images to check its applicability to the Sri Lankan context. The efficacy of each model was assessed using performance evaluation matrices such as precision-recall curves, mean Average Precision (mAP@0.5), inference time, and loss function values. The results show that YOLOv9-s outperformed the other models in overall performance even though, it went through the highest number of epochs during training. Future work can explore enhancing the YOLOv9-s model performance by integrating motion detection through IoT devices, allowing for more precise tracking of PPE compliance and reducing false detections of idle or stored PPE. This approach could significantly improve real-time monitoring of PPE compliance for worker safety at construction sites.

Keywords: *Construction; Object Detection; Personal Protective Equipment; Safety; Workers.*

1. INTRODUCTION

Personal Protective Equipment (PPE) compliance is essential in construction sites to mitigate workplace hazards and ensure worker safety since the construction industry is one of the most hazardous sectors, with a high incidence of workplace injuries and fatalities (Ajmani et al., 2024). Construction sites often present hazards including falls, slips, trips, exposure to hazardous materials, operating heavy machinery (Bedi et al., 2021) and working at significant heights (Nadhim et al., 2016). Consequently, hazards caused by these at construction sites can be severe for construction workers. Therefore,

[1] Lecturer, Department of Building Economics, University of Moratuwa, Sri Lanka, sivanrajs.19@uom.lk
[2] Senior Lecturer, Department of Building Economics, University of Moratuwa, Sri Lanka, nuwanthas@uom.lk
[3] PhD Student, School of ICT, Thammasat University, Thailand, d6722300024@g.siit.tu.ac.th

safety regulations mandate the compliance of PPE to mitigate the risks (Muntiyono et al., 2021). PPE compliance refers to the proper use of PPE, which significantly reduces the likelihood of workplace hazards; however, ensuring compliance remains a key challenge (Trubetskov et al., 2023). Despite safety regulations, many workers fail to wear the required PPE, partly due to the inefficiency of site supervision and shortcomings in safety management systems (Wong et al., 2020).

According to Trubetskov et al. (2023), site supervision plays a crucial role in maintaining PPE compliance; however, manual supervision is often inefficient and costly (Raoofi et al., 2024). According to Tian (2023), traditional manual supervision is costly and time-consuming for construction sites, which makes it difficult for supervisors to monitor every worker in real-time, leading to increased workplace risks. As a result, there is a growing demand for automated solutions that can enhance PPE compliance monitoring without the need for constant human involvement.

Recent advancements in deep learning and computer vision provide significant alternatives to manual monitoring. According to Pal & Hsieh (2021), deep learning enables the analysis of complex visual data, improving the accuracy and efficiency of PPE detection. Similarly, Bai et al. (2023) found that automated PPE detection using deep learning ensures more consistent and efficient monitoring. Object detection models, particularly those based on the You Only Look Once (YOLO) architecture, have demonstrated high accuracy and speed in real-time applications within the construction industry (Bai et al., 2023; Lin, 2024; L. Wang et al., 2023)

Although several studies have evaluated the performance of YOLO models in detecting PPE compliance at construction sites, there remains a lack of research focusing on the detection of multiple PPE types using the latest YOLO iterations. This study introduces the PPE detection model using YOLOv11-s algorithm and compares its performance with the models developed using YOLOv8 and YOLOv9 algorithms to determine their effectiveness in the real-time detection of helmets, high-visibility vests, gloves and boots. Each model was trained, tested and validated on construction site images.

2. LITERATURE REVIEW

2.1 SIGNIFICANCE OF SAFETY AND PPE COMPLIANCE

Construction sites inherently pose significant risks due to the nature of activities performed, the wide range of equipment in operation, and the dynamic work environment (Sanni-Anibire et al., 2020). According to recent data on occupational hazards, the construction industry continues to experience disproportionately high levels of workplace incidents. Statistics indicate that in 2022, 22.9% of all fatal accidents within the European region occurred in the construction sector. The construction risks are linked to multiple factors, including working in areas prone to trips and slips, operating heavy machinery (Bedi et al., 2021), working at heights (Nadhim et al., 2016), exposure to hazardous materials (Blaauw & Maina, 2022) and exposure to electrocution. Many of these hazards can be avoided if construction sites adhere to PPE compliance.

PPE compliance refers to the proper use of essential safety equipment such as helmets, high-visibility vests, gloves, and boots. These PPEs play a vital role in protecting workers from the construction site hazards (Al-Bayati et al., 2023). Ensuring compliance is crucial in reducing workplace hazards, yet many construction workers fail to adhere to PPE

requirements. Al-Bayati et al. (2023) identified 16 factors that influence PPE compliance in construction, with supervision being a key determinant in ensuring adherence to safety protocols. To enhance PPE compliance monitoring, effective supervision is necessary. However, in large-scale construction projects, manually monitoring compliance can be often costly, time-consuming, and inefficient (Raoofi et al., 2024). Manual monitoring also introduces challenges such as human error, resource constraints, and the inability to track every worker in real-time (Tian, 2023). Due to these limitations, researchers have explored alternative solutions through computer vision-based systems to improve PPE compliance monitoring.

2.2 AUTOMATED PPE COMPLIANCE MONITORING

The use of computer vision has revolutionised construction site monitoring using real-time object detection (Yang et al., 2024). According to Ahmed et al. (2023), computer vision can be used to automate PPE monitoring and to detect compliant and non-compliant workers in real-time. Several studies have demonstrated the effectiveness of computer vision models in detecting PPE compliance. For instance, Delhi et al. (2020), developed a Convolutional Neural Network (CNN) model to detect the presence of construction helmets and high-visibility vests. Similarly, Ahmed et al. (2023) developed a CNN model to detect helmets of different colours, high-visibility vests and safety boots.

The CNN object detection models are broadly categorised into two major types: (1) one-stage object detection and (2) two-stage object detection. The key distinction between these types lies in their approach to regional proposal generation (García et al., 2021). Two-stage detection models, such as the Region-based Convolutional Neural Network (R-CNN), first generate region proposals before classifying and localising objects (Du et al., 2020). This additional step enhances detection accuracy, making them ideal for tasks that require precise object identification. However, one-stage object detection models such as YOLO directly classify and localise the object with high speed in a single pass (García et al., 2021). This approach significantly improves detection speed, making one-stage models highly suitable for real-time applications where fast inference has a higher priority.

2.3 YOLO MODELS FOR OBJECT DETECTION

YOLO is a single-stage object detection algorithm known for its speed and accuracy in detecting multiple objects. It consists of three major components: (1) backbone, (2) neck and (3) head. According to Mahasin & Dewi (2022), the backbone acts as a feature extractor, the neck is the feature fusion module finally, the head is responsible for the classification and localisation of the object.

The YOLO family has undergone several advancements and has been widely applied for PPE compliance monitoring. Research by Delhi et al. (2020) developed a construction safety helmet and high visibility vest detection model using YOLOv3. Similarly, Velasco & Marasigan (2024) investigated the use of YOLOv7 for PPE compliance detection validating its effectiveness in real-time safety monitoring. Additionally, Guney et al. (2024) conducted a comparative study to evaluate the performance of YOLO-NAS, YOLOv8 and YOLOv9 for PPE compliance monitoring. According to Rasheed & Zarkoosh (2024), YOLOv11 is one of the latest versions of the YOLO family. Each YOLO model has variants tailored to diverse computational and application needs (Kishor, 2024). The variants of YOLOv8, YOLOv9 and YOLOv11 vary from nano to

extra-large allowing users to balance speed, accuracy and computational efficiency. The nano variant models are particularly useful for low-power devices which focus on speed, whereas the extra-large variants are useful for high-performance systems. The small model offers a balance between speed and accuracy which is suitable for medium-scale applications such as real-time PPE compliance monitoring.

2.4 SIGNIFICANCE OF THE STUDY

Studies have shown that manual monitoring of construction sites can be costly, time-consuming and inefficient. The challenges of manual construction site monitoring can be overcome by automated techniques especially by using single-stage object detection models such as YOLO algorithms. There are various advancements in the YOLO family, each consisting of a different capability. Several studies have attempted to use YOLO models for PPE compliance detections (Delhi et al., 2020; Velasco & Marasigan, 2024) and a few have carried out comparisons of YOLO models such as Guney et al. (2024). However, no studies have attempted to compare the latest iterations of YOLO, such as YOLOv11, to assess its performance in terms of PPE compliance detections. Therefore, this study was carried out to determine the best YOLO iteration for PPE compliance monitoring.

3. METHODOLOGY

According to Wang et al. (2025), YOLOv10 is an efficient model designed for general object detection; however, it does not inherently support oriented bounding box detection, limiting its suitability for applications requiring precise object localisation. Therefore, this study used three variants of YOLO algorithms such as YOLOv8-s, YOLOv9-s, and YOLOv11-s to develop three PPE detection models. The models detect humans and four essential PPEs such as helmets, high-visibility vests, gloves and boots. The models were developed on an online platform, and their performance was evaluated and compared based on PR values, mAP, and inference speed.

3.1 DATASET

One of the critical stages in developing a model using the YOLO algorithm is collecting a relevant dataset. In this study, 2,092 images were obtained from an online database called RoboFlow. This online database of images was captured from various construction sites without any specific consideration for the geographical location of the site. Therefore, a previously available dataset was suitable for this study. The same dataset of images was consistently used across training, testing, and validation phases of the model to ensure fair comparison. As such, environmental factors such as lighting conditions, weather, or site-specific occlusions were not varied or controlled. However, all photographs were taken during the day, under natural lighting conditions. The dataset was split into 70% for training, 20% for internal validation, and 10% for testing. Furthermore, custom images from local construction sites were collected to externally validate and assess the applicability of the model to the Sri Lankan projects. To maintain consistency across training, validation, and testing phases, all images were pre-processed to a fixed resolution of 640×640×3 in the Red-Green-Blue (RGB) colour scale.

3.2 MODEL TRAINING

This section outlines the training methodology applied in this study. By maintaining the same training parameters and procedures, the evaluation aimed to identify the most optimal YOLO model for PPE detection without bias from differing training setups.

3.2.1 Hyperparameter

Within YOLO models, hyperparameters play a crucial role in balancing model performance and computational efficiency. Several key hyperparameters were utilised in this study, including batch size, learning rate, momentum, and weight decay.

The dataset images were fed into the model batch-wise, with a batch size of 64 images, allowing for efficient training. The learning rate, which determines how quickly a model learns during training, was set at 0.01. The momentum, which helps to accelerate learning and smooth gradient updates, was set at 0.937. Weight decay, a regularisation technique used to prevent overfitting was set at 0.0005. These hyperparameters are preset within the model but can be fine-tuned for optimisation.

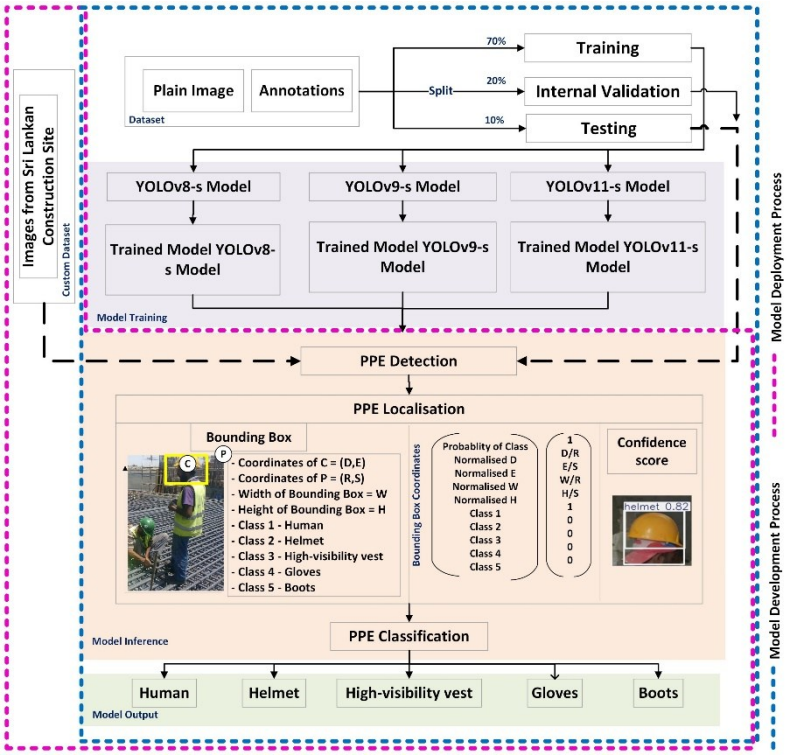

Figure 1: Overall architectural diagram for PPE detection models

Additionally, the number of epochs, which determines how many times the dataset passes through the entire training cycle, was set to 300 epochs in this study. However, prolonged training can lead to overfitting, where the model learns patterns specific to the training data but fails to generalise well to new inputs. To mitigate this, early stopping with patience set to 20 epochs was implemented. This technique prevents overfitting by stopping further training of the model when the model's performance stops improving. The patience value of 20 epochs specifies that the training will automatically stop if there is no improvement in evaluation metrics for 20 consecutive epochs. Accordingly,

YOLOv8-s, YOLOv9-s and YOLOv11-s models went through total epochs of 164, 300 and 160, respectively, during the training and validation process.

3.2.2 Training Process

After splitting the dataset, the three models were trained separately, as shown in Figure 1. The blue dotted line illustrates the model development process, while the pink dotted line represents the model deployment process. In the model development, the training phase utilised two types of input data: (1) plain images and (2) annotations. The annotation provides coordinates of the localised objects such as humans, helmets, high-visibility vests, gloves, and boots. With each training epoch, the system underwent validation. Upon completion of training, the models were further tested to determine and compare the key evaluation matrices. The time to train YOLOv8-s, YOLOv9-s, and YOLOv11-s was 1.273 hours, 3.612 hours and 1.420 hours, respectively.

3.3 MODEL VALIDATION

The validation step in YOLO models is crucial for assessing model performance and generalisation. In this study, the proposed three models underwent two phases of validation: (1) internal validation and (2) external validation. Internal validation was performed using 20% of the dataset from an online database, ensuring that the models were validated in parallel for each training iteration. This process allowed for continuous performance monitoring and determined the optimal number of epochs required for training. By analysing validation results, training was halted at the appropriate stage to prevent overfitting and ensure model generalisation.

Once the models were fully trained and tested, external validation was conducted to evaluate their adaptability to the Sri Lankan context. This phase involved testing the models on custom images collected from Sri Lankan construction sites. These images, captured using mobile phones, provided a diverse and realistic dataset which included various angles, lighting conditions, and image qualities, including both daytime and nighttime scenarios.

3.4 MODEL INFERENCE

Once the models were fully trained, they were tested using 10% of the dataset through the inference pipeline. Model inference refers to the process of using a trained model to predict outputs for an unseen dataset. In the inference pipeline, each model first detected the presence of PPE in the input image. The models then generated bounding boxes with coordinates and confidence scores, allowing them to localise humans and four PPE classes: helmets, high-visibility vests, gloves, and boots. The bounding box coordinates were determined based on key parameters, including the bounding box size, the centre coordinates of the bounding box, and the classes of detected objects within the image with human as class 1, helmet as class 2, high-visibility vest as class 3, gloves as class 4 and boots as class 5.

As illustrated in Figure 1, Point C represents the centre of the bounding box, while Point P denotes the top-right corner of the image. The confidence score associated with each bounding box indicates the model's certainty in detecting the object. While a higher confidence score signifies greater certainty, the detection threshold must be adjusted based on object size to ensure accurate classification. In this study, a confidence threshold

of 25% (or 0.25) was used to balance detection accuracy and false positives. Following localisation, the models classified the detected objects into five classes.

4. RESULTS AND PERFORMANCE EVALUATION

4.1 CONFUSION MATRIX

According to Caelen (2017), a confusion matrix evaluates the performance of classification by a model. It provides the breakdown of the model's prediction against the actual outcomes. When there are six distinct classification categories, a 6×6 confusion matrix can be used to analyse the model's performance. A 2x2 confusion matrix has been given in Table 1.

Table 1: Standard confusion matrix

	Actual Positive	**Actual Negative**
Predicted Positive	True Positive (TP)	False Positive (FP)
Predicted Negative	False Negative (FN)	True Negative (TN)

In the PPE detection models, a true positive would mean correctly detecting a class of Human or PPE, whereas a true negative would indicate correctly recognising the absence of PPE. A false positive would occur if the model mistakenly detected a different class. In the proposed models, there are five primary classes: human, helmet, high-visibility vest, gloves, and boots. These classes represent different PPE components and human presence detected within construction site images. Additionally, when none of these five classes is present in an image, it is labelled as background. During inference, if an image contains no PPE or human presence, it is classified under this background category.

Figures 2, 3, and 4 present the normalised confusion matrices for YOLOv8-s, YOLOv9-s, and YOLOv11-s, respectively. The values have been normalised to ensure comparability across different class distributions. These matrices show the accuracy of true positive classifications, reflecting how well each model classified the presence of different PPE classes. The overall accuracy of each model in correctly classifying PPE items is summarised in Table 2.

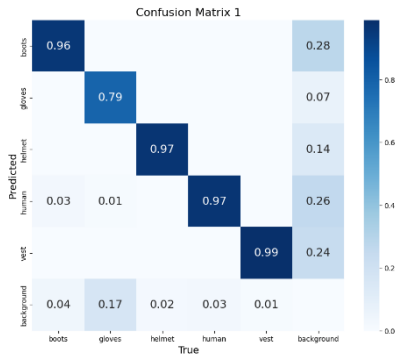

Figure 2: Normalised confusion matrix for YOLOv8-s model

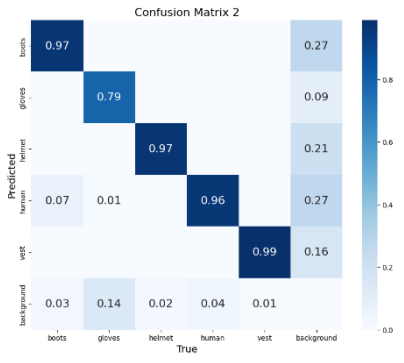

Figure 3: Normalised confusion matrix for YOLOv9-s model

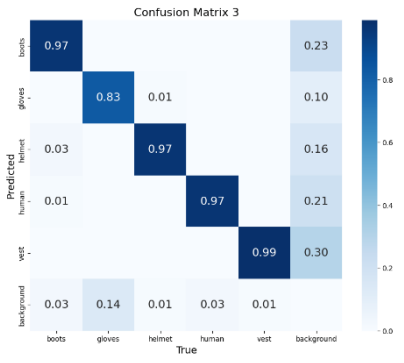

Figure 4: Normalised confusion matrix for YOLOv11-s model

Table 2: Comparison of true positive accuracy of the three models

	YOLOv8-s	YOLOv9-s	YOLOv11-s
Human	0.97	0.96	0.97
Helmet	0.97	0.97	0.97
High-visibility vest	0.99	0.99	0.99
Gloves	0.79	0.79	0.83
Boots	0.96	0.97	0.97

Across all three models, the detection accuracy for humans, helmets, high-visibility vests, and boots remained consistently high, with the high-visibility vest class demonstrating the best performance. However, gloves presented a greater challenge, as all three models exhibited comparatively lower accuracy in detecting them. Instances of misclassification were observed, where images containing no PPE classes were incorrectly classified as containing boots, gloves, helmets, humans, or high-visibility vests. This misclassification occurred across all models, indicating difficulties in distinguishing background elements from actual PPE instances. The highest misclassification rate was recorded in the YOLOv11-s model, where 30% of the misclassified cases involved the incorrect detection of a high-visibility vest in images where none were present, suggesting potential overfitting in the model.

4.2 PRECISION-RECALL CURVE

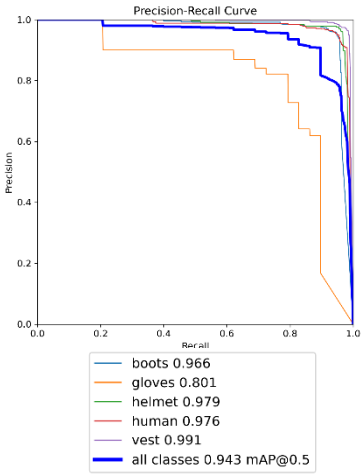

Figure 5: YOLOv8-s PR curve

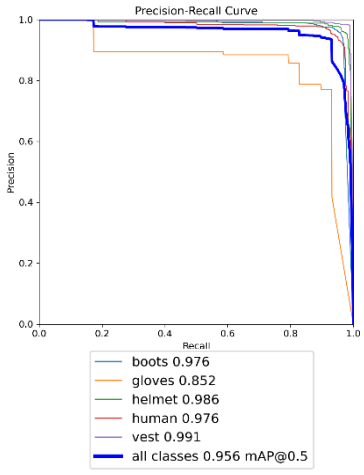

Figure 6: YOLOv9-s PR curve

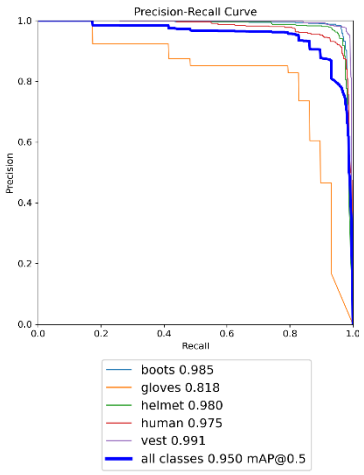

Figure 7: YOLOv11-s PR curve

The models' performance was further tested using precision and recall which are two key metrics for evaluating the performance of the YOLO models. Precision (Equation 1) measures how accurately the model identifies the true positive and shows how often the model is correct when making the classification. Recall (Equation 2), on the other hand, measures the model's ability to detect all relevant instances in a dataset and shows how well a model classifies all the actual classes.

$$Precision = \frac{TP}{TP+FP} \quad Equation\ (1) \qquad Recall = \frac{TP}{TP+FN} \quad Equation\ (2)$$

There is often a trade-off between precision and recall, where improving one may negatively impact the other. A higher precision ensures fewer false positives, while a higher recall reduces false negatives. Therefore, a Precision-Recall (PR) curve was

generated to analyse the trade-off between these two metrics and assess the overall performance of each model.

Figures 5, 6, and 7 present the precision-recall curves for YOLOv8-s, YOLOv9-s, and YOLOv11-s, respectively, in PPE detection. These curves illustrate how well each model maintained a balance between precision and recall. A higher area under the PR curve (AUC-PR) indicates better performance, as it reflects strong classification capabilities. The PR values for YOLOv8-s, YOLOv9-s, and YOLOv11-s models have been summarised in Table 3.

Table 3: Comparison of PR values of the three models

	YOLOv8-s	YOLOv9-s	YOLOv11-s
Human	0.976	0.976	0.975
Helmet	0.979	0.986	0.980
High-visibility vest	0.991	0.991	0.991
Gloves	0.801	0.852	0.818
Boots	0.966	0.976	0.985

Overall, the YOLOv9-s had the highest PR value for 4 out of 5 classes compared to the other two models which had the highest PR values for 2 out of 5 classes. This makes YOLOv9-s the most effective model in terms of precision-recall performance.

4.3 MEAN AVERAGE PRECISION (MAP)

The mAP (Equation 3) measures the average precision at a single Intersection over Union (IoU) threshold, typically set at 50% (mAP@0.5). This provides a single metric that can be used to directly compare the models' performance.

$$mAP@0.5 = \frac{1}{number\ of\ classes\ (n)} \sum_{i=1}^{n.} Average\ Precision \quad Equation\ (3)$$

The mAP scores for YOLOv8-s, YOLOv9-s, and YOLOv11-s were 0.943, 0.956, and 0.950, respectively. These results indicate that YOLOv9-s achieved the highest overall performance, followed closely by YOLOv11-s, while YOLOv8-s had the lowest mAP.

4.4 INFERENCE TIME

The inference time refers to the total time a model takes to process a new unseen image and generate a detection and classification output. A lower inference indicates faster processing, which is crucial for real-time detection applications. The inference times for the three models were 4.4 ms for YOLOv8-s, 7.3 ms for YOLOv9-s, and 5.4 ms for YOLOv11-s. These results indicate that YOLOv8-s is the fastest model, making it the most suitable for real-time PPE detection.

4.5 TRAINING AND VALIDATION LOSS

The YOLO models utilise a unified loss function to evaluate performance and prevent overfitting during training. This loss function consists of three primary components: box loss, classification loss, and distribution focal loss (DFL). The box loss ("box_loss") measures the accuracy of bounding box predictions by evaluating how well the detected object's bounding box aligns with the ground truth. This ensures the precise localisation

of PPE items. The classification loss ("cls_loss") assesses the accuracy of class predictions, determining whether the model correctly identifies PPE categories such as helmets, vests, gloves, and boots. Finally, the distribution Focal Loss ("dfl_loss") enhances refined classification by improving object boundary precision, ensuring that predicted bounding boxes align more accurately with their respective objects.

Figures 8, 9, and 10 illustrate the loss function trends for the three models. As the number of epochs increased, the loss function steadily decreased. Similarly, the precision and recall curves showed a consistent upward trend with increasing epochs. The number of epochs for each model was limited by a patience level of 20 epochs. The training and validation loss values of the three models are summarised in Table 4.

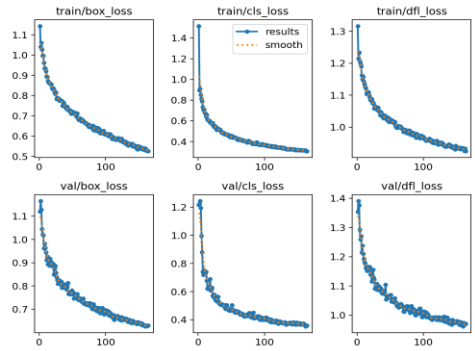

Figure 8: Loss function curve for YOLOv8-s Model for 164 epochs

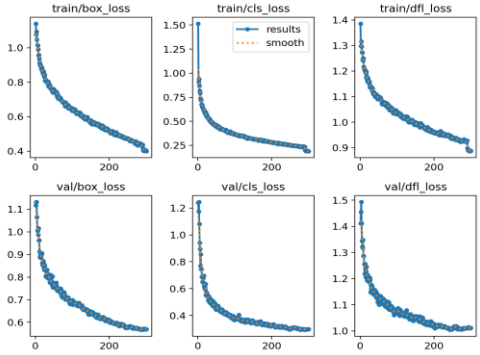

Figure 9: Loss function curve for YOLOv9-s Model for 300 epochs

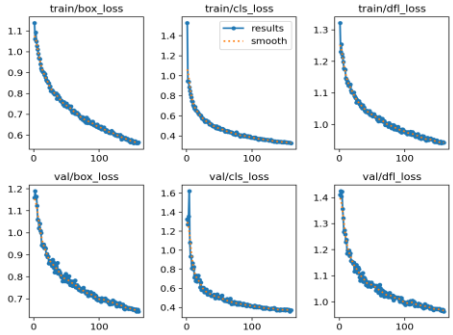

Figure 10: Loss function curve for YOLOv11-s Model for 160 epochs

YOLOv9-s, which was trained for the entire 300 epochs, achieved the lowest box loss, classification loss, and DFL loss among the three models. This suggests that YOLOv9-s had the best bounding box accuracy and classification performance. In contrast, YOLOv8-s and YOLOv11-s were trained for fewer epochs and exhibited higher loss values. The slightly higher loss values in YOLOv8-s and YOLOv11-s suggest they did not reach the same level of optimisation as YOLOv9-s.

Table 4: Comparison of training and validation loss

	YOLOv8-s	**YOLOv9-s**	**YOLOv11-s**
Training Epochs	164	300	160
Box Loss	0.5251	0.3991	0.5654
Classification Loss	0.3071	0.1903	0.3274
DFL Loss	0.9243	0.8883	0.9428

4.6 Overall Comparison of the Models

Table 5 presents the best-performing model among YOLOv8-s, YOLOv9-s, and YOLOv11-s for each performance evaluation metric. The results indicate that YOLOv9-s achieved the highest performance across all metrics, including True Positive Accuracy, PR Value, mAP@0.5 scores, and Training & Validation Loss. However, in terms of Inference Time, YOLOv8-s performed better, demonstrating a faster processing speed.

Table 5: Overall comparison of the model

	YOLOv8-s	YOLOv9-s	YOLOv11-s
True Positive accuracy	✓	✓	✓
PR Value		✓	
mAP@0.5 scores		✓	
Inference Time	✓		
Training and validation loss		✓	

4.7 Model Detections

The detections made by the models are presented in Figures 11 to 16, where each detection is marked with a bounding box and confidence score, indicating the model's level of certainty in its predictions. Comparatively higher confidence scores can be observed for the detections made by YOLOv9-s, indicating its superior performance in identifying PPE with greater certainty. Some inaccuracies were observed in the detections, suggesting areas for improvement. These limitations can be addressed by further training the models using a larger and more diverse dataset to enhance detection accuracy.

Figure 11: Detection by YOLOv8-s

Figure 12: Detection by YOLOv9-s

Figure 13: Detection by YOLOv11-s

Figure 14: Detection by YOLOv8-s

Figure 15: Detection by YOLOv9-s

Figure 16: Detection by YOLOv11-s

5. CONCLUSION AND RECOMMENDATIONS

This study explored object detection for PPE compliance monitoring using YOLOv8-s, YOLOv9-s, and YOLOv11-s models. A dataset from an online database was utilised for training, enabling the models to detect the presence of humans and four essential PPE items, including helmets, high-visibility vests, gloves, and boots. The models were evaluated based on precision-recall performance, mean Average Precision (mAP@0.5), inference time, and loss function analysis to determine their effectiveness in real-time PPE detection. Among the three models, YOLOv9-s demonstrated the best overall performance, achieving the highest mAP@0.5 and the lowest loss values, making it the most optimised for PPE compliance monitoring. However, YOLOv8-s had the fastest inference time. This study contributes to construction safety by identifying an optimal YOLO-based model for PPE detection in construction safety monitoring. However, the model is limited to detecting only helmets, gloves, high-visibility vests, and safety boots, as other essential PPE, such as eye or hearing protection, was not considered. Additionally, factors such as varying lighting and weather conditions were not considered, which may affect generalisability.

While the proposed models effectively detected PPE compliance, future research should focus on enhancing PPE monitoring by incorporating personalised PPE requirements based on specific worker roles rather than relying on generalised detection using YOLOv9-s. Additionally, integrating motion analysis would improve real-time tracking accuracy. This enhancement would also help distinguish actively worn PPE from idle or stored equipment, reducing false detections and improving overall system reliability.

6. REFERENCES.

Ahmed, M. I. B., Saraireh, L., Rahman, A., Al-Qarawi, S., Mhran, A., Al-Jalaoud, J., Al-Mudaifer, D., Al-Haidar, F., AlKhulaifi, D., Youldash, M., & Gollapalli, M. (2023). Personal protective equipment detection: A deep-learning-based sustainable approach. *Sustainability*, *15*(18). https://doi.org/10.3390/su151813990

Ajmani, F., Belamine, F., Derqaoui, M. A., Abouzahir, H., Belhouss, A., & Benyaich, H. (2024). Fatal workplace accidents. *Occupational Medicine*, *74*(Supplement_1). https://doi.org/10.1093/occmed/kqae023.0915

Al-Bayati, A. J., Rener, A. T., Listello, M. P., & Mohamed, M. (2023). PPE non-compliance among construction workers: An assessment of contributing factors utilizing fuzzy theory. *Journal of Safety Research*, *85*, 242–253. https://doi.org/10.1016/j.jsr.2023.02.008

Bai, R., Wang, M., Zhang, Z., Lu, J., & Shen, F. (2023). Automated construction site monitoring based on improved YOLOv8-seg instance segmentation algorithm. *IEEE Access*, *11*, 139082–139096. https://doi.org/10.1109/ACCESS.2023.3340895

Bedi, J. K., Rahman, R. A., & Din, Z. (2021). Heavy machinery operators: Necessary competencies to reduce construction accidents. *IOP Conference Series: Earth and Environmental Science*, *641*(1), 012007. https://doi.org/10.1088/1755-1315/641/1/012007

Blaauw, S., & Maina, J. (2022). Exposure of construction workers to hazardous emissions in highway rehabilitation projects. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4086555

Caelen, O. (2017). A Bayesian interpretation of the confusion matrix. *Annals of Mathematics and Artificial Intelligence*, *81*, 429–450. https://doi.org/10.1007/s10472-017-9564-8

Delhi, V. S. K., Sankarlal, R., & Thomas, A. (2020). Detection of personal protective equipment (PPE) compliance on construction site using computer vision based deep learning techniques. *Frontiers in Built Environment*, *6*. https://doi.org/10.3389/fbuil.2020.00136

Du, L., Zhang, R., & Wang, X. (2020). Overview of two-stage object detection algorithms. *Journal of Physics: Conference Series*, *1544*, 012033. https://doi.org/10.1088/1742-6596/1544/1/012033

García, M. C., Mateo, J. T., Benítez, P. L., & Gutiérrez, J. G. (2021). On the performance of one-stage and two-stage object detectors in autonomous vehicles using camera data. *Remote Sensing*, *13*(1), 89. https://doi.org/10.3390/rs13010089

Guney, E., Altin, H., Esra Asci, A., Bayilmis, O. U., & Bayilmis, C. (2024). YOLO-based personal protective equipment monitoring system for workplace safety. *JITSI: Jurnal Ilmiah Teknologi Sistem Informasi*, *5*(2), 77–85. https://doi.org/10.62527/jitsi.5.2.238

Kishor, R. (2024). Performance benchmarking of YOLOv11 variants for real-time delivery vehicle detection: a study on accuracy, speed, and computational trade-offs. *Asian Journal of Research in Computer Science*, *17*(12), 108–122. https://doi.org/10.9734/ajrcos/2024/v17i12532

Lin, B. (2024). Safety helmet detection based on improved YOLOv8. *IEEE Access*, *12*, 28260–28272. https://doi.org/10.1109/ACCESS.2024.3368161

Mahasin, M., & Dewi, I. A. (2022). Comparison of CSPDarkNet53, CSPResNeXt-50, and EfficientNet-B0 Backbones on YOLO V4 as object detector. *International Journal of Engineering, Science, and Information Technology*, *2*(3), 64-72. https://doi.org/10.52088/ijesty.v2i3.291

Muntiyono, Haris, S., Nuryono, B., Mahardika, A. G., Fadriani, H., & Hidayat, I. (2021). Protection of construction workers with personal protective equipment. In *Advances in Social Science, Education and Humanities Research: Proceedings of the 1st Paris Van Java International Seminar on Health, Economics, Social Science and Humanities*, (pp. 388-391). Atlantis press. https://doi.org/10.2991/assehr.k.210304.085

Nadhim, E.A., Hon, C., Xia, B., Stewart, I., & Fang, D. (2016). Falls from height in the construction industry: A critical review of the scientific literature. *International Journal of Environmental Research and Public Health*, *13*(7), 638. https://doi.org/10.3390/ijerph13070638

Pal, A., & Hsieh, S.-H. (2021). Deep-learning-based visual data analytics for smart construction management. *Automation in Construction*, *131*, 103892. https://doi.org/10.1016/j.autcon.2021.103892

Raoofi, H., Sabahnia, A., Barbeau, D., & Motamedi, A. (2024). Deep learning method to detect missing welds for the joist assembly line. *Applied System Innovation*, *7*(1), 16. https://doi.org/10.3390/asi7010016

Rasheed, A. F., & Zarkoosh, M. (2024). YOLOv11 optimization for efficient resource utilization. *Computer Vision and Pattern Recognition*. https://doi.org/10.48550/arXiv.2412.14790

Sanni-Anibire, M. O., Mahmoud, A. S., Hassanain, M. A., & Salami, B. A. (2020). A risk assessment approach for enhancing construction safety performance. *Safety Science*, *121*, 15–29. https://doi.org/10.1016/j.ssci.2019.08.044

Tian, Y. (2023). Design and research on supervision project management system based on system engineering. In *Proceedings of the 2023 4th International Conference on Computer Science and Management Technology*, (pp. 20–24). https://doi.org/10.1145/3644523.3644528

Trubetskov, A. D., Makhonko, M. N., Shkrobova, N. V., & Shelekhova, T. V. (2023). Problems of using personal protective equipment in modern conditions. *Russian Journal of Occupational Health and Industrial Ecology*, *63*(5), 336–343. https://doi.org/10.31089/1026-9428-2023-63-5-336-343

Velasco, M. A., & Marasigan, R. I. (2024, July). *Enhancing personal protective equipment compliance and usage accuracy on construction sites through YOLOv7 detection*. 2024 IEEE 7th International Conference on Big Data and Artificial Intelligence (BDAI), 205–212. https://doi.org/10.1109/BDAI62182.2024.10692912

Wang, L., Zhang, X., & Yang, H. (2023). Safety helmet wearing detection model based on improved YOLO-M. *IEEE Access*, *11*, 26247–26257. https://doi.org/10.1109/ACCESS.2023.3257183

Wang, X., Han, C., Huang, L., Nie, T., Liu, X., Liu, H., & Li, M. (2025). AG-Yolo: Attention-guided yolo for efficient remote sensing-oriented object detection. *Remote Sensing*, *17*(6), 1027. https://doi.org/10.3390/rs17061027

Wong, T. K. M., Man, S. S., & Chan, A. H. S. (2020). Critical factors for the use or non-use of personal protective equipment amongst construction workers. *Safety Science*, *126*, 104663. https://doi.org/10.1016/j.ssci.2020.104663

Yang, J., Lee, E., Kwon, J., Lee, D., Kim, Y., Park, C., & Lee, D. (2024). A weakly supervised learning framework utilizing enhanced class activation map for object detection in construction sites. *IEEE Access*, *12*, 99989–100004. https://doi.org/10.1109/ACCESS.2024.3423697