 ISSN:2302-8149
e-ISSN:2540-9719

# Deep Learning-based Identification of Personal Protective Equipment in Construction Area

**[1]Mutia Fadhilla*, [2]Sapitri, [3]Rizky Wandri**
[1,3]Program Studi Teknik Informatika, Fakultas Teknik, Universitas Islam Riau
[2]Program Studi Teknik Sipil, Fakultas Teknik, Universitas Islam Riau
[1,2,3]Jl. Kaharuddin Nst No.113, Simpang Tiga, Kec. Bukit Raya, Pekanbaru, Riau, Indonesia
*e-mail: tiafadhilla@eng.uir.ac.id

(***received*:** 5 May 2025, ***revised*:** 20 May 2025, ***accepted*:** 21 May 2025)

## Abstract

Work safety in the construction environment is highly dependent on the correct use of personal protective equipment (PPE). This study aims to develop an automatic PPE detection system using a YOLO-based deep learning model to improve supervision and compliance with PPE use in the field. Two variants of the YOLO model, namely YOLOv10 and YOLOv11, were tested and their performance was compared through a fine-tuning process using custom dataset consisting of 16,568 annotated images of construction workers wearing various types of PPE. The model was evaluated using precision, recall, and mAP50. The results showed that the YOLOv11s model performed the best with with an mAP50 of 0.718 and a precision score of 0.804, indicating good detection and classification ability. This model is able to detect various types of PPE effectively, so it can be used as a tool in real-time occupational safety monitoring. This study proves that the application of YOLO-based deep learning technology can be an effective solution to improve compliance with PPE use and reduce the risk of work accidents in the construction sector. The implications of this study open up opportunities for the development of more sophisticated and adaptive automatic monitoring systems in the future such as deploying the model on edge devices for real-time inference and expanding detection capabilities to include additional safety violations such as the absence of safety harnesses or proximity to hazardous zones.

**Keywords:** PPE detection, deep learning, computer vision, YOLO

## 1 Introduction

Work safety in the construction industry is the foundation and main priority in determining the success and sustainability of projects. This is caused by the construction work environment which is complex and full of potential dangers. Workers in this sector are faced with various risks that threaten their safety. These risks include the use of sharp equipment such as saws, knives, and other cutting tools that can cause serious injuries[1], [2]. In addition, there are heavy machines such as cranes and excavators where small errors can be fatal, both for operators and workers around them. The threat of falling objects such as building materials or work tools can also be a cause of accidents at construction sites. Therefore, the use of Personal Protective Equipment (PPE) is very crucial for construction workers. PPE such as helmets, safety shoes, and protective glasses are designed to provide extra protection to workers from various potential hazards. Proper and consistent use of PPE can significantly reduce the risk of accidents and injuries in the workplace.

Although various safety regulations have been implemented and various types of PPE are available, ensuring that PPE is used correctly and consistently on construction sites remains a significant challenge. This challenge arises due to several factors, one of which is that the safety monitoring process, such as the use of PPE, is still carried out manually. This conventional method certainly requires a lot of labor to supervise and ensure that each worker complies with the established safety protocols. In addition, this method is inadequate to handle large-scale construction operations, where the number of workers and the complexity of the project increase significantly, as well as the wide area coverage and various activities that take place simultaneously. This can cause safety aspects to be neglected, thereby increasing the risk of workplace accidents and injuries. Therefore, technological innovation is needed to monitor compliance with safety standards, so that challenges in

 ISSN:2302-8149
e-ISSN:2540-9719

ensuring workers use consistent and correct PPE can be overcome and a safe work environment created.

The development of technology in the field of computer vision has made rapid progress in recent years. This is driven by significant advances in deep learning and the availability of large data sets and powerful computing resources. One of the key elements in this progress is the development of the Convolutional Neural Network (CNN) model which revolutionizes the way computers process and understand visual information. In computer vision, object recognition and detection in visual data, such as images and videos, is a fundamental task that has been applied in various industries. CNN-based object detection allows computers to interpret and understand visual information more accurately and efficiently. This approach can be used to detect PPE use in real-time and automatically and provide a reliable solution to ensure consistent use of work safety protocols on construction sites. The system can be trained using CNN to recognize various types of PPE with a relatively high level of accuracy. Apart from that, this system can also be developed by integrating it into surveillance cameras or drones to monitor the construction work environment in a wider and more detailed manner. Thus, reducing dependence on manual supervision which is susceptible to human error.

The implementation of deep learning with the CNN model approach to detect the completeness of PPE has been widely carried out in previous studies. Most previous studies focus on building a CNN model to detect one PPE tool, for example a safety helmet [3], [4] and gloves[5], [6]. This is certainly not effective when applied to real situations where many types of PPE must be recognized simultaneously. Several other studies have built CNN models to detect several PPE tools, but the number of tools detected is still limited to certain body parts such as head safety[7]. However, this approach does not have comprehensive results to be applied to the real context of the construction industry. The construction environment has a high risk of work accidents. Therefore, occupational safety and health regulations require the use of various types of PPE simultaneously, such as safety helmets, safety vests and safety gloves. If the system only detects one type of PPE, for example a helmet, then violations of the obligation to use other PPE may not be detected, even though the risk remains high. In other words, the system does not reflect the completeness of PPE as a whole, which is an important requirement in work safety standards in the field. Therefore, it is necessary to develop a deep learning model for multi-PPE detection.

Based on the problems and limitations of previous research discussed above, in this research a deep learning model will be built to detect PPE in construction areas using a CNN architecture. In this research, the CNN model will be trained using a fine-tuning approach by comparing two types of architecture, namely YOLOv10 and YOLOv11, to detect several PPE that are commonly used by construction workers, including safety helmets, gloves and vests.

## 2 Literature Review

This study aims to detect the use of Personal Protective Equipment (PPE) in construction areas using the Convolutional Neural Network (CNN) model. This section will discuss previous studies that are relevant to object detection, especially PPE using a deep learning approach. Where the focus is the implementation of CNN model in the context of work safety.

Many studies related to computer vision in detecting PPE by implementing CNN architecture have been conducted previously. Most of the studies focus on detecting one of the PPE, especially safety helmets, in industrial and construction areas. The study conducted by [8] implemented CNN with the YOLOv3 model transfer learning approach in detecting safety helmets with the aim of improving supervision of their use. However, the accuracy level obtained was still below 90%. With this level of accuracy for one type of PPE indicates that the system is not yet robust enough even for one element, especially if extended to multi-class or multi-label detection. This indicates limitations in the generalization of the model. Other studies used other versions of YOLO in detecting safety helmets, namely YOLOv4 [9], YOLOv5 [10], [11], and YOLOv7 [12], with an increase in the level of accuracy in the evaluation results. Meanwhile, [11]conducted fine tuning modeling of YOLOv5 by adding architectural blocks for the attention model that separates foreground and background to improve detection accuracy. In addition, other study [4] conducted a comparative analysis using the YOLOv5, YOLOv6, and YOLOv7 models in carrying out safety helmet detection tasks.

 ISSN:2302-8149
e-ISSN:2540-9719

Furthermore, other studies focus on deep learning modeling for other PPE detection, namely glove usage detection. [13]conducted research to recognize PPE gloves using an object recognition approach. In this study, there are several pre-trained CNN models used to classify workers using or not using gloves, including VGG-19 and ResNet50. Other related study [6] adapted the YOLOv4 model to detect glove usage, especially for workers in the laboratory. While the research conducted by [5] developed a damage detection system for gloves being used by workers using a fine-tuning approach on the YOLOv5 model. These previous studies show that the system developed only focuses on building a CNN model to detect one piece of PPE, which is certainly not effective in ensuring the complete use of PPE in complex environments such as industrial and construction areas, so further research needs to be carried out to detect the completeness of various types of PPE.

Research in detecting more than one type of PPE has also been done a lot before. Some of them [7], [14] build a system that can detect the completeness of PPE that focuses on head safety equipment, such as helmets and masks. Study proposed by [14] built system based on image processing and object recognition using CNN architecture to detect and monitor worker discipline in using PPE on the head, while [7] conducts similar research but already uses the CNN object detection approach with the YOLOv5 model so that it can detect several PPE tools on the head simultaneously and in real-time. Meanwhile, another study [3] develops a safety helmet and vest detection system in the construction, mining, and law enforcement areas. This study implements the object detection approach with the YOLOv8 model in the developed detection system. Other studies have created a system for detecting multiple PPE with the CNN deep learning approach, but there are still few studies that focus more on PPE detection systems on construction sites. Most of the previous studies that are currently developing PPE detection systems in the health sector[15], [16], [17], especially to prevent the spread of COVID-19 or similar viruses. Research conducted by [18] builds a detection system for the completeness of PPE use in heavy equipment workshops. However, in this study, images were inputted via webcam, which has limitations in overall and complex vision. Meanwhile, other research developed detection of complete use of PPE among workers in the oil and gas industry [19] and the firefighting sector [20].

In conclusion, based on the literacy studies carried out, there are still several limitations that previous studies have. Most previous studies are still limited to the detection system of one type of PPE such as safety helmet or safety gloves. Several other studies focused on creating a system for detecting PPE use on the head. Other studies that have built a complete or multi-PPE detection system are still focused on sectors other than construction, such as the oil and gas, mining, firefighting and health sectors. In addition, a computer vision-based approach by implementing the YOLO model offers the potential to build an efficient real-time detection system. This research will explore its application in the context of accurate and efficient PPE detection to improve work safety in construction areas.

## 3 Research Method

The main process carried out in this study can be seen in Figure 1. The first stage carried out is data collection. The data that has been collected is then annotated and labeled, then divided into three parts, namely training data, evaluation, and testing. Next, a model training process was carried out using a fine-tuning approach using two versions of You Only Look Once (YOLO) models which are YOLOv10 and YOLOv11, then ended with model evaluation.

 ISSN:2302-8149
e-ISSN:2540-9719

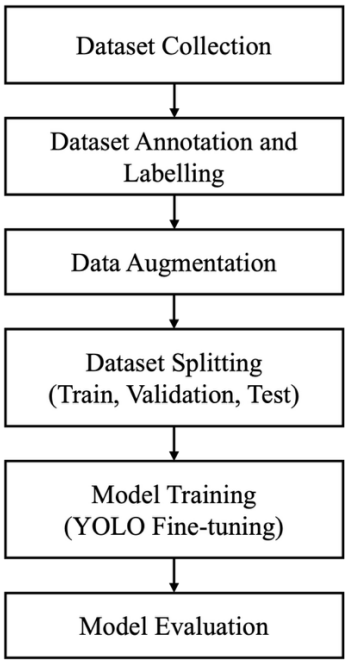


**Figure 1. Stages of research process**

### 3.1 Dataset Collection

In this initial stage, data is collected from two main sources to build a comprehensive and representative dataset. First, data is obtained from videos accessed through social media platforms. These videos are selected based on their relevance to the research object, namely personal protective equipment (PPE). The process of retrieving data from social media videos involves extracting image frames containing PPE, so that they can be used as sample images for model training. Second, data from open source sources from kaggle that are already available is also utilized. These open source dataset usually already have good annotations and quality, so they can enrich data variations and improve the generalization ability of the model. The datasets from these two sources are then combined into one complete dataset. The amount of data collected at this stage is 4142 images that have been separated from the video. To be relevant and of high quality, the selected videos follow the following criteria:

a) The videos clearly show construction activities or work environments.
b) The videos show workers with or without PPE, to allow classification of PPE completeness.
c) The resolution must be at least 720p so that objects such as helmets, vests, and safety shoes can be visually recognized.
d) Videos have various camera angles (top view, frontal, side) and natural or artificial lighting are prioritized to increase data diversity.
e) The videos must be at least 10 seconds long so that enough frames can be captured.

After the videos are collected, the frame extraction process is carried out using video processing tools, which is OpenCV. Frames are extracted at intervals of every 0.5 seconds or 2 fps, to avoid taking consecutive frames that are too similar or redundant. After extraction, a manual frame selection process is carried out by considering the clarity of the main object to be detected, the diversity of contexts, and class distribution.

### 3.2 Dataset Annotation and Collection

At this stage, the dataset obtained from a collection of various videos on social media is further processed by annotating and labeling using the Roboflow tool. The annotation process is carried out frame-by-frame, namely each image frame extracted from the video is annotated by creating a bounding box around the detected personal protective equipment (PPE) object. Each bounding box is then labeled according to the type of PPE, such as Glove, Helmet, No Helmet and Vest. With thorough and consistent annotations, models can learn to recognize and differentiate between different types of PPE with greater accuracy during the training process. The annotation and labeling process is carried out by one annotator with a basic understanding of personal protective equipment

 ISSN:2302-8149
e-ISSN:2540-9719

classification. To ensure label accuracy and consistency, all annotation results are then validated by an expert who has expertise in Occupational Safety and Health (OHS). This validation includes reviewing the bounding box and the suitability of the object class to the context of PPE use in the construction field.

### 3.3 Data Augmentation

After the data is labeled accordingly, the annotated image data is further processed to be ready for use in model training. One of the important processes in this stage is image augmentation. Augmentation aims to increase the variety of data by modifying the original image without changing its label [21], so that the model can learn from a wider range of conditions and variations. Some of the augmentation techniques used in this study are image rotation with a certain image angle, horizontal and vertical flip, and changes in brightness and contrast in the image. By performing data augmentation, deep learning models such as YOLOv10 and YOLOv11 can become more robust and resistant to real variations in the field, such as different viewing angles, changing lighting, or varying object positions. Image rotation is performed randomly with a rotation angle between -15 to +15 degrees. This limit is chosen to keep the object structure realistic and recognizable by the model, considering that extreme rotation can cause shape distortion that is not representative of the actual conditions. Furthermore, the image is flipped horizontally and vertically with a probability of 0.5 each. This augmentation helps the model learn to recognize objects even when they appear in different orientations. Additionally, lighting level adjustments were applied to strengthen the model against lighting variations common in open fields. The brightness and contrast of the image are changed randomly within a range of ±20% of the original value. In addition, the amount of data after the augmentation process is 16,568 images.

### 3.4 Data Splitting

After the data has been processed and annotated, the dataset is then divided into three main parts with the following proportions:

a) Training Data (80%): The largest part of this dataset is used to train the model. This training data functions so that the model can learn to recognize patterns and features of various types of personal protective equipment (PPE).
b) Validation Data (15%): This data division is used during the training process to monitor model performance periodically. With this data, hyperparameter adjustments and overfitting prevention are carried out so that the model not only memorizes the training data, but is also able to generalize to new data.
c) Test Data (5%): This data is used to test the final performance of the model after the training process is complete. This data is not used during training so that it can provide an objective picture of the model's ability to detect PPE on new data that has never been seen before.

The aim of dividing the data in this proportion is to ensure the model gets enough data to learn, as well as being tested and validated effectively so that PPE detection results are accurate and reliable.

### 3.5 YOLO Fine-Tuning

Fine tuning is a deep learning model training technique where a pre-trained model is used as a starting point, then retrained (tuned) on a new dataset specific to a particular task [22]. This approach is very effective in saving training time and improving model performance, especially when the new dataset is relatively small or similar to the dataset the pre-trained model was originally trained on. The fine-tuning technique used in this study is transfer learning. This technique is part of transfer learning, which utilizes knowledge that has been learned by previous models for new tasks [23]. The pre-trained model's weights are adjusted (retrained) with new data so that the model can recognize specific objects or patterns in the research dataset. In this study, personal protective equipment detection uses You Only Look Once (YOLO) models, which are YOLOv10 and YOLOv11 models. Fine tuning allows models that have been trained on general datasets to recognize PPE objects specifically. This process involves adjusting the model's weights so that it can optimally detect glove, helmet, no helmet, and vest classes according to the characteristics of the data collected from social media videos and open source datasets.

 ISSN:2302-8149
e-ISSN:2540-9719

In the fine-tuning process, all weights of the model are unfrozen so that they can be updated during training. This means that no layers are frozen; all layers of the model are retrained to adjust to the distribution and unique features of the APD dataset used. This approach was chosen because the baseline models, YOLOv10 and YOLOv11, were trained on different domains, so full fine-tuning is considered more effective for adaptation to specific construction domains. The training process was carried out using the AdamW optimizer which is known to be able to provide stable and effective weight updates, especially on large-scale models. The learning rate was set at 0.001 to maintain stability during training while ensuring an optimal convergence rate. In addition, a momentum value of 0.9 was used to help speed up the training process by reducing gradient oscillations. The number of epochs used in training was 100, so that the model had enough iterations to learn the relevant feature representations from the dataset. Meanwhile, the batch size used was 32, as a compromise between computational efficiency and stability in gradient calculations.

## A. YOLOv10

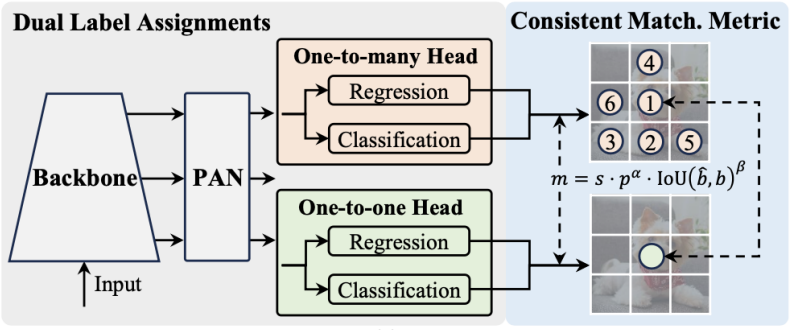


**Figure 2. Architecture of YOLOv10**[24]

YOLOv10 is one of the variants of the YOLO (You Only Look Once) family designed for real-time object detection with better efficiency and accuracy than previous versions. YOLOv10 combines the latest techniques in network architecture and optimization to improve detection performance, especially on resource-constrained devices. The main architecture of YOLOv10 can be shown in figure 2. YOLOv10 adopts a modular structure consisting of three main components[24] :

a) Backbone: This components is responsible for extracting important features from the input image. YOLOv10 uses a lightweight and efficient backbone, such as a variant of CSPDarknet that has been optimized for speed and accuracy. This backbone is capable of capturing features at various scales, from coarse features to fine details.
b) Neck: It functions to combine and amplify features from various resolution levels produced by the backbone. YOLOv10 uses techniques such as PANet (Path Aggregation Network) which helps the model understand spatial context and feature hierarchy, so that object detection becomes more accurate, especially for objects of different sizes.
c) Head: This part that makes the final prediction in the form of bounding boxes, confidence scores, and object classification. YOLOv10 optimizes the head to produce precise and fast predictions, using customized anchor boxes and improved loss functions.

In this study, the YOLOv10 architecture used is the small (YOLOv10s) and nano (YOLOv10n) versions. YOLOv10n is the lightest variant of YOLOv10, designed for devices with limited resources such as mobile devices or embedded systems. While YOLOv10s has a more complex architecture than YOLOv10n, with more layers and parameters.

## B. YOLOv11

YOLOv11 is the latest iteration of the YOLO series, bringing a number of architectural and performance improvements over previous versions. It is designed to improve object detection accuracy while maintaining high inference speed, making it ideal for real-time applications such as personal protective equipment detection. The main architectural of YOLOv11 can be shown in figure 3 which the components of the model are[25]:

 ISSN:2302-8149
e-ISSN:2540-9719

a) C3k2 Block (Cross Stage Partial with kernel size 2): A convolution block optimized for more efficient feature extraction with a smaller kernel size, reducing computational complexity without sacrificing feature quality.
b) SPPF (Spatial Pyramid Pooling - Fast): A pooling technique that enhances the model's ability to capture features at various object scales in a faster and more efficient way than traditional pooling.
c) C2PSA (Convolutional block with Parallel Spatial Attention): A parallel spatial attention module that helps the model focus on important parts of the image, improving detection capabilities especially for difficult or overlapping objects.

The YOLOv11 architecture used is also a small version (YOLOv11s) and nano (YOLOv11n). The YOLOv11n architecture uses a simpler network with fewer parameters, so the training and inference process becomes more efficient. While YOLOv11s has a small variant that has a more complex architecture than nano, with a larger network capacity.

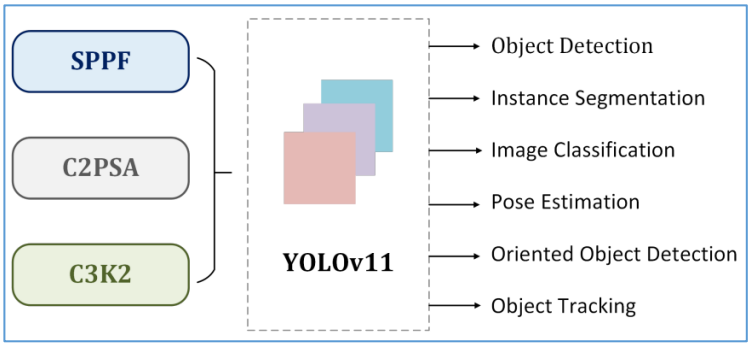


**Figure 3 Architecture of YOLOv11**[25]

### 3.6 Model Evaluation

After the model has been trained, the evaluation stage is carried out to measure how well the model detects objects in data that has never been seen before (test data). This evaluation is important to ensure the model can perform accurately and reliably in real-world conditions. The metrics used in the evaluation stage are precision, recall, and mean average precision (mAP50) [26]:

a) Precision: a metric that measures how accurate the positive predictions made by a model. In the context of object detection, precision indicates the proportion of true positive bounding box predictions out of all predictions made by the model.
b) Recall: a metric that measures the ability of a model to find all objects that are actually present in the data. In object detection, recall indicates the proportion of objects that are successfully detected (true positives) out of all objects that should have been detected.
c) mAP50: a key metric in object detection evaluation that measures the average precision at an Intersection over Union (IoU) threshold of 0.5. IoU is a measure of the overlap between the prediction bounding box and the ground truth bounding box. If IoU $\geq$ 0.5, the prediction is considered correct (True Positive). mAP50 is calculated by taking the average of the Average Precision (AP) for each object class at an IoU threshold of 0.5. This metric provides an overall picture of the model's performance in detecting and classifying objects with a fairly tight location accuracy.

## 4 Results and Analysis

The evaluation results using precision, recall, and map50 metrics can be seen in figures 4, 5, and 6. Based on the evaluation results using precision metrics as shown in figure 4, YOLOv10n has the lowest precision of 0.709. This model is the lightest and fastest, so it is possible to have slightly lower precision. YOLOv10s shows an increase in precision to 0.731, indicating that this model is more accurate in predicting objects than the nano variant. YOLOv11n and YOLO v11s have precisions of 0.740 and 0.804. Based on these results, it can be concluded that models with larger architectures and sizes (small) tend to have higher precision than the nano variant, because the larger model capacity allows for more complex and accurate feature learning.

 ISSN:2302-8149
e-ISSN:2540-9719

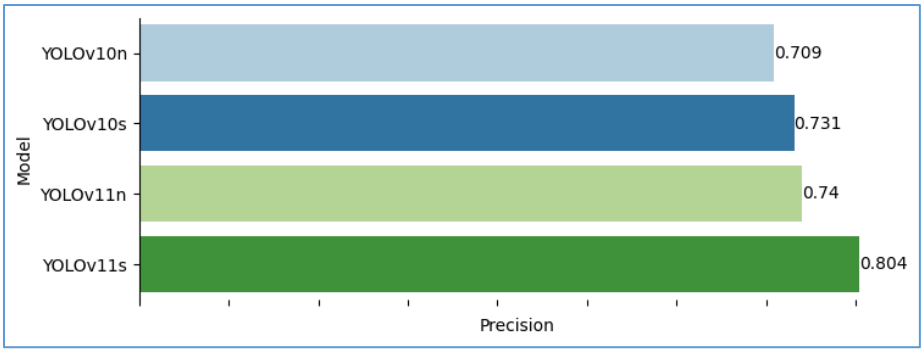


**Figure 4. Evaluation results using precision metrics**

Based on the evaluation results using the recall metric as seen in Figure 5, YOLOv10n has the lowest recall value of 0.671, meaning that this model successfully detects around 67.1% of the total objects that are actually in the test data. YOLOv10s is slightly better with a recall of 0.682, indicating a better ability to find existing objects. YOLOv11n has the highest recall value of 0.684, indicating that this model is most effective in detecting objects that are actually in the dataset. Meanwhile, YOLOv11s with a larger architecture size than YOLOv11n has the highest precision (0.804), but also has the lowest recall (0.669). This may be because large models such as YOLOv11s tend to produce predictions with higher confidence, but narrower distributions, so the model only detects predictions with very high confidence. This evaluation results indicate a natural trade-off between precision and recall. More selective models such as YOLOv11s produce highly accurate predictions but miss some objects due to low recall, while more permissive models such as YOLOv11n are able to capture more objects but with a slight decrease in precision. For the case of detecting the completeness of PPE in construction areas, a model with high recall such as YOLOv11n is more suitable, because it is able to detect more violations that actually occur, which is crucial for occupational safety. However, precision is still important, and the system should not ignore real violations to avoid false positives.

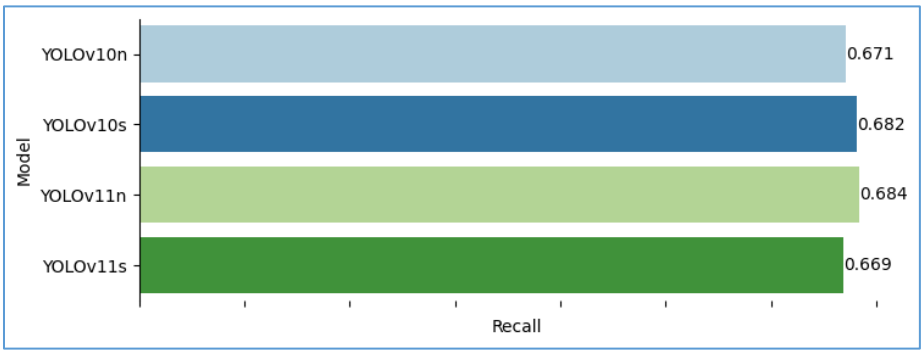


**Figure 5. Evaluation results using recall metrics**

Based on the evaluation results using the map50 metric which can be seen in figure 6, YOLOv10n has the lowest mAP50 value of 0.691, which means that this model has the lowest object detection performance compared to other variants in terms of the balance between precision and recall at the IoU threshold of 0.5. YOLOv10s is slightly better with an mAP50 value of 0.698, indicating an increase in overall object detection capabilities. YOLOv11n shows a more significant increase with an mAP50 value of 0.711, indicating that this model is more accurate in detecting and classifying objects. YOLOv11s has the highest mAP50 value of 0.718, indicating the best performance among the four models. This model is able to detect objects with better accuracy and consistency. So it can be concluded that the higher mAP50 value in YOLOv11s indicates that this model is the most effective in detecting personal protective equipment (PPE) in the dataset used. This means that YOLOv11s is able to provide more precise bounding box predictions and more accurate object classifications compared to other variants. Thus, the use of YOLOv11s in this study can provide more reliable and dependable detection results for real applications.

 ISSN:2302-8149
e-ISSN:2540-9719

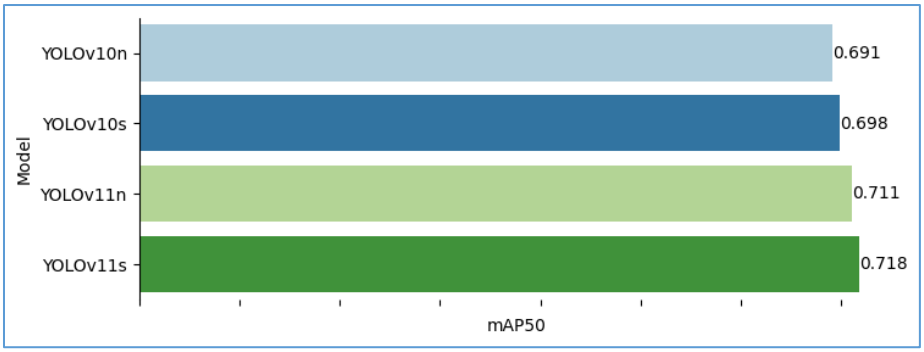


**Figure 6. Evaluation results using mAP50 metrics**

In addition, based on the overall evaluation results, the model with the YOLOv11s architecture has the best evaluation results compared to other architectures. This can be seen from the best results of precision and mAP evaluation obtained by the YOLOv11s architecture. However, this architecture has the lowest recall value compared to other architectures. Although the recall of YOLOv11s is lower, the high precision can compensate in the calculation of mAP50. This shows that most of the predictions made are correct and with high confidence. In conclusion, YOLOv11s is very good at making accurate predictions, but tends to be conservative—only detecting objects if it is absolutely sure. This can be beneficial for applications that require high accuracy and low tolerance for detection errors, such as occupational safety monitoring (PPE detection). However, it can be less than optimal if the main goal is to detect as many objects as possible, for example in a search or rescue system.

(a) (b)

**Figure 7. Examples of PPE detection results using the YOLOv11s architecture: (a) safety helmet and vest detection; (b) gloves, safety helmet and vest detection**

Figure 7 shows an example of detection results using the model architecture that has the best evaluation, namely YOLOv11s. In figure 7 (a), object detection successfully identified the personal protective equipment used by the two people, namely helmets and safety vests, with a fairly high level of confidence, around 78% to 85%. Then in figure 7 (b), object detection successfully identified the personal protective equipment used by the two individuals, namely helmets, safety vests, and gloves, with a very high level of confidence, around 71% to 93%. From these two examples of detection results, it can be concluded that the deep learning model used in this detection shows very good performance in identifying various types of personal protective equipment (PPE) such as helmets, safety vests (vests), and gloves (gloves) in the images provided.

## 5 Conclusion

This research succeeded in developing a personal protective equipment (PPE) detection model in a construction environment using a CNN architecture with a fine-tuning approach on two YOLO variants, namely YOLOv10 and YOLOv11. Based on the evaluation results using precision, recall, and mAP50 metrics, the YOLOv11s model showed the best performance with the highest mAP50 value of 0.718, indicating a more accurate and consistent object detection and classification capability

 ISSN:2302-8149
e-ISSN:2540-9719

compared to other variants. Although YOLOv11n has the highest recall value, this model must be balanced with precision so as not to produce too many false positives. These results indicate that the increased model capacity and more complex architecture in YOLOv11s significantly contribute to improving the accuracy of PPE detection, so this model has great potential to be applied in automatic occupational safety monitoring in the construction field. In further research, it is recommended to develop models with more diverse datasets and more complex environmental scenarios, including lighting conditions and varying shooting angles. In addition, the integration of this detection system with real-time hardware such as CCTV cameras and automatic warning systems can be an important step for practical implementation in the field.

## Acknowledgement

The authors would like to thank the Direktorat Penelitian dan Pengabdian Masyarakat Universitas Islam Riau (DPPM UIR) for providing financial support so that this research can be conducted well.

## Reference

[1] J. H. Lo, L. K. Lin, and C. C. Hung, "*Real-Time Personal Protective Equipment Compliance Detection based on Deep Learning Algorithm*," *Sustainability (Switzerland)*, Vol. 15, No. 1, Jan. 2023, doi: 10.3390/su15010391.

[2] M. I. B. Ahmed *et al.*, "*Personal Protective Equipment Detection: A Deep-Learning-based Sustainable Approach*," *Sustainability (Switzerland)*, Vol. 15, No. 18, Sep. 2023, doi: 10.3390/su151813990.

[3] 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*, 2023. doi: 10.1109/NMITCON58196.2023.10275958.

[4] N. D. T. Yung, W. K. Wong, F. H. Juwono, and Z. A. Sim, "*Safety Helmet Detection using Deep Learning: Implementation and Comparative Study using YOLOv5, YOLOv6, and YOLOv7*," in *2022 International Conference on Green Energy, Computing and Sustainable Technology, GECOST 2022*, 2022. doi: 10.1109/GECOST55694.2022.10010490.

[5] Y. C. How, A. F. Ab. Nasir, K. F. Muhammad, A. P. P. Abdul Majeed, M. A. Mohd Razman, and M. A. Zakaria, "*Glove Defect Detection Via YOLO V5*," *MEKATRONIKA*, Vol. 3, No. 2, 2022, doi: 10.15282/mekatronika.v3i2.7342.

[6] A. K. Al Bari, E. Rachmawati, and G. Kosala, "*Glove Detection System on Laboratory Members using Yolov4*," *Journal of Information System Research (JOSH)*, Vol. 4, No. 4, 2023, doi: 10.47065/josh.v4i4.3806.

[7] V. Isailovic *et al.*, "*Compliance of Head-Mounted Personal Protective Equipment by using YOLOv5 Object Detector*," in *International Conference on Electrical, Computer, and Energy Technologies, ICECET 2021*, 2021. doi: 10.1109/ICECET52533.2021.9698662.

[8] D. Arya, A. Prayoga, M. Taufiqurrochman, A. A. Zein, and M. Khanif, "Deteksi *Helm Safety* menggunakan *Convolutional Neural Networks* dan *YOLOv3-Tiny*," Vol. 4, No. 1, pp. 1276–1286, Dec. 2024, doi: https://doi.org/10.20895/centive.v4i1.299.

[9] Y. Liu and W. Jiang, "*Detection of Wearing Safety Helmet for Workers based on YOLOv4*," in *Proceedings - 2021 International Conference on Computer Engineering and Artificial Intelligence, ICCEAI 2021*, 2021. doi: 10.1109/ICCEAI52939.2021.00016.

[10] F. Zhou, H. Zhao, and Z. Nie, "*Safety Helmet Detection based on YOLOv5*," in *Proceedings of 2021 IEEE International Conference on Power Electronics, Computer Applications, ICPECA 2021*, 2021. doi: 10.1109/ICPECA51329.2021.9362711.

[11] Z. P. Xu, Y. Zhang, J. Cheng, and G. Ge, "*Safety Helmet Wearing Detection based on YOLOv5 of Attention Mechanism*," in *Journal of Physics: Conference Series*, 2022. doi: 10.1088/1742-6596/2213/1/012038.

[12] Hadi Supriyanto, Sarosa Castrena Abadi, and Aliffa Shalsabilah, "Deteksi Helm Keselamatan menggunakan *Jetson Nano* dan *YOLOv7*," *Journal of Applied Computer Science and Technology*, Vol. 5, No. 1, 2024, doi: 10.52158/jacost.v5i1.637.

 ISSN:2302-8149
e-ISSN:2540-9719

[13] M. Gugssa, A. Gurbuz, J. Wang, J. Ma, and J. Bourgouin, "PPE-Glove *Detection for Construction Safety Enhancement based on Transfer Learning*," in *Computing in Civil Engineering 2021 - Selected Papers from the ASCE International Conference on Computing in Civil Engineering 2021*, 2021. doi: 10.1061/9780784483893.008.
[14] M. Ulum, M. Zakariya, A. Fiqhi, and H. Haryanto, "Rancang Sistem Pendeteksi Alat Pelindung Diri (APD) berbasis *Image Prosessing*," *Informatics, Electrical and Electronics Engineering (Infotron)*, Vol. 1, No. 1, 2021, doi: 10.33474/infotron.v1i1.11236.
[15] I. A. Rachimi and F. Utaminingrum, "Deteksi Masker dan Suhu Tubuh untuk Kendali Portal Otomatis menggunakan CNN sebagai Pencegahan Penularan SARS-CoV-2," 2021. [Online]. Available: http://j-ptiik.ub.ac.id
[16] M. L. R. Collo, J. Richard M. Esguerra, R. V. Sevilla, J. Merin, and D. C. Malunao, "*A COVID-19 Safety Monitoring System: Personal Protective Equipment (PPE) Detection using Deep Learning*," in *2022 International Conference on Decision Aid Sciences and Applications, DASA 2022*, 2022. doi: 10.1109/DASA54658.2022.9765088.
[17] S. Khosravipour, E. Taghvaei, and N. M. Charkari, "*COVID-19 Personal Protective Equipment Detection using Real-Time Deep Learning Methods*," Mar. 2021, [Online]. Available: http://arxiv.org/abs/2103.14878
[18] M. Reza Kusuma, A. Azis Abdillah, D. Junaedi, D. Yanti Liliana, S. Arifin, and dan Zahran Muzakki, "Sistem Deteksi Alat Pelindung Diri di Workshop Alat Berat Politeknik Negeri Jakarta menggunakan *Teachable Machine*," 2022. Accessed: May 20, 2025. [Online]. Available: https://prosiding.pnj.ac.id/sntm/article/view/517/365
[19] Z. Syifa Juanda and Z. Khairil Simbolon, "Mengintegrasikan Metode YOLO (You Only Look Once) dalam Deteksi APD (Alat Pelindung Diri) pada Industri Migas," Vol. 4, No. 1, 2024.
[20] A. Sesis *et al.*, "*A Robust Deep Learning Architecture for FireFighter PPEs Detection*," in *2022 IEEE 8th World Forum on Internet of Things, WF-IoT 2022*, 2022. doi: 10.1109/WF-IoT54382.2022.10152263.
[21] A. Mumuni and F. Mumuni, "Data *Augmentation: A Comprehensive Survey of Modern Approaches*," *Array*, Vol. 16, Dec. 2022, doi: 10.1016/j.array.2022.100258.
[22] K. W. Church, Z. Chen, and Y. Ma, "*Emerging Trends: A Gentle Introduction to Fine-Tuning*," *Nat Lang Eng*, Vol. 27, No. 6, pp. 763–778, Nov. 2021, doi: 10.1017/S1351324921000322.
[23] S. Panigrahi, A. Nanda, and T. Swarnkar, "*A Survey on Transfer Learning," in Smart Innovation, Systems and Technologies, Springer Science and Business Media Deutschland GmbH*, 2021, pp. 781–789. doi: 10.1007/978-981-15-5971-6_83.
[24] A. Wang *et al.*, "*YOLOv10: Real-Time End-to-End Object Detection*," May 2024, [Online]. Available: http://arxiv.org/abs/2405.14458
[25] R. Khanam and M. Hussain, "YOLOv11: *An Overview of the Key Architectural Enhancements*," Oct. 2024, [Online]. Available: http://arxiv.org/abs/2410.17725
[26] P. Fränti and R. Mariescu-Istodor, "*Soft Precision and R*ecall," *Pattern Recognit Lett*, vol. 167, 2023, doi: 10.1016/j.patrec.2023.02.005.