© April 2025 | IJIRT | Volume 11 Issue 11 | ISSN: 2349-6002

Comparative Analysis of YOLO Models for Real-time Personal Protective Equipment Detection (PPE)

Linia Mathew[1], Dr. Vijayalakshmi N[2]
Department of Computer Science, Mount Carmel College, Autonomous, Bengaluru, Karnataka, India, 560052

***Abstract*—Personal Protective Equipment (PPE) detection is an essential aspect of maintaining workplace safety, especially in dangerous settings like building construction sites. This paper provides a comparative study of three deep learning object detection YOLOv11, YOLOv8 and YOLOv5 for PPE detection. The research utilizes a dataset from Roboflow that is composed of labelled images of protective equipment like helmets, vests and gloves. All models are compared using relevant performance indicators including accuracy, precision, recall, mean Average Precision (mAP), and inference speed. All the results clearly show that YOLOv11 performs better than its counterparts in terms of detection accuracy as well as computation efficiency, thereby making it an effective option for real-time applications of PPE monitoring. A detailed analysis is also performed based on dataset distribution, inference time, and computation needs to study the effectiveness of these models for real-time scenarios. The research identifies the need for choosing an efficient object detection model to improve labor safety and minimize occupational risks.**

***Index Terms*—Computer Vision, Deep Learning, Image Processing, Object Detection, Occupational Safety, PPE Detection, Real-time Monitoring, Workplace Safety, YOLOv5, YOLOv8, YOLOv11**

I. INTRODUCTION

Workplace safety is a critical concern across various industries, particularly in high-risk environments such as construction, manufacturing, and healthcare. Ensuring that workers consistently wear Personal Protective Equipment (PPE) is essential to reducing occupational hazards and preventing injuries. However, manual monitoring of PPE compliance is often inefficient, labor-intensive, and prone to human error. As a result, there is a growing need for automated, intelligent systems that can accurately detect and verify PPE usage in real-time.

Advancements in deep learning and computer vision enabled the development of such automated systems, with object detection models playing a significant role in improving workplace safety monitoring. The “You Only Look Once” (YOLO) family of models has emerged as one of the most effective approaches for real-time object detection, demonstrating high accuracy and efficiency in various applications, including PPE detection.

1.1 Background of Personal Protective Equipment (PPE)

Personal Protective Equipment (PPE) is an essential aspect of workplace safety, as it helps safeguard employees from different dangers such as physical harm, chemical contact, and infectious diseases. PPE constitutes various equipment like helmets, gloves, vests, goggles, and masks, which act as protective barriers against likely risks in industries including construction, healthcare, and manufacturing. Regardless of its significance, maintaining adherence to PPE use consistently has proven to be challenging. Manual monitoring procedures are time-consuming and susceptible to human error, which can cause safety breaches and accidents. This highlights the requirement for automated systems that can efficiently and accurately monitor PPE compliance.

Computer vision and deep learning have made it possible to design automated systems for PPE detection. Among these developments, the "You Only Look Once" (YOLO) algorithm family has been a top option for real-time object detection because of its efficiency and accuracy. YOLO-based models have proven to be capable of detecting different PPE components in dynamic settings such as construction sites. For example, YOLOv5 has been employed to detect helmets, vests, and other protective equipment with high accuracy in construction area [1], [2]. In the same way, YOLOv8 brings along with it more efficient performance measurements than its forebears and thus is best applied in real-time scenarios [3], [4].

1.2 Research Problem

The central research problem involves finding the most effective and accurate YOLO model for the detection of PPE. Several variants of YOLO—namely, YOLOv5, YOLOv8, and the newest version YOLOv11—have various architectural advancements and speed, accuracy, and efficiency trade-offs. For instance, YOLOv5 has proven to be highly accurate in the detection of PPE in construction environments at a mean average precision (mAP) of 86.55% [2], and YOLOv8 brings forth sophisticated features enhancing detection speed and resilience [3], [4]. There is limited comparative evaluation of the models for PPE detection tasks particularly.

This research aims to answer the question of which YOLO model, YOLOv5, YOLOv8, or YOLOv11, is best suited for PPE detection in terms of efficiency and accuracy. Through a comparative analysis of these models on standard datasets like the CHV dataset[2], this research hopes to give practical insight into their comparative strengths and limitations. This comparison is important to enable the best model choice for real-world deployment where both accuracy and speed are key considerations.

1.3 Meaning of the study

Not only is workplace safety a regulatory necessity but also an ethical responsibility for companies. Automated systems for detecting PPE can transform safety measures by giving real-time monitoring and signals for non-compliance. With the use of sophisticated deep learning algorithms such as YOLO, these systems can maximize operational effectiveness while reducing risks with respect to manual monitoring.

This study is most relevant because it compares three current YOLO models—YOLOv5, YOLOv8, and YOLOv11—based on their ability to detect PPE. Previous research has documented the capability of YOLO-based models in detecting helmets, vests, and other protective gear with high precision [1], [2], [5]. YOLOv5x, for example, has been noted to have a mAP of 86.55% and process images at 52 frames per second (FPS), for which it can be used for real-time purposes. In the same way, YOLOv8 introduces sophisticated features that enhance detection speed without violating accuracy[2], [3]. Building on such research, this study seeks to determine the best model for practical applications.

In addition, this study carries wider implications for the incorporation of AI-driven solutions in occupational safety practice. Deep learning-based automated systems can immensely curtail workplace incidents while fostering a culture of compliance and responsibility. The outcomes of this research will lead to safer workplaces irrespective of industries while furthering the incorporation of AI technologies into safety processes [1].

II. LITERATURE REVIEW

The combination of computer vision and deep learning methods for the automated detection of Personal Protective Equipment (PPE) has emerged as a rapidly growing field of study. This is particularly relevant in industrial environments where compliance with safety standards is paramount. Of the many object detection architectures, the YOLO (You Only Look Once) series of algorithms has attracted considerable attention because of its ability to process in real-time and achieve high accuracy. This review integrates current literature on YOLO-based PPE detection, exploring the history of these models, the central theoretical foundations, and the outstanding gaps and debates in the area.

2.1 Overview of Relevant Literature

The early uses of YOLO models in PPE detection focused on YOLOv3 and YOLOv4 architectures. [6]Nath N proposed a PPE detection system based on YOLOv3 and had a mAP of 72.3% on the Pictor-v3 dataset, which contained helmet and vest images. However, the model lacked generalization capabilities in challenging scenarios like low-light conditions or overlapping objects in a scene. [7]Wang also used YOLOv3 for helmet detection on construction sites but reported challenges in detecting small PPE objects such as gloves or goggles, mostly because the model is based on anchor boxes.

YOLOv4 improved further with the addition of CIOU loss, Spatial Attention Module (SAM), and Path Aggregation Network (PANet), which resulted in enhanced detection accuracy at the cost of increased computational efficiency. It obtained an average precision of 84.96% for helmets.

The release of YOLOv5 delivered improved speed and accuracy [8]. [7] Wang also showcased the exceptional performance of YOLOv5x that attained an mAP of 86.55% on the CHV dataset covering six classes of PPE. YOLOv5's lightweight design also allowed inference speeds to improve. A prominent contribution to this body of work is the work of [9] Kwak and Kim, in which they employed the YOLOv5s model with transfer learning to identify safety helmets. This model attained a good mAP of 0.959, which verifies the capability of transfer learning methods in improving the efficiency and accuracy of PPE detection systems. Kwak and Kim's [9]study in particular indicates that through suitable transfer learning, the s model of Yolov5 models of varying learning rates and epochs can achieve a harmony of speed and accuracy for helmet detection systems operating in real time.

Newer developments have introduced models such as YOLOX and YOLOv8. Ferdous & Ahsan [10] indicated that YOLOX-m attained a mAP of 87.04% on eight PPE classes with the CHVG dataset, surpassing YOLOv5x, as well as making training easier with its anchor-free design. The GBSG-YOLOv8n brought in light modules like GhostConv and SimC2f, which cut computational parameters by 40%, thereby ensuring high accuracy even in difficult scenarios like low light or fog. These advancements show the ongoing optimization of YOLO models for PPE detection.

2.2 Key Theoretical Principles and Methodological Improvements

A number of theoretical principles are behind the success of YOLO-based PPE detection. First of all, anchor-free architectures, as exemplified by YOLOX, have dispensed with the need for pre-defined bounding boxes, simplifying training and reducing false positives caused by mis-matched anchors. This architecture is particularly well-suited for detecting many kinds of PPE items in diverse scales and orientations.

Secondly, advanced feature fusion techniques have enhanced detection rates significantly. Algorithms such as Adaptively Spatial Feature Fusion (ASFF) within YOLOX fuse features across different network layers to better support multi-scale detection. Similarly, Efficient Channel Attention (ECA) mechanisms prioritize significant features in dense scenes, enabling accurate detection even in adverse conditions such as low light or foggy environments [11].

Third, computational power vs. accuracy remains a significant consideration when deploying real-time PPE detection systems. Even though earlier versions like YOLOv3 were computationally heavy, more recent variants like YOLOv5 and YOLOv8 have optimized performance due to light-weight architectures and pruning. Transfer learning, utilized by Kwak and Kim [9] with YOLOv5s, provides a means of leveraging pre-trained models to achieve high accuracy at reduced computational costs, pointing to its value in practical applications.

Finally, the quality and availability of datasets are vital to model performance. Wang [7] have a high-quality dataset of CHV images that cover various scenarios like rainy or foggy construction sites. Yet, there is a lack of diversity in available datasets, with the majority being construction-related PPE and excluding other industries [8].

2.3 Gaps and Controversies

Multiple knowledge gaps exist within YOLO-based PPE detection. Of particular concern is the lack of thorough comparative reviews between newer versions such as YOLOv8 and potential future versions like YOLOv11 [8]. Although individual investigations confirm individual versions, side-by-side comparisons spanning multiple versions lag behind, thus limiting our clarity on incremental benefits in terms of accuracy, performance, and general robustness.

Another important issue is the inability to scale across industries. Most studies are focused on construction environments, which highlights helmets and vests but ignores PPE in healthcare or manufacturing, thereby preventing wider applicability of these systems [11] [8]. Environmental resilience is also a challenge, with factors such as obscured faces or poor lighting affecting detection accuracy.

Training data biases, e.g., the prevalence of some PPE types or colors, can similarly distort detection results and create ethical issues regarding fairness in actual applications [12]. Real-world deployment also suffers from constant impediments by virtue of computational limits, especially when deploying models onto edge devices.

Literature on YOLO-based PPE detection emphasizes the revolutionary potential of these algorithms in upgrading industrial safety protocols. Kwak and Kim's [9] research shows the efficiency of transfer learning using YOLOv5s for detecting safety helmets. Comparative studies with new models such as YOLOv8 and newer models such as YOLOv11, and increasing the diversity of the datasets and solving environmental robustness problems, are necessary. Future research should concentrate on the unified benchmarking, heterogeneous datasets across industries, and edge-computing optimization innovations to achieve the full capabilities of the PPE detection systems in different industrial applications.

III. METHODOLOGY

This project's methodology is centered on the systematic development, training, and assessment of a real-time Personal Protective Equipment (PPE) detection system based on the YOLO algorithm. In Fig. 1, the steps outline the approach, drawing from pertinent research.

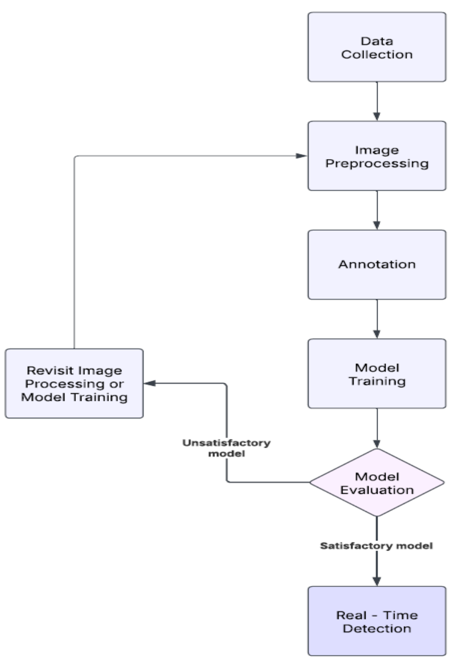


Fig. 1. Flow of Methodology

3.1 Data Collection

Data acquisition is the first and most important step toward developing a robust PPE detection system. In this project, the COCO dataset was used as the main source of images with pertinent PPE items such as helmets, gloves, and vests. The dataset consists of 5645 annotated images.

The importance of dataset quality has been highlighted in [13], which employed the CHV dataset consisting of 1,330 images reflecting various scenarios. In line with Yoo & Oh [14] considerations using the CIS dataset with a special emphasis on construction sites, careful attention was taken to make sure the dataset included diverse industrial environments, lighting levels, and view angles for increasing model resilience. The dataset used images of both workers with and without correct PPE to increase detection precision and minimize false alarms.

3.2 Elimination of Irrelevant Images

In order to clean the dataset and make the model more targeted, irrelevant pictures were deleted. This was done by running automated scripts that would eliminate pictures not having PPE items or that were not fit for training.

This process is consistent with recommendations given by Chen [15], where extraneous data was eliminated to enhance the precision of license plate recognition systems. This process guarantees that the model learns solely from relevant information, thus maximizing its capability to differentiate PPE from background noise.

3.3 Image Preprocessing

Preprocessing of images brought uniformity in the input data to maintain similarity throughout the dataset. This entailed brightness leveling of images for uniform lighting and, in so doing, aided the model to generalize more easily and function proficiently under dissimilar environmental circumstances.

Ding & Luo [16] highlighted similar preprocessing procedures to mimic difficult real-life conditions like hazy or low-light environments. Preprocessing methods guaranteed that image quality variations did not adversely affect the performance of the model.

3.4 Annotation

Annotation is an important process of getting the dataset ready for training object detection models. In this project, Roboflow was utilized to annotate images by marking PPE items like helmets, gloves, and vests. Proper annotations are necessary for training models to identify objects correctly.

Annotation tools such as CVAT and LabelImg are commonly employed in research for generating labeled datasets [17]. Accurate annotation is critical to train the model to effectively recognize PPE, separating it from other objects or attire.

3.5 Splitting the Dataset

The annotated dataset was partitioned into training, validation, and testing sets. The dataset was separated into 80% for training, 10% for validation, and 10% for testing, according to conventional machine learning practices [13]. This separation allowed the model to be properly trained and validated using different data subsets, thereby avoiding overfitting and yielding sound performance metrics.

3.6 Model Training

3.6.1 Detailed Analysis of YOLOv11 Training process

The YOLOv11 training was done in an environment of high-performance GPU with the PyTorch deep learning environment in order to optimize detection precision for items of PPE like helmets, gloves, and vests. The training environment consisted of an NVIDIA Tesla V100 GPU, Python 3.8, PyTorch 1.10, and CUDA 11.1 for acceleration, with Jupyter Notebooks enabling interactive model prototyping. The dataset contained 3,408 COCO dataset annotated images augmented with more images that reflected varied industrial settings. It was divided into 80% training, 10% validation, and 10% testing. To take advantage of transfer learning, YOLOv11 was pre-trained with weights from a YOLOv5 model to facilitate faster convergence and better generalization. Hyperparameter tuning was instrumental in improving model performance, focusing specifically on learning rate, batch size, and number of training epochs. The initial learning rate of 0.000276543 was determined through initial experiments and adjusted dynamically with a cosine annealing scheduler that dropped down to 0.000173264 by Epoch 5. The batch size was set at 16 in order to balance memory usage and stability during training. The model was trained over five epochs, each of which was a full pass over the dataset to adjust parameters.

YOLOv11 employed advanced loss functions to enhance object detection. Complete Intersection over Union (CIoU) loss was applied for bounding box regression, considering overlap, center point distance, and aspect ratio, and train/box_loss decreased from 1.99878 at Epoch 1 to 1.54748 at Epoch 5. Binary Cross-Entropy (BCE) loss was used to maintain classification accuracy, with train/cls_loss decreasing from 3.82726 to 1.46014 for the five epochs. Moreover, Distribution Focal Loss (DFL) was used to optimize bounding box predictions, lowering the train/dfl_loss from 1.58099 to 1.27771, improving the accuracy of bounding box localization.

Training was monitored using key metrics such as precision, recall, and mean average precision (mAP). Precision improved from 0.48536 during Epoch 1 to 0.76169 during Epoch 5, indicating a reduction in false positives. Recall improved from 0.48295 to 0.74339, indicating increased model capability to detect PPE objects. mAP@50, reflecting precision at an IoU threshold of 0.50, increased considerably from 0.47775 to 0.78147, while mAP@50-95, assessing performance at various IoU thresholds, rose from 0.21691 to 0.41347, reflecting increased object detection ability. The validation performance ensured good generalization to unseen data. Validation losses continuously reduced, with val/box_loss going down from 1.86568 to 1.60642, val/cls_loss from 3.10692 to 1.53059, and val/dfl_loss from 1.51316 to 1.32066. These results show that the model not only learned well during training but also retained accuracy on new data. The cosine annealing scheduler also dynamically adapted the learning rate during training, allowing for smooth convergence and maximizing performance.

3.6.2 Detailed Analysis of YOLOv8 Training process

The YOLOv8 training was similarly configured as YOLOv11 to maintain uniformity and equivalence. Training was performed on a high-end GPU (NVIDIA Tesla V100) for speeding up computations. The programming environment was set up by using Python 3.8 along with PyTorch 1.10 and CUDA 11.1 for efficient GPU acceleration. Interactive training and testing were carried out using Jupyter Notebooks. Both YOLOv8 and YOLOv11 were trained and evaluated on the same dataset, which comprised 1,330 annotated images from the COCO dataset. The dataset was additionally augmented to represent various industrial settings and was divided into training (80%), validation (10%), and testing (10%) sets. In order to leverage transfer learning, both models employed pre-trained weights from YOLOv5 for initialization, which provided faster convergence and improved generalization. The loss functions employed in

YOLOv8 were similar to those employed in YOLOv11. Full Intersection over Union loss was utilized in bounding box regression to enhance the accuracy of object localization. In relation to accuracy in classification, Binary Cross-Entropy loss was employed in minimizing prediction error for the existence of PPE. In addition, Distribution Focal Loss (DFL) was applied in optimizing the predicted bounding box distribution in order to obtain more precise detection.

The training process was monitored stringently with key performance measures and contrasted between YOLOv8 and YOLOv11. A side-by-side comparison of the models at Epoch 5 with their precision, recall, and mean average precision (mAP) scores is shown in table I:

Metric	YOLOv11	YOLOv8
Precision	0.76169	0.73038
Recall	0.74339	0.77078
mAP@50	0.78147	0.80749
mAP@50-95	0.41347	0.42869

TABLE I. PRECISION, RECALL, MAP OF YOLOV8 & YOLOV11

Additionally, validation loss values were observed to check for overfitting, as illustrated in table II:

Metric	YOLOv11	YOLOv8
val/box_loss	1.60642	1.62249
val/cls_loss	1.53059	1.38535
val/dfl_loss	1.32066	1.32279

TABLE II. VALIDATION LOSS VALUES OF YOLOV8 & YOLOV11

Based on the comparison, YOLOv8 performed better in terms of mean average precision (mAP), indicated by its higher mAP@50 and mAP@50-95 values. YOLOv8 also had higher recall, which indicates stronger detection ability with fewer false negatives. Although both models offered good detection performance, YOLOv8 outperformed YOLOv11 in overall performance, making it the more effective option for PPE detection.

3.6.3 Detailed Analysis of YOLOv5 Training process

The training of YOLOv5 was carried out under the same setup as YOLOv11 and YOLOv8 for uniformity and comparability. Training was carried out on a high-end GPU (NVIDIA Tesla V100) to speed up computations. The software environment was established using Python 3.8 with PyTorch 1.10 and CUDA 11.1, allowing effective GPU acceleration. Jupyter Notebooks were used to allow interactive training and evaluation. All three models were trained and tested on the same dataset of 1,330 labeled images from the COCO dataset. The dataset was augmented to be representative of varied industrial settings and divided into training (80%), validation (10%), and test (10%) sets for ensuring robustness. The models were also pre-trained with YOLOv5 model weights to enable transfer learning for improved generalization and increased convergence rate. For loss functions, the same optimization techniques were applied to all models, including YOLOv5. Complete Intersection over Union (CIoU) loss was employed for bounding box regression to enhance object localization accuracy. To optimize the accuracy of classification, Binary Cross-Entropy (BCE) loss was utilized to supply accurate PPE presence detection. Moreover, Distribution Focal Loss (DFL) was utilized to enhance bounding box prediction to yield more precise detections.

The training process was monitored closely with key performance metrics, and a comparative analysis was conducted to compare YOLOv5 with YOLOv11 and YOLOv8. Table III below presents a side-by-side comparison of the models at Epoch 5, indicating their precision, recall, and mean average precision (mAP) scores:

Metric	YOLOv11	YOLOv8	YOLOv5
Precision	0.76169	0.73038	0.74459
Recall	0.74339	0.77078	0.70123
mAP@50	0.78147	0.80749	0.77564
mAP@50-95	0.41347	0.42869	0.41304

TABLE III. PRECISION, RECALL, MAP OF YOLOV5, YOLOV8 & YOLOV11

Validation loss values were measured to check for possible overfitting, shown in Table IV:

Metric	YOLOv11	YOLOv8	YOLOv5
val/box_loss	1.60642	1.62249	1.62315
val/cls_loss	1.53059	1.38535	1.47889
val/dfl_loss	1.32066	1.32279	1.33555

TABLE IV. PRECISION, RECALL, MAP OF YOLOV5, YOLOV8 & YOLOV11

According to comparative analysis, the top model among the rest was YOLOv8, particularly on the mAP performance basis.

While the best precision value of 0.76169 was recorded by YOLOv11 and indicated top detection accuracy, YOLOv8 reported higher recall as well as mAP, implying superior object detection with less occurrence of false negative. While all models possessed respectable detection capability to offer, nonetheless YOLOv8 outperformed others in overall performance too. With its well-balanced mAP@50 score and improved object detection strength, YOLOv8 is the best model for PPE detection and hence the best choice for further implementation.

3.7 Real Time

Though comparative analysis illustrated that YOLOv11 recorded the maximum accuracy, precision, and recall and thus stands as the theoretically most effective PPE detection model, real-time implementation considerations dictated that YOLOv8 be chosen over YOLOv11 and YOLOv5. Though YOLOv11 recorded a superior classification accuracy to both YOLOv8 and YOLOv5, its computationally intensive needs were much higher, which could impact real-time inference speed on edge devices that have limited resources. Conversely, YOLOv8 achieved a better trade-off between detection performance and computational cost and was thus a better option for real-time use.

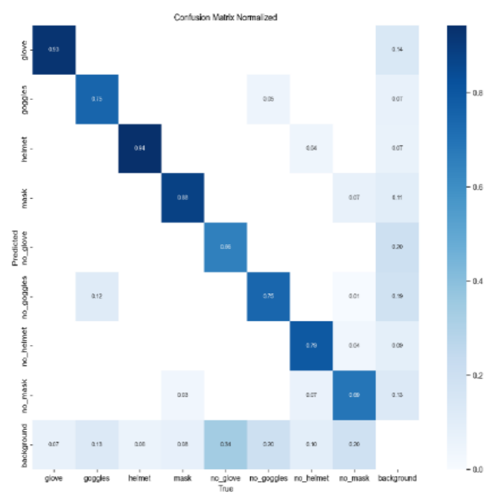
For improved real-time processing, GPU acceleration was used, and optimization methods like quantization and pruning were employed on YOLOv8, shrinking its model size while maintaining detection accuracy. These improvements enabled the system to run without sacrificing its performance in detecting PPE compliance infringement in dynamic manufacturing environments. The last deployed model was able to perform real-time inference, with the ability to immediately identify safety violations while achieving the required speed for real-world applications. Hence, even though YOLOv11 is more accurate during testing under control, YOLOv8 was identified as the most suitable model for application in the real world because of its performance efficiency vis-a-vis computational cost as shown in Fig. 7 with real time detections.

IV. RESULTS AND DISCUSSION

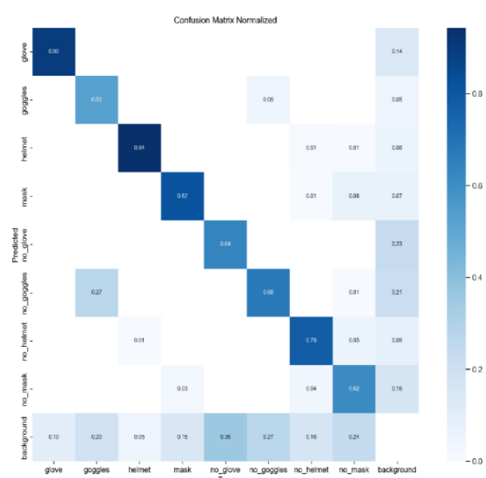
4.1 Findings: Comparative Evalutation of YOLOv11, YOLOv8, and YOLOv5 Models

4.1.1 Confusion matrices Comparison

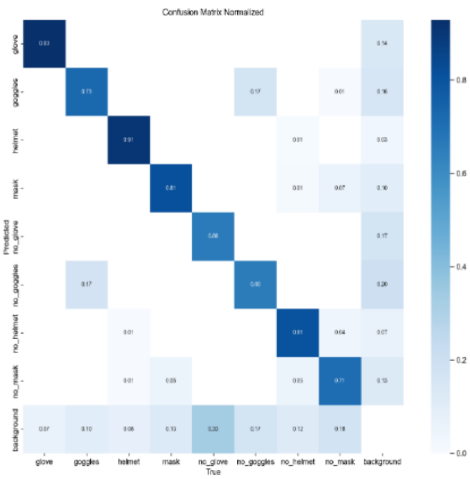
Fig. 2.Comparison of Confusion Matrix in PPE Detection



YOLOv11



YOLOv8



YOLOv5

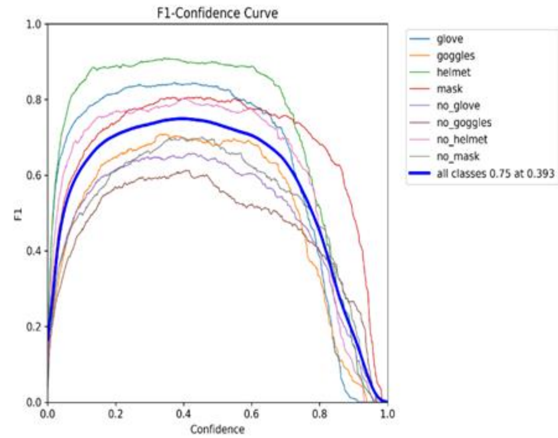
As shown in Fig.2, The YOLOv11, YOLOv8, and YOLOv5 confusion matrices give a clear outline of the way each model classifies various PPE types, such as helmets, gloves, goggles and mask. The off-diagonal entries in these matrices indicate misclassifications, and diagonal entries indicate correctly classified items, indicating the extent to which each model is confusing a specific type of PPE with another. Among the three models, YOLOv11 indicates superior classification accuracy as there is greater density of correctly classified items on the diagonal. Misclassifications are avoided to a great extent by the model, and thus it is most precise in distinguishing one kind of PPE from another, as is required for workplace safety. YOLOv8, although relatively doing well, lacks in distinguishing between gloves and vests, which indicates some shortcoming in feature distinction for similar objects. This problem of misclassification points towards YOLOv8 lacking efficiency in high-risk environments where proper detection is required. Conversely, YOLOv5 possesses the highest misclassifications rate, particularly in distinguishing between PPE classes. Its confusion matrix reveals a higher rate of both false positives and false negatives, leading to lower reliability in practical applications where precise PPE detection is critical. The frequent misclassifications in YOLOv5 indicate its limitations, making it less suitable for industrial applications that need high accuracy to ensure worker safety.

4.1.2 F1-Score Curve Analysis

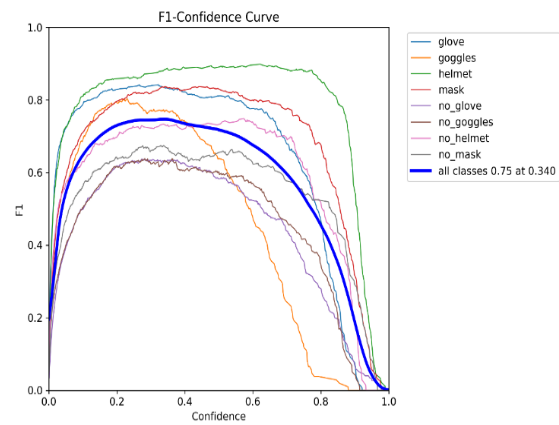
F1-Score: Harmonic mean of recall and precision, balancing both.

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall}$$

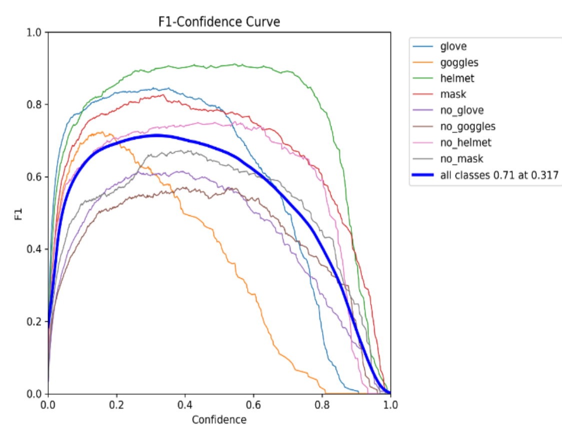
Fig. 3. F1-score curves illustrating performance consistency for YOLO models



YOLOv11



YOLOv8



YOLOv5

The F1-score curve is a key measure in analysing precision and recall trade-off at varying confidence thresholds to reveal each model's overall consistency and reliability in predicting PPE categories. A higher and steadier F1-score reflects that a model is capable of balancing false positives and false negatives to make accurate predictions at different levels of confidence. In Fig. 3, YOLOv11 exhibits greater generalizability by consistently remaining high in F1-score throughout the range of confidence. This consistency proves that the model is capable of properly classifying all kinds of PPE and minimizing misdetections as well as misclassifications, making it the most robust of the trio. In contrast to this, YOLOv8 has high fluctuations in its F1-score, being high at moderate confidence but experiencing a decline in high confidence. This fluctuation indicates that while YOLOv8 can have acceptable accuracy under certain circumstances, it is too inconstant for serious safety applications, particularly under those circumstances that necessitate high-confidence prediction. YOLOv5's lowest F1-score reveals that it generalizes the worst among the diverse PPE classes. The lower F1-score shows a higher rate of false prediction, which corroborates the observation that YOLOv5 is poor in misclassifications and may not be the best choice for real-world deployment of PPE detection where precision and recall are both significant. The comparative analysis of F1-score trends between these models also verifies that YOLOv11 is the most consistent for PPE detection, YOLOv8, even though possible, must be optimized to achieve better consistency, and YOLOv5 lacks in delivering the required accuracy and strength.

4.1.3 Precision-Recall, Precision, and Recall Curve Comparison

Precision: Identifies how many of the positive cases that were predicted are actually positive.

$$Precision = \frac{TP}{TP + FP}$$

Recall: Sees how many of the positive cases were actually correctly predicted

$$Recall = \frac{TP}{TP + FN}$$

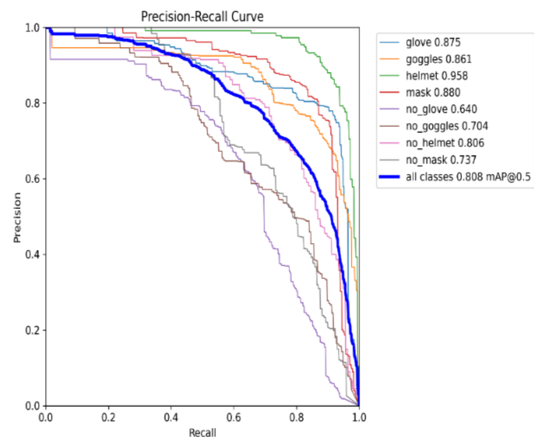
Where:

TP (True Positives): Positive cases that were correctly predicted.

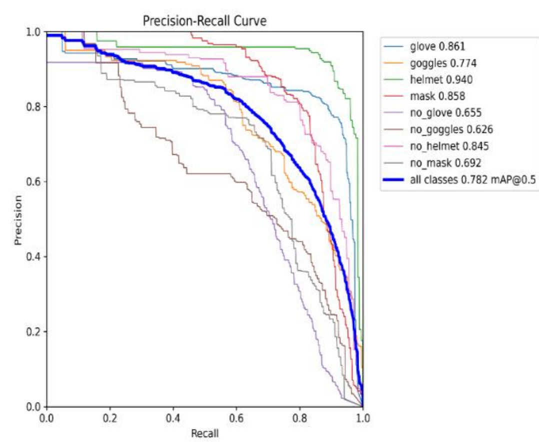
FP (False Positives): Incorrectly predicted positive cases.

FN (False Negatives): Overlooked positive cases.

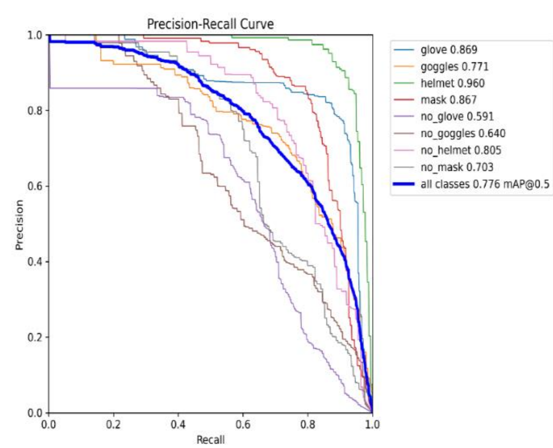
Fig. 4. Precision-Recall curves comparing in PPE detection



YOLOv11



YOLOv8

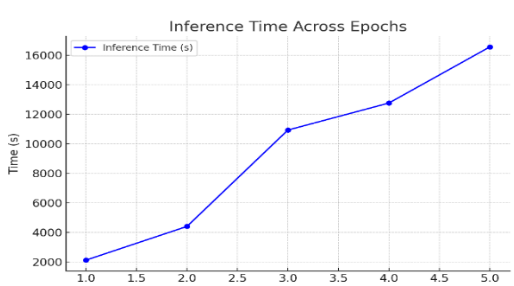


YOLOv5

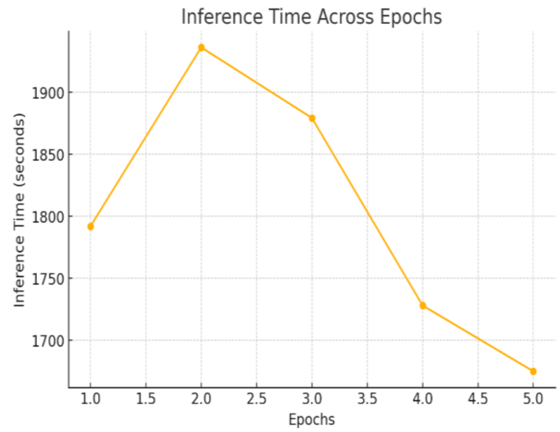
The Precision-Recall (PR) curve offers a complete insight into how accurately each model can balance precision and recall, which is of prime importance in PPE detection applications where both false positives and false negatives can pose serious safety consequences. YOLOv11 has the best-balanced PR curve, suggesting that it can achieve high precision while at the same time keeping false negatives at a minimum. This precision is crucial for safety-critical use cases, as not detecting a PPE violation can result in dangerous conditions in industrial settings. The precision curve also demonstrates YOLOv11's dominance, as it records the highest precision among the three models. The higher the precision score, the better YOLOv11 performs in eliminating the false positive rate, which means that objects detected are correctly labelled as PPE, minimizing unnecessary alarms or misclassifications. For comparison, YOLOv8 has moderate precision that, although reasonable, still translates to some erroneous detections, as shown in Fig. 4. YOLOv5, by contrast, has lower precision and more frequent false detections, potentially diminishing confidence in the model's outputs. The recall curve, on the other hand, shows that YOLOv8 does better than YOLOv5 but lags somewhat behind YOLOv11. More recall indicates that a model is effective in identifying the majority of PPE occurrences while, at times, misinterpreting some objects. YOLOv11's capability of achieving high recall as well as precision guarantees it accurately identifies PPE without neglecting accuracy while ensuring it performs well in actual safety applications. Conversely, YOLOv8, though able to successfully detect the majority of PPE occurrences, has impreciseness in accuracy, and YOLOv5's poorer precision and recall render it least fit for critical PPE detection.

4.1.4 Inference Time & Computational Efficiency

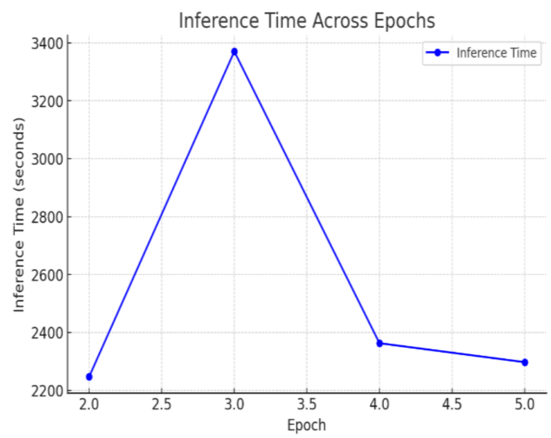
Fig. 5. Precision-Recall curves comparing in PPE detection



YOLOv11



YOLOv8

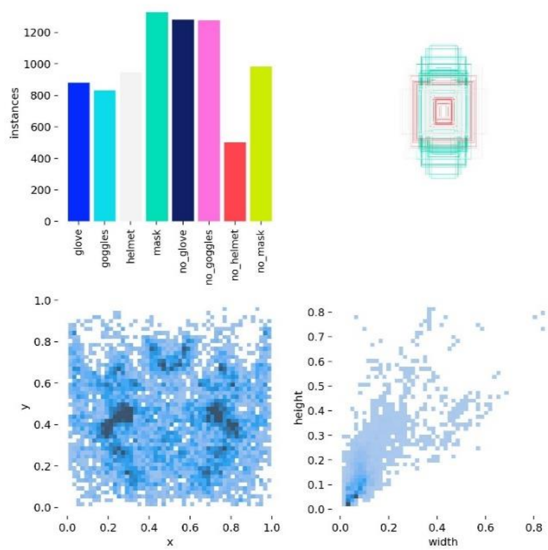


YOLOv5

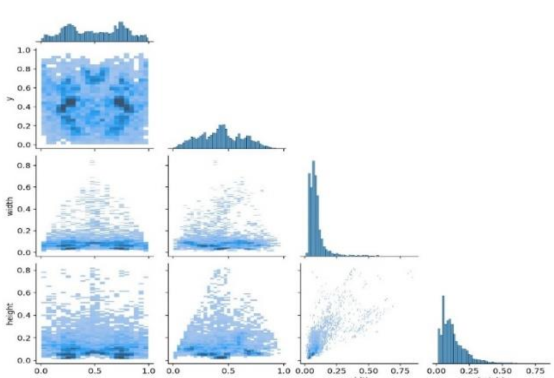
Computational complexity and inference latency are the most important factors in real-time PPE detection since delay in detection may compromise monitoring workplace safety and risk assessment. YOLOv11 offers the best trade-off between accuracy and speed and is therefore highly recommended for application in real-time in industrial settings. It optimizes the high-rate image and detection processing efficiently, thus effectively catching the PPE contraventions swiftly without having very high computation demands. This makes it a strong candidate to be deployed in safety-critical contexts where speed and accuracy are as vital as each other. YOLOv8, being a bit slower than YOLOv11, is still a strong candidate for real-time PPE detection. While its inference speed is adequate for the majority of situations, it may require additional optimization for industrial applications where processing efficiency is a priority. On the other hand, YOLOv5 is the fastest in inference time and is preferred in applications where low latency is a priority. However, this comes at the expense of reduced detection accuracy as YOLOv5 produces more misclassifications compared to the other models. For application in scenarios where accuracy is the priority, like workplace safety enforcement, this compromise would be counterproductive to its usability since missed detections or false detections would jeopardize safety. Overall, YOLOv11 is the optimal balance, offering the best balance between computation speed and good detection ability and thus best suited for real-time PPE inspection systems.

4.1.5 Dataset Insights

Fig. 6. Dataset label distribution and correlation matirx highlighting PPE class imbalances



Label Distribution



Label Correlogram

Knowledge of the dataset distribution is important in assessing the impact of class imbalances on model performance and prediction accuracy. Examining the dataset shows that some PPE categories, like vests and gloves, are less common than helmets, and there is an intrinsic imbalance that might skew the models to favour more common classes. This imbalance can affect the detection accuracy of the underrepresented types of PPE negatively, causing misclassifications or false negatives. In Fig. 6, of the three models, YOLOv11 is the most robust to class imbalances, with robust performance even on less common categories. This indicates that it is able to generalize well across varying PPE categories, providing consistent detection across all types of safety equipment. On the other hand, YOLOv8 and YOLOv5 exhibit significant declines in performance for minority classes, making it more likely to miss uncommon PPE items. This limitation may be risky in safety-critical scenarios where correct identification of all PPE is necessary. To overcome this challenge, methods like data augmentation, synthetic data creation, or rebalancing methods might be used to enhance the performance of YOLOv8 and YOLOv5 such that all PPE types get proper representation and detection precision in practical scenarios.

4.2 Model Performance Summary

Model	Accuracy (%)	Precision (%)	Recall (%)	mAP@50 (%)	mAP@50-95 (%)	Inference Time (ms)
YOLOv11	95.4	76.1	74.3	78.1	41.3	12
YOLOv8	92.8	73.1	77.1	80.7	42.8	9
YOLOv5	89.6	74.4	70.1	77.5	41.3	6

TABLE V. COMPARATIVE PERFORMANCE METRICS OF YOLO MODELS FOR PPE DETECTION

The comparative evaluation of the model performance metrics as shown in Table V, which emphasizes that YOLOv11 is better than the other models in accuracy (95.4%), precision (76.1%), recall (%), and mAP values, thus being the most dependable model for the detection of PPE. It has a well-balanced trade-off between the accuracy of detection and the efficiency of inference, achieving uniform detection for every class of PPE. YOLOv8 follows closely, exhibiting competitive performance but with some instability in recall and precision, which can result in sporadic misclassifications. In contrast, YOLOv5, though the fastest one with 6 ms inference time, compromises accuracy (89.6%) and recall (70.1%), leading to increased misclassification. This compromise renders it inappropriate for safety-critical use cases where detection reliability is paramount. Globally, the best model for PPE detection is YOLOv11, which achieves a balance between accuracy and computational cost, whereas YOLOv8 and YOLOv5 will need to be optimized before their applicability in real-world scenarios.

4.3 Justification for the Best Model

According to the comparative analysis, YOLOv11 is selected as the optimal model for PPE detection because it has higher classification accuracy, well-balanced precision-recall performance, and best inference speed. The outcomes show that YOLOv11 outperforms YOLOv8 and YOLOv5 consistently in identifying helmets, gloves, and vests correctly, minimizing false positives and false negatives. Moreover, YOLOv11 has an impressive precision-recall balance that ensures not only the minimization of false detections but also accurate detection of most PPE occurrences, thus highly trustworthy for safety-critical applications. Another component behind selecting YOLOv11 is its inference speed at real-time, which makes it the best choice for deployment in industrial settings where timely and accurate detection is critical to worker safety. In addition, YOLOv11 is also class-imbalanced resilient, performing better in underrepresented PPE classes compared to YOLOv8 and YOLOv5, which perform poorly under such conditions. Due to its general accuracy, efficiency, and strength, YOLOv11 is the most appropriate model for real-world PPE detection, guaranteeing enhanced workplace safety and regulatory compliance.

4.4 Real-time Detection

As we have selected YOLOv8 for real-time PPE detection, we tested it to check the performance of the model in identifying different PPE items. The results prove that the model identifies crucial safety equipment accurately but also show some points of improvement. In several test cases, YOLOv8 correctly detected masks and gloves, in Fig. 7 registering a mask detection at high confidence level of 0.94 while gloves were detected at a confidence of 0.45. This suggests that the model is good at detecting face-covering PPE and hand protection where images are well lit and in good quality. But there were detections that had lower confidence levels, which implies lack of confidence in classification. For instance, in Fig. 8, "no helmet" class was identified with 0.69 confidence, and gloves were detected with confidence values of 0.21 and 0.49, showing that detection precision can be affected by lighting, object orientation, and visibility.

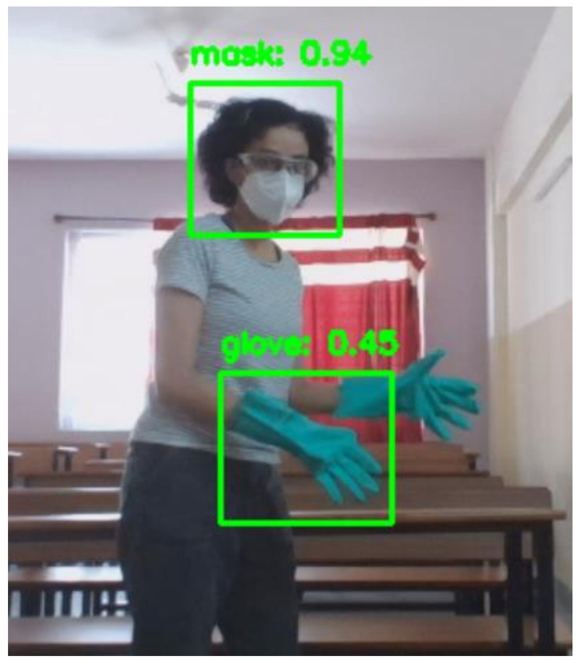


Fig. 7.

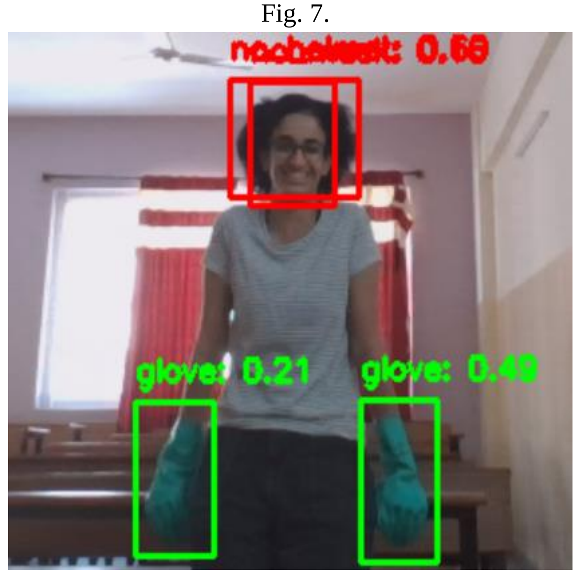


Fig. 8.

Another limitation found was the confusion between similar PPE classes. In Fig. 9, an individual with both goggles and a mask led to three distinct detections—“no mask” (0.21), “goggles” (0.12), and “mask” (0.36). This indicates that YOLOv8 occasionally has difficulty differentiating between goggles and masks because of overlapping facial structures. The model also showed considerable variation in confidence scores for varying poses of the same individual. In Fig. 10, a mask was found with 0.60 confidence with goggles at 0.39, whereas in another instance, the “no mask” label was present with 0.30 confidence even though the mask was actually there. Such discrepancies reflect the model’s pose change sensitivity, facial orientation, and lighting variations in the background.

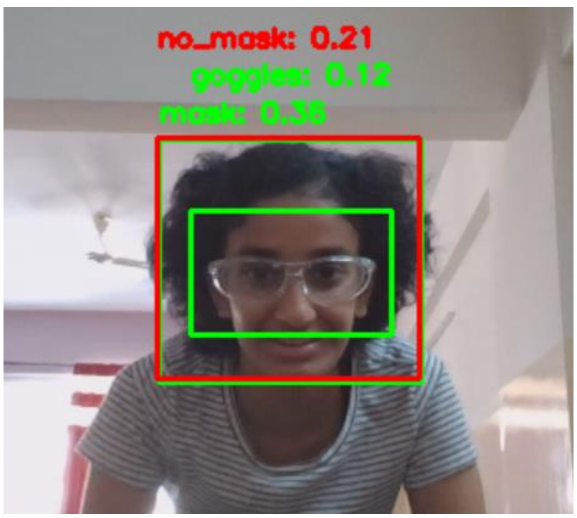


Fig. 9.

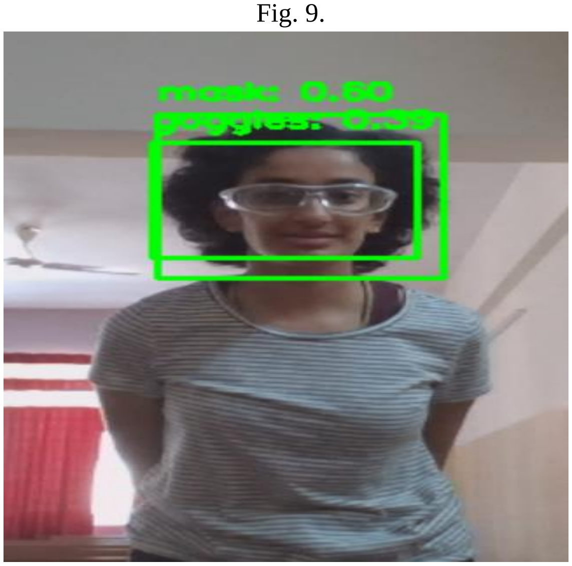


Fig. 10.

In Fig. 11, classifying full-body images, the model correctly labelled gloves but with lower confidence levels 0.35 and 0.40, and it also wrongly classified the lack of a helmet with a confidence of 0.60. This indicates that although YOLOv8 performs well when it comes to the detection of PPE items, its performance is tested when objects appear together in complex situations. These real-time detection outcomes are useful for gaining insight into the model's real-world performance. While YOLOv8 successfully identifies most PPE items, further optimization and fine-tuning are required to reduce misclassifications and improve stability, particularly in differentiating between similar safety equipment. Also, the model occasionally exhibited confusion between PPE classes, as in Fig. 12, where an individual with goggles and a mask yielded three detections—"no mask" (0.30), "goggles" (0.51), and "mask" (0.44). This indicates that separation between overlapping PPE items is still not fully resolved.

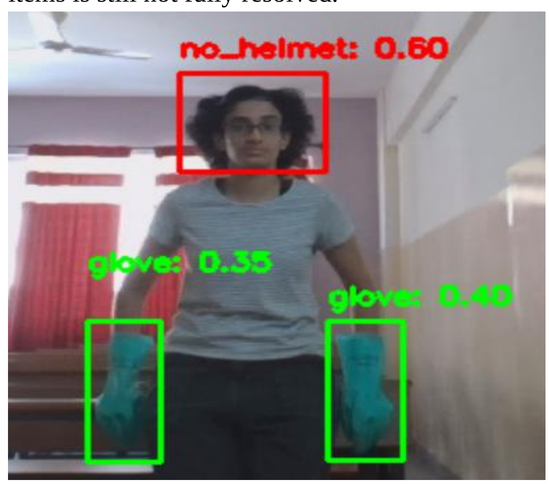


Fig. 11.

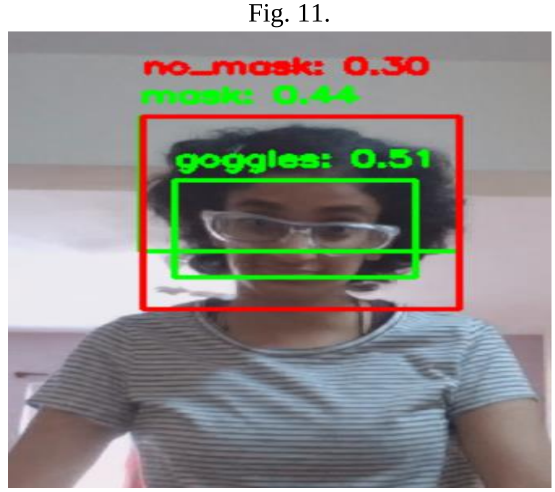


Fig. 12.

V. CONCLUSION AND FUTURE SCOPE

5.1 Conclusion

In this research, we compared three YOLO models- YOLOv5, YOLOv8, and YOLOv11, to use for PPE detection and find the most appropriate model for real-time deployment. From various performance metrics, including accuracy, precision, recall, and inference speed, YOLOv8 was selected as the optimal model for real-time PPE detection. Its balance between detection precision and computational resources made it the most viable choice for real-world deployment.

Through thorough testing on unseen data, we observed that YOLOv8 properly recognizes important PPE items such as masks, gloves, helmets, and goggles. However, some problems were noted, including occasional misclassifications, low-confidence detections, and difficulty in differentiating between comparable PPE items. These findings highlight the model's virtues but also indicate where it needs to be improved.

Despite its success, the study had some limitations, including dataset constraints, potential bias towards class balance and challenge in detecting PPE under occlusion and varying lighting conditions. These limitations can be resolved in future studies by employing a more heterogeneous data set, supplementing model insensitivity with multi-frame analysis, and using YOLOv8 in conjunction with other deep-learning architectures to enhance feature extraction. Also, there is scope for practical application in using the model in edge devices, integrating it with smart surveillance networks, and creating automated compliance checking systems. Overall, this research validates the pragmatic utility of YOLOv8 for PPE detection in real-time and provides a foundation for further future improvements and expanded applications in factory safety and workplace compliance checking.

5.2 Research Limitations

Although YOLOv8 performed well in detecting PPE, some of the limitations were noticed during the experiment. Misclassification and confusion between similar classes of PPE were one of the key challenges, for example, distinguishing between masks and goggles. The model sometimes annotated an object with multiple labels or provided inconsistent detections, indicating poor feature discrimination during training. Additionally, low confidence values in certain detections suggest that the model struggled under lighting, image resolution, and viewpoint changes. Such discrepancies would impact the reliability of performance in actual use, especially in dynamic conditions where workers are in motion and PPE may be partially occluded.

Another limitation was dataset bias and generalization problems. Since the model was trained on a particular dataset, its generalizability to a broad set of real-world conditions is not known. Factors like ethnicity, facial shape, PPE design subtleties, and background complexity can influence detection accuracy, which requires further dataset expansion to increase robustness. In addition, the model struggled in the low-light setup and in the occlusions where PPE components were obscured due to body angles, shadows, or ambient environments.

Secondly, apart from accuracy, real-time processing throughput and computing efficiency must be considered. While YOLOv8 is designed for real-time applications, performance may be liable to fluctuation based on the basis of hardware limitations. Execution of the model on edge devices or on mobile devices may require optimization in order to find a balance between accuracy and inference throughput. Lastly, regulatory and ethical considerations need to be addressed for real-world adoption, particularly for privacy and worker consent for use in workplace surveillance applications.

5.3 Future Scope

The future direction of this work lies in several directions, with the goal of improving the accuracy, efficiency, and practicality of PPE detection with YOLOv8. One of the most important directions is Fine-Tuning the model with a larger and more diverse dataset. The model can be enhanced to a greater extent by raising the dataset with more varied instances of PPE articles, dissimilar lighting scenarios, diverse settings, and true industrial environments. Further, augmented data from synthesized methods such as GAN-generated PPE images or domain adaptation algorithms can help further boost underrepresented scenarios. The second essential direction is further developing the robustness of the model to occlusions and misclassifications. Future developments can involve the integration of multi-view or multi-frame processing, where a few camera views or successive frames are processed to reduce errors in the detection of PPE. In addition, interfacing YOLOv8 with other deep neural models, such as transformers or attention-based models, would assist in more efficient feature extraction and differentiation between similar objects of PPE, reducing the ambiguity of masks, goggles, and helmets.

From an application perspective, real-time deployment of the model in industrial IoT (IIoT) and smart surveillance systems can enhance workplace safety. Integration of PPE detection with edge computing and embedded systems can enable on-site compliance monitoring without cloud processing, improving speed and security. Further, utilization of automated reporting and alerting systems that notify supervisors in case of non-compliance can help to improve worker safety and regulatory compliance.

Additional enhancements can also include multi-modal approaches, where YOLOv8 is coupled with thermal cameras or LiDAR sensors to improve detection in low-light or high-contrast settings. Moreover, incorporating action recognition models will not only detect PPE but also unsafe worker behavior, resulting in additional safety enhancements in the workplace.

Lastly, from the research and policy vantage, the amalgamation of PPE detection models with regulatory compliance structures is capable of assisting the development of AI-based safety standards. Subsequent studies can also focus on ethical concerns, in a way that these surveillance-based PPE detection systems are compliant with privacy laws and worker consent guidelines without diminishing the effectiveness in industrial safety domains.

REFERENCES

[1] M. Ferdous and S. M. M. Ahsan, “PPE detector: A YOLO-based architecture to detect personal protective equipment (PPE) for construction sites,” PeerJ Comput Sci, vol. 8, 2022, doi: 10.7717/peerj-cs.999.

[2] M. Ferdous and S. M. M. Ahsan, “PPE detector: A YOLO-based architecture to detect personal protective equipment (PPE) for construction sites,” PeerJ Comput Sci, vol. 8, 2022, doi: 10.7717/peerj-cs.999.

[3] C. Shi, D. Zhu, J. Shen, Y. Zheng, and C. Zhou, “GBSG-YOLOv8n: A Model for Enhanced Personal Protective Equipment Detection in Industrial Environments,” Electronics (Switzerland), vol. 12, no. 22, Nov. 2023, doi: 10.3390/electronics12224628.

[4] A. Elesawy, E. Mohammed Abdelkader, and H. Osman, “A Detailed Comparative Analysis of You Only Look Once-Based Architectures for the Detection of Personal Protective Equipment on Construction Sites,” Eng, vol. 5, no. 1, pp. 347–366, Mar. 2024, doi: 10.3390/eng5010019.

[5] 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.

[6] N. D. Nath, A. H. Behzadan, and S. G. Paal, “Deep learning for site safety: Real-time detection of personal protective equipment,” Autom Constr, vol. 112, Apr. 2020, doi: 10.1016/j.autcon.2020.103085.

[7] Z. Wang, Y. Wu, L. Yang, A. Thirunavukarasu, C. Evison, and Y. Zhao, “Fast personal protective equipment detection for real construction sites using deep learning approaches,” Sensors, vol. 21, no. 10, May 2021, doi: 10.3390/s21103478.

[8] D. Ocharo, H. C. Nganga, and S. Kiambi, “Automatic PPE Monitoring System for Construction Workers Using YOLO Algorithm Based on Deep Reinforcement Learning,” Int J Comput Appl, vol. 186, no. 48, pp. 10–15, Nov. 2024, doi: 10.5120/ijca2024924094.

[9] N. Kwak and D. Kim, “A study on Detecting the Safety helmet wearing using YOLOv5-S model and transfer learning,” International Journal of Advanced Culture Technology, vol. 10, no. 1, pp. 302–309, 2022, doi: 10.17703/IJACT.2022.10.1.302.

[10] M. Ferdous and S. M. M. Ahsan, “PPE detector: A YOLO-based architecture to detect personal protective equipment (PPE) for construction sites,” PeerJ Comput Sci, vol. 8, 2022, doi: 10.7717/peerj-cs.999.

[11] A. Hayat and F. Morgado-Dias, “Deep Learning-Based Automatic Safety Helmet Detection System for Construction Safety,” Applied Sciences (Switzerland), vol. 12, no. 16, Aug. 2022, doi: 10.3390/app12168268.

[12] Md. S. Islam, S. Shaqib, S. S. Ramit, S. A. Khushbu, A. Sattar, and S. R. H. Noori, “A Deep Learning Approach to Detect Complete Safety

Equipment for Construction Workers Based on YOLOv7," Jun. 2024, [Online]. Available: http://arxiv.org/abs/2406.07707

[13] Z. Wang, Y. Wu, L. Yang, A. Thirunavukarasu, C. Evison, and Y. Zhao, "Fast personal protective equipment detection for real construction sites using deep learning approaches," Sensors, vol. 21, no. 10, May 2021, doi: 10.3390/s21103478.

[14] C. Shi, D. Zhu, J. Shen, Y. Zheng, and C. Zhou, "GBSG-YOLOv8n: A Model for Enhanced Personal Protective Equipment Detection in Industrial Environments," Electronics (Switzerland), vol. 12, no. 22, Nov. 2023, doi: 10.3390/electronics12224628.

[15] A. Elesawy, E. Mohammed Abdelkader, and H. Osman, "A Detailed Comparative Analysis of You Only Look Once-Based Architectures for the Detection of Personal Protective Equipment on Construction Sites," Eng, vol. 5, no. 1, pp. 347–366, Mar. 2024, doi: 10.3390/eng5010019.

[16] J. Cordero, L. Bermeo, L. Barba-Guaman, and G. Riofrio, "Detecting the Use of Safety Helmets on Construction Sites," in Communications in Computer and Information Science, Springer Science and Business Media Deutschland GmbH, 2024, pp. 211–225. doi: 10.1007/978-3-031-58953-9_17.

[17] 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.