Research Article

Leveraging YOLO Models for Safety Equipment Detection on Construction Sites

Melike Çiftçi [1,*]**, Mehmet Uğur Türkdamar** [1]**, and Celal Öztürk** [2]

[1]   Computer Engineering Department, Erciyes University, Turkey; e-mail : mlkcftc01@gmail.com, mturkdamar@erciyes.edu.tr
[2]   Software Engineering Department, Erciyes University, Turkey; e-mail : celal@erciyes.edu.tr
*   Corresponding Author : Melike Çiftçi

Abstract: Occupational safety encompasses a range of practices adopted to protect the health and safety of employees. In the construction and industrial sectors, employees may be exposed to various risks such as falls, impacts, temperature changes and the effects of chemical substances. For this reason, personal protective equipment (PPE) is an important element for protecting employees against risks. The effective use of equipment such as a hardhat, mask, and vest makes an important contribution to the prevention of occupational accidents and health problems by ensuring the safety of employees. This study conducted three separate experiments investigating the potential of deep learning methods on occupational safety. In the first experiment, the YOLOv5n and YOLOv8n models were trained on the same data set with ten classes, and their performance was compared. In the second experiment, the YOLOv8n model was trained on a 2-class dataset to examine how the number of classes affected the model's performance. As a result of the experiments, it was seen that it emphasized the potential of deep learning and object detection methods to quickly and effectively monitor and evaluate the use of personal protective equipment.

Keywords: Deep Learning; Object Detection; Personal Protective Equipment; YOLOv5n; YOLOv8n.

1. Introduction

Construction sites are naturally dangerous and have a high potential for accidents and injuries[1]. Occupational accidents are among the main issues that we need to take precautions because every year, 2.3 million men and women suffer from work accidents or occupational diseases[2], and accordingly, more than 6000 fatal work accidents occur every day. The construction sector is one of the areas where employee safety is a priority. If employees use appropriate personal protective equipment such as a hardhat, mask, and vest, accidents such as injury, illness, and death can be prevented[3]. For this reason, it has become important to develop a widely available system that can detect workers' safety equipment in real-time and alert managers [4]. Figure 1 gives a statistical graph of the major types of fatal accidents that occurred in the United Kingdom between 2018 and 2019.

As a result of rapid advances in artificial intelligence (AI) technology, AI is widely applied to various industry fields. Although the use of artificial intelligence in the construction industry is slower than in other industries, AI; it has been easily integrated into areas such as tender, design, construction, safety and maintenance[5].

Object recognition, an important component of digital image processing applications, has been studied by researchers for many years[6]. Nowadays, object detection in computerized vision is the process of determining the positions of objects (varieties) in real-time images or video in variable conditions. With the use of the neural network, the ability to detect predetermined categories of objects and quickly adapt to various environmental conditions has been provided[7]. Algorithms such as R-CNN, YOLO, SSD are object detection methods that use deep learning architecture, and in the studies using these algorithms, the authors were able to achieve faster and more accurate results[8].
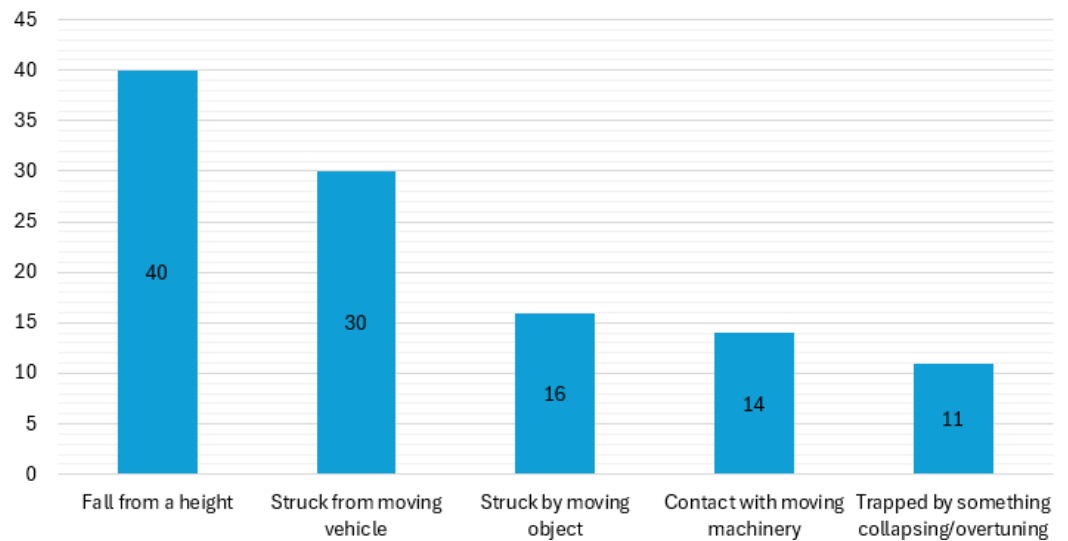
Figure 1. The main types of fatal accidents of workers working in the UK in 4542018-2019[3].

The You Only Look Once (YOLO) model was preferred for object detection purposes in this study. In images, YOLO represents an advanced approach to object detection. YOLO employs a single Convolutional Neural Network (CNN) to process the entire image, subdividing it into grids for analysis[9]. This streamlined architecture contributes to YOLO's rapid processing speed since the algorithm only requires a single pass through the network to obtain the final detection results[10]. Consequently, YOLO is capable of real-time detection, making it suitable for applications such as video processing, where time-sensitive analysis is essential.

In the health sector, in a range of applications such as monitoring of hospital equipment, placement of medical devices, and patient monitoring; in the automotive industry, in the improvement of parts recognition, quality control processes, and production processes; in the energy sector, in, monitoring of devices in smart energy networks, optimizing maintenance and repair processes, and finally in high-risk areas such as the construction industry, which are the subject of this work, artificial intelligence-assisted object detection and recognition technologies play an important role, especially in improving the use of personal protective equipment (PPE).

To improve the safety of workers on construction sites, AI systems can detect PPE in real-time and issue warnings to workers. This contributes to the prevention of potential accidents, the reduction of injuries, and the overall improvement of occupational safety. In this way, artificial intelligence technology is making working conditions in the construction industry safer and raising occupational safety standards. The following sections of this study are shaped as Related Studies, Materials and Methods, Experimental Study, and Conclusion.

2. Literature Review

Many studies determine the use of personal protective equipment in the construction industry. Junjie Bao et al. [4] carried out an object detection study using the YOLOv8 model on a data set consisting of 3 classes: human body, human head, and whether the helmet is worn on the head or not.

Shichu Li et al. [11] conducted a study using the YOLOv8-AFPN-M-C2f algorithm to detect gloves of different colors. They emphasized the algorithmic performance of this algorithm compared to the YOLOv3, YOLOv5, YOLOv8n, YOLOv8s, LSKnet, Fasternet, EfficientViT and Efficientformerv2 algorithms tested on the same data set.

In their study, Adem Korkmaz et al.[2] aimed to verify the presence of helmets and whether workers are wearing them in order to ensure the safety of construction sites with automated methods. They used a data set of 7036 images in the study. This data set; includes the worker's body, worker's head, and helmet classes. In addition, using various data augmentation techniques, they increased the 7036 image data set to 16867 images. As seen in this study, the use of 3-class data sets is common in studies on object detection.

Zijian Wang et al.[3] Their study used a dataset that included human head, vest, and blue, red, white, and yellow helmet classes. Using a single data set, they evaluated the performance of YOLOv3, YOLOv4 and YOLOv5 models in object detection.

Muneerah M. Alateeq et al. [12] used two different data sets in their study. The first data set detects personal protective equipment (PPE), and the other data set detects heavy equipment such as bulldozers, excavators, and graders. The data set determined to detect PPE consists of 3 classes: workers, safety helmets, and reflective vests. There are a total of 5241 instances of these classes. In the same study, the YOLOv5 model was used for object detection. As a result of the study, they found that the general precision value of YOLOv5 in this data set was approximately 90%, and the recall value was 77%.

In his study, M. Chithik Raja [13] used a 2-class dataset containing helmet and human head images. By dividing the data set into 60% training, 20% validation, and 20% testing, YOLOv8x, YOLOv7, and YOLOv5 models were used for training and evaluating the performance of the models on the same data set.

As seen in the current literature, there have not been ten multi-class studies on detecting personal protective equipment.

3. Materials and Methods

Safety measures at construction sites are of increasing importance day by day. Effective implementation of these measures is critical to reducing work accidents, ensuring worker safety, and improving site management. In this context, technologies such as deep learning, computer vision, and object detection play an important role in closing the vulnerabilities in the construction site and taking more effective measures.

3.1. Deep Learning

Deep learning, a sub-branch of machine learning, makes sense out of high amounts of data using artificial neural networks that exist at its core. The input layer from the layers is the first layer, and the output layer is the last. The layers between the input and the output layer are also called hidden layers[14]. The output of the previous layer becomes the input to the current layer. The general representation of a multi-layer artificial neural network used in this study is presented in Figure 2.
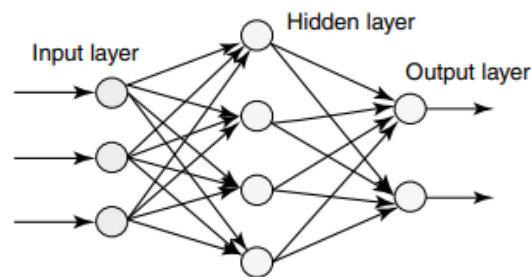

Figure 2. Multi-layer artificial neural network.

Artificial neural networks (ANN) are computer software that perform basic functions such as learning, remembering, generalizing, and producing new data from collected data by imitating the human brain's learning [15]. Mathematical modeling of the nervous system and decision-making mechanisms of living things is the main subject of artificial neural networks. The learning structure of the living brain is imitated thanks to artificial neural networks that can be trained, self-organized, and able to make adaptive evaluations[16].

Deep learning architecture Convolutional Neural Networks(CNN), which are widely used in image analysis processes, consist of at least one or more convolution layers. In addition, real-life problems can be easily solved due to the presence of nonlinear functions in the convolution layers. CNN architecture comprises layers of input, convolution, pooling, activation, and classification [17].

3.2 Computerized Vision and Object Detection

Computer vision is the performance of computer analysis of one or more images using various techniques in a time sequence with one or more main processors. Computer vision is

a science that enables information to be extracted and examined theoretically and algorithmically by a computer through images or image sets. Regarding the object and objects on the image, it includes concepts such as location, orientation, and size[18].

The object detection process in the image is expressed as the presence of the object in the image and the determination of the object class and its position in the image if it exists [19]. Object detection aims to develop computational techniques that provide one of the most fundamental pieces of information needed by computer vision applications. Object detection forms the basis of many computer vision tasks and is used in many fields[20]. Object recognition systems often make use of deep learning and convolutional neural network techniques and use algorithms such as Haar-Cascade, R-CNN, YOLO, and SSD.

3.3. You Only Look Once (YOLO)

Joseph Redmon is a computer vision family member, highly acclaimed for Santosh Divvalai, Ross Girshick, and Ali Farhadi, promoting YOLO and winning OpenCV's People Choice Award[21]. YOLO algorithm is a state-of-the-art (SOTA) object recognition system designed for real-time processing using CNN principles[22]. In Figure 3, we see the general structure of the YOLO architecture.
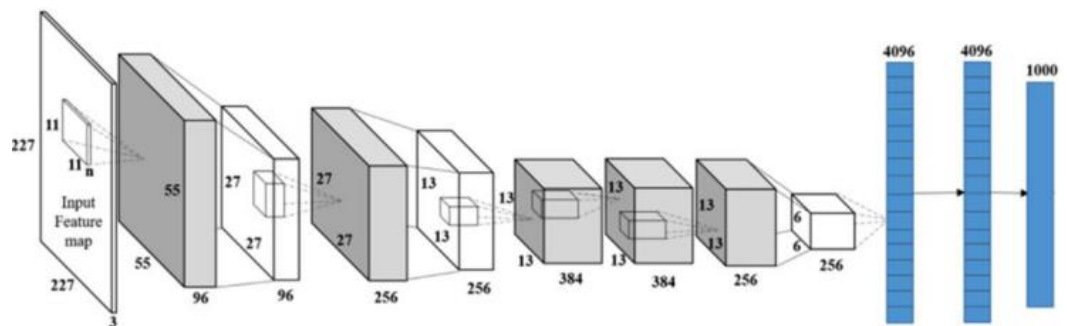

Figure 3. The architectural structure of the YOLO algorithm [23].

Unlike Faster R-CNN, YOLO can detect objects faster by not needing possible areas where the object can be found. YOLO draws the bounding box coordinates for each class in the image due to its regression process. Attributed to the classification layer, the probability of the predicted class is annotated on the bounding box. The difference between YOLO and most object detection algorithms is that it performs the regression and classification processes described above in one go. In other words, instead of performing separate operations for each class in the image, it looks at the image only once to create the bounding box coordinates of all classes and the possibilities of the classes[24].

The descriptive attributes of each bounding box in the image are the center position (Bx, By), the width of the object (Bw), the height of the object (Bh), and the predicted class (obj). The probability of estimation of the presence of an object in the grid (P(obj)) is a confidence score that reflects the presence and absence of the object[25]. It shows how similar the predicted object is to the real object. The confidence score of the object is formed between 0 and 1. It is calculated as in Equation (1) [26].

$$\text{Confidence Score} = \text{P(Obj)} * \text{IoU} \tag{1}$$

The operations described are performed when an object is located in its bounding box. If a box situation containing more than one object is encountered, this situation is solved with the "anchor box" method, and separate boxes are created for each object. The values explained above are calculated for these boxes, and "IoU" (Intersection over Union) is used in this calculation[27]. The YOLO network used in the study was used for detection purposes. The IoU value was calculated for each detected object in the test images. That is, the IoU metric calculation was used in the experiments. The IoU value is calculated according to the formula in Figure 4.

Finally, the YOLO algorithm eliminates bounding boxes that do not contain objects or contain the same object as other bounding boxes with the NMS (Non-Maximum Suppression) technique, with the help of the IoU (Intersection over Union) threshold value given in

Figure 5. In this way, while the calculation cost of the system is reduced, a result that does not tire the eyes is obtained[28].

From its emergence in 2016 until today (2023), the YOLO family has continued to evolve rapidly. Although first author Joseph Redmon ended work in computer vision at YOLO-v3, various authors have further developed the core 'combined' concept's effectiveness and potential. The latest version of the YOLO family appeared in the form of YOLO-v8[29]. Figure 5 gives the timeline for the development of the YOLO algorithm.
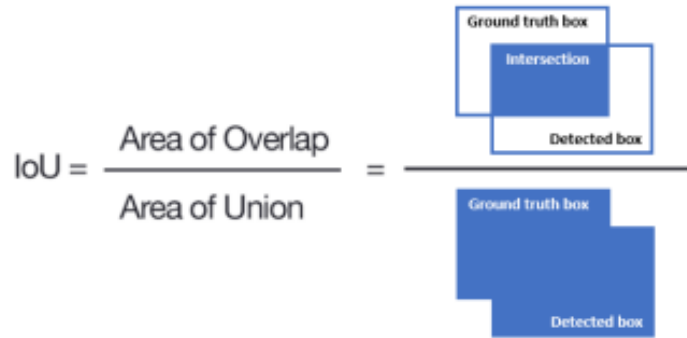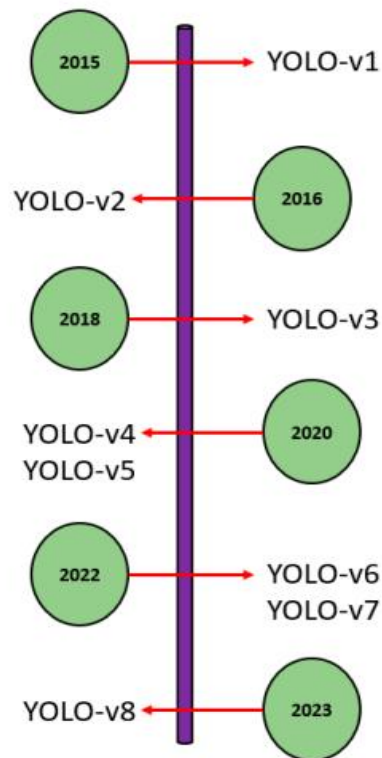

Figure 4. The IoU Formula.

Figure 5. Timeline of the YOLO development process [29].

In the field of agriculture, YOLO models have been used to detect and classify plants, detect pests and diseases, which has aided precision farming techniques and automated farming processes. They are also adapted for facial detection tasks in biometric, security, and facial recognition systems[30]. Bibliometric network visualization related to creating YOLO algorithms is shown in Figure 6.
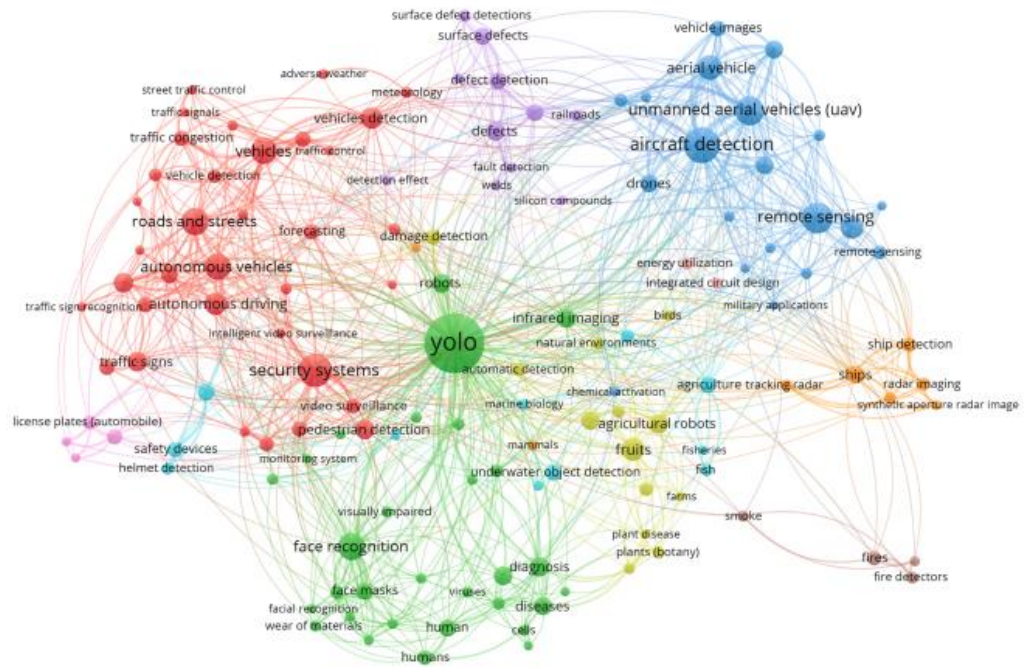
Figure 6. Bibliometric network visualization concerning the creation of the main YOLO applications (31)

4. Experimental Study

4.1. Data Set

We used two different data sets for this study. The first data set, Construction Site Safety Image [31], is a rich resource used to evaluate and optimize security measures at the construction site. There are a total of 2801 images in the dataset. Of these, 2605 are for training, 114 are for validation, and 82 are for testing. The ten classes in the dataset have a (.txt) structure labeled in YOLO format. These classes are Hardhat, Mask, NO-Hardhat, NO-Mask, NO-Safety Vest, Person, Safety Cone, Safety Vest, machinery, and vehicle. This labeling ensures accurate identification of the classes required for training the object detection model. Additionally, the Personal Protective Equipment (PPE) Class Map has paired each class with a number, making it easier to understand class relationships when evaluating object detection results. Table 1 shows the PPE class map of the dataset.

Table 1. PPE Class Map.

Class	Equipment
0	Hardhhat
1	Mask
2	NO-Hardhat
3	NO-Mask
4	NO-Safety Vest
5	Person
6	Safety Cone
7	Safety Vest
8	machinery
9	vehicle

The second data set, YOLO Helmet/Head [32], was employed to compare the model's performance based on the number of classes. This data set, structured in YOLO format, labels 'helmet' as 1 and 'head' as 0. The data split ratio for training, validation, and testing was approximately 0.7/0.2/0.1, with each .jpg image file accompanied by a corresponding .txt file

in the 'labels' directory. The .txt file contains object number and coordinates for each object detected in the image, listed on separate lines.

4.2. Determination of Training Models

4.2.1. YOLOv5

Introduced in 2020 by Ultralytics, creators of the earlier versions YOLOv1 and YOLOv3, YOLOv5 is a model for detecting objects that achieves state-of-the-art performance on the COCO benchmark dataset. It is known for its speed and efficiency both during training and in deployment[33].

The architecture of YOLOv5 includes three main components: the CSP-DarkNet backbone, the FPN+PAN neck, and the prediction head [34]. The backbone serves to extract features from the input images. The neck module consolidates these features and produces feature maps at three different scales. The prediction head then uses these maps to detect objects [35].

There are four versions of the YOLOv5 model—YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x—each offering different depths and widths of feature maps to accommodate various user needs [36]. For our application in detecting equipment at construction sites, we opted for the YOLOv5n model, which is characterized by its real-time detection capabilities and ease of deployment. This version is the smallest and fastest among the YOLOv5 models.

4.2.2. YOLOv8

YOLOv8 represents the newest installment in the YOLO series of object detection algorithms, which are renowned for their accuracy and real-time detection capabilities [37]. The core architecture of YOLOv8 resembles that of YOLOv5, including a Feature Pyramid Network (FPN) that is composed of spine, neck, and head components. However, the components have undergone significant modifications[38]. A key innovation in YOLOv8 is the introduction of the C2F module, or Cross-Stage Partial Bottleneck Dual Convolution, which enhances the model's accuracy by blending high-level features with contextual information [39].

YOLOv8 also incorporates several major advancements, such as a redesigned partial bottleneck, a new anchor-free approach, and a modified activation function in the top layer. The model uses a variety of loss functions to optimize performance, including binary cross entropy for classifying objects, Combined Intersection over Union (CIoU) for precise localization, and Distribution Focused Loss (DFL) for optimizing location predictions [40]. The YOLOv8 series is available in five different sizes, ranging from YOLOv8n, the smallest, to YOLOv8x, the largest. Each model variant is designed to meet different computational and performance requirements, as detailed in the referenced table with all models set for an input image size of 640 × 640[39].

Table 2. Parameter and FLOP numbers of YOLOV8 Models.

Model	Number of Params (M)	Flops (B)
YOLOv8n	3.2	8.7
YOLOv8s	11.2	28.6
YOLOv8m	25.9	78.9
YOLOv8l	43.7	165.2
YOLOv8x	68.2	257.8

Floating Point Operations Per Second (FLOP) represents the number of calculations required to train or predict a model. The higher the FLOP number of a model, the more computational power it requires, which means the more complex the model is.

Parameters (Params) refer to a model's number of learnable parameters. More parameters usually mean more learning capacity, but it can also mean that the model is larger and more complex.

Different sizes (n, s, m, l, x) of the YOLOV8 model have different parameters and FLOP values. "The model "YOLOV8n" has fewer parameters and a lower FLOP value than others. This means that the YOLOv8n model can be lighter and faster. Therefore, the YOLOv8n

model was used in this study because it was preferred in cases requiring less computational power and less memory usage.

4.3. Training and Testing of the Model

In this section, the training and evaluation processes of YOLOv8n and YOLOv5n models were carried out using the Ultralytics YOLO library. Each experiment was run for 20 epochs, and the batch size was set to 16. The image size was determined as 640 pixels. During training, the IOU (Intersection over Union) value was set to 0.2 for YOLOv5, while it was set to 0.7 for YOLOv8. After the training process, the model's performance was evaluated on the validation set, and then the trained model performed object detection on the test data sets. Visualizations were created using OpenCV and PIL libraries and saved in a specific folder.

The main purpose of these operations is to train the YOLO models on the data sets and then test the model's performance. We used a computer with an AMD Ryzen 5 7600X processor, NVIDIA GeForce RTX 3060 graphics card and 32 GB of RAM was used in this study.

4.4. Performance Metrics of the Models

Confusion matrix is a tool used to evaluate model performance in classification problems. This matrix visualizes how accurately or incorrectly the model classifies the actual and predicted classes. Four values need to be known to create the confusion matrix. These:

- TP (True Positive): The number of situations that the model correctly predicted as positive.
- TN (True Negative): The number of situations that the model correctly predicted as negative.
- FP (False Positive): The number of situations that the model incorrectly predicted as positive.
- FN (False Negative): The number of situations that the model incorrectly predicted as negative.

The success parameters of YOLO models can be calculated based on the results obtained using the confusion matrix. These are precision, recall, and F1 score.

Precision: It is a performance metric that measures how many of the instances that the classification model predicts to be positive are actually positive. The precision value is calculated as in Equation (2).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

Recall: It is the true positive number of a class, is the ratio of the total number of instances of that class. Recall value is calculated as in Equation (3).

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

F1 Score: It is the harmonic mean of precision and recall metrics. F1 score is calculated as in Equation (4).

$$\text{F1 Score} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} \tag{4}$$

The F1 score attempts to achieve a balance by considering the impact of both false positives and false negatives.

mAP: It is a metric used to measure the performance of object detection models. This metric gives the average of the model's accuracy and precision for all classes, thus evaluating the overall success of the model. A higher mAP value indicates a better model performance.

4.5. Experimental Results

The three experiments presented in this study evaluate the performance of the YOLOv5 and YOLOv8 models on the same data set and the performance of YOLOv8 on data sets with different class numbers. The confusion matrix visually presents the results of each

experiment, allowing us to compare model performance. Figures 7, 9, and 11 present these experiments' confusion matrices.

In addition, the class-based precision, recall, and F1 scores obtained in the experiments help us evaluate the model's performance for each class in detail. These metrics measure the model's ability to recognize specific classes by offering each class's accuracy, precision, and recall values. In the experiments, graphs of recall and f1 score values of each class were presented in Figures 8,10, and 12.
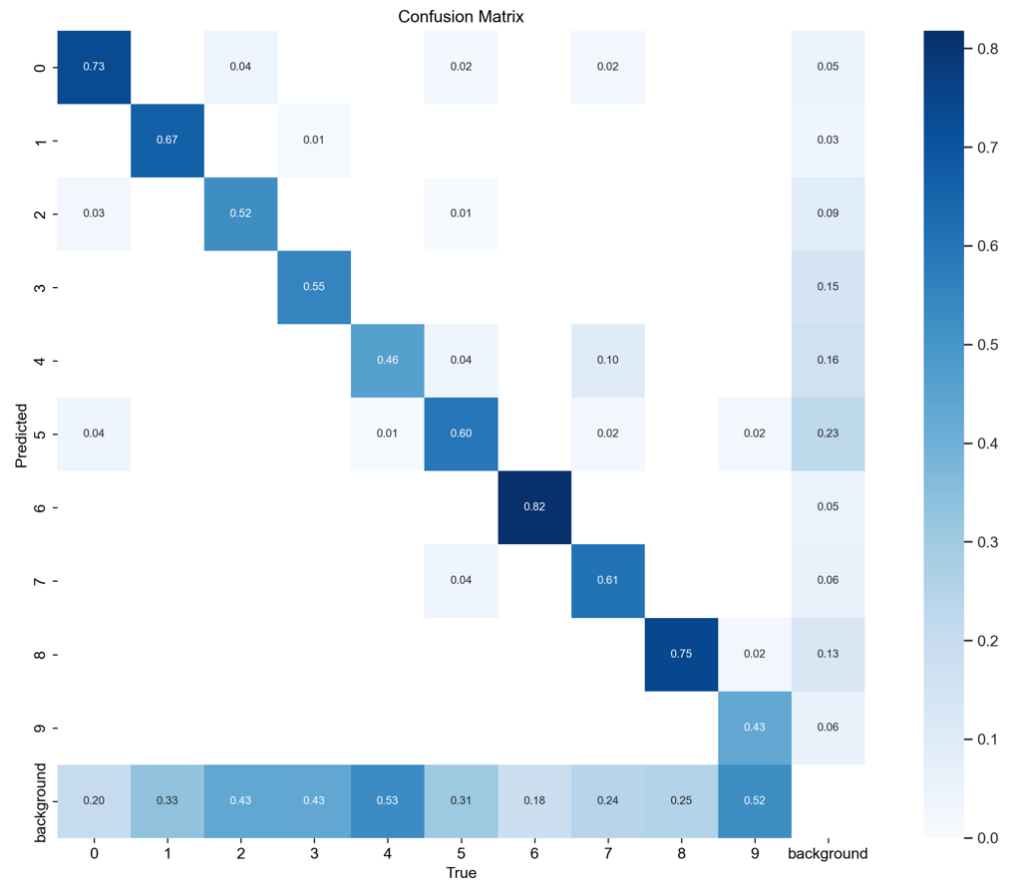
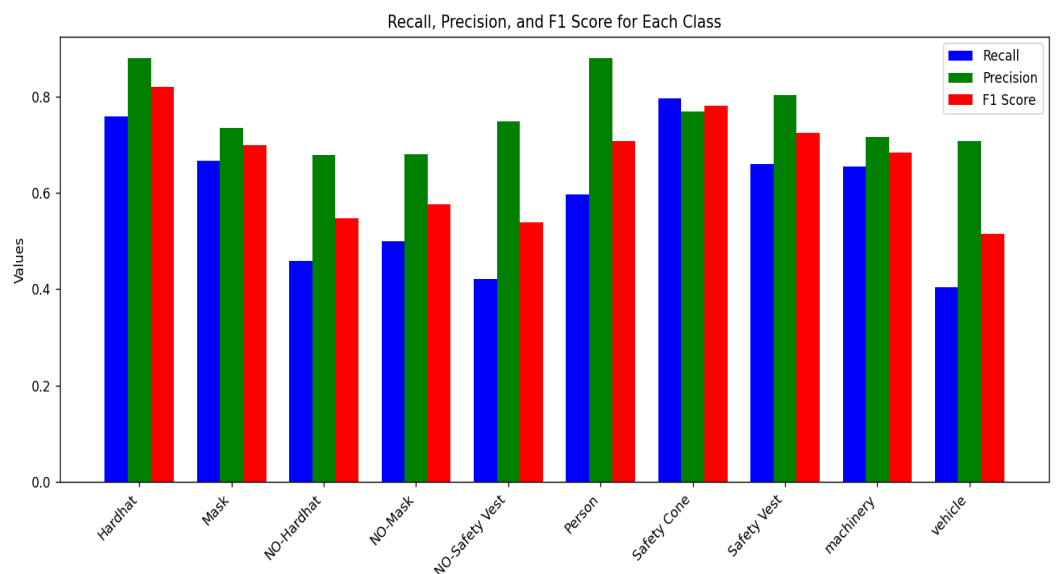**Figure 7.** Confusion Matrix of YOLOv5 Model on Construction Safety Data Set

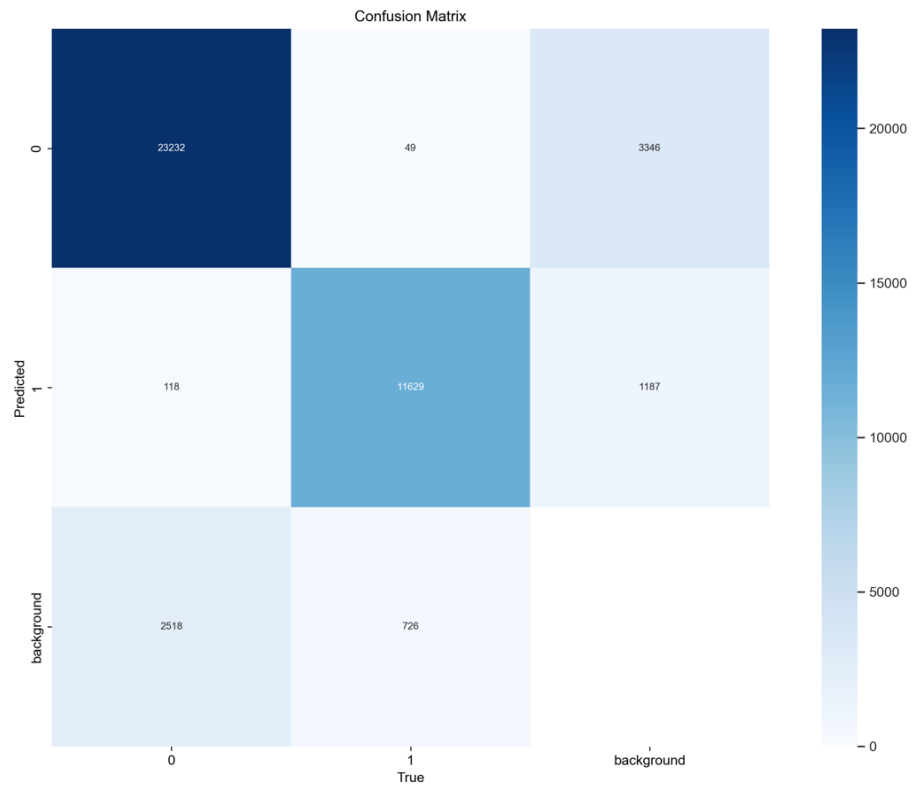**Figure 8.** Precision, Recall, and F1 Scores of YOLOv5 Model on Construction Site Safety Data Set

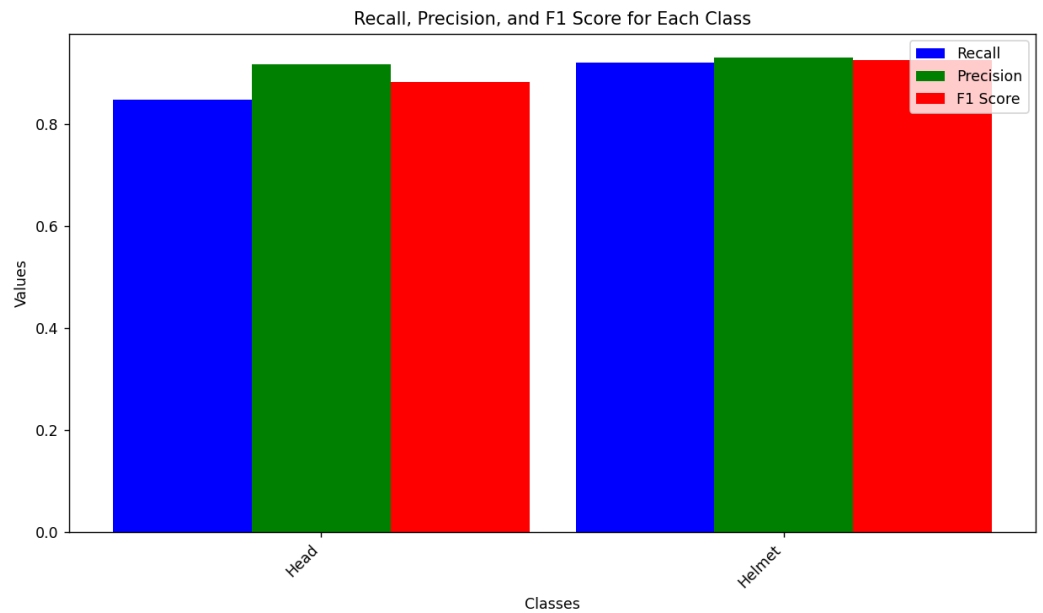**Figure 9.** Confusion matrix of YOLOv8 Model on YOLO Helmet/Head Data Set

Figure 10. Precision, Recall and F1 Scores of YOLOv8 Model on YOLO Helmet/Head Data Set

The metrics used to monitor the model's performance are a critical component of the training process. These metrics are used to measure the success and error rates of the model on the training dataset. Two important metrics are called "box_loss" and "cls_loss".

"box_loss" (box loss) is a metric that assesses the model's success in predicting object locations. It specifies how compatible the box coordinates calculated by the model are with their actual position to predict exactly where an object is located. A lower box loss indicates that the model more accurately predicts object positions.
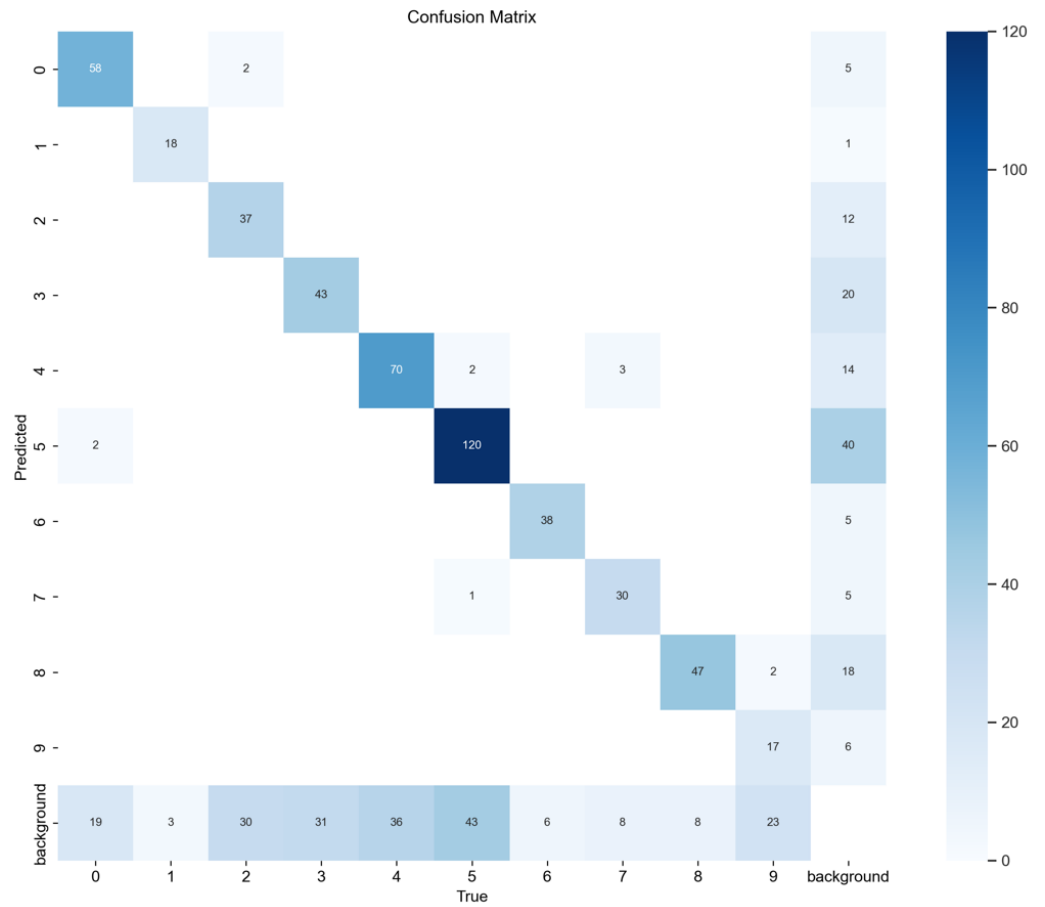
Figure 11. Confusion matrix of YOLOv8 Model on Construction Safety Data Set
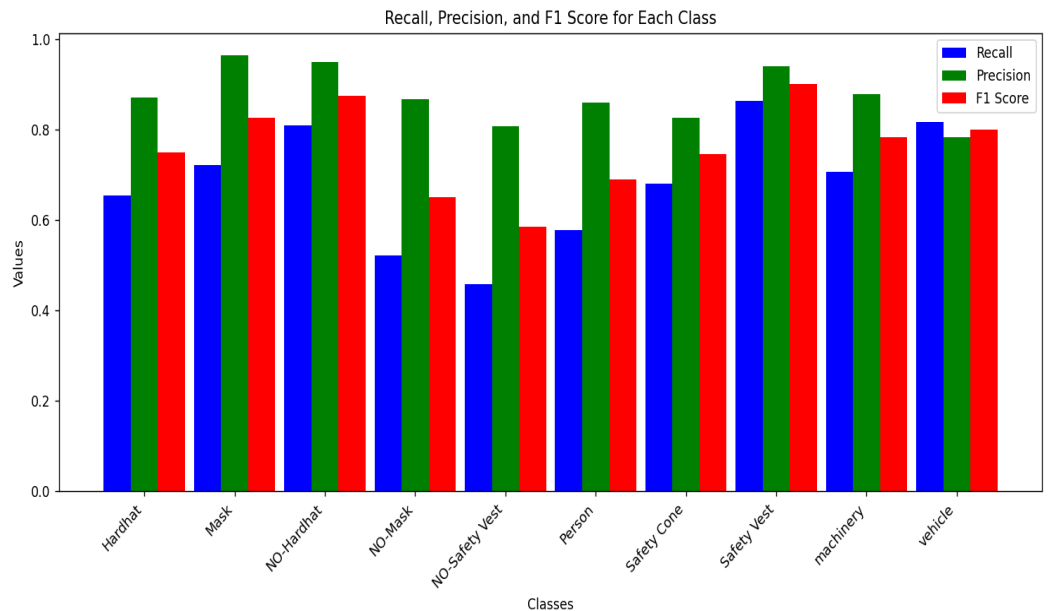

Figure 12. Precision, Recall and F1 Scores of YOLOv8 Model on Construction Site Safety Data Set

On the other hand, "cls_loss" (loss of classification) measures the ability of the model to classify objects accurately. That is, whether the model accurately predicts an object's class. A low classification loss indicates that the model can perform the classification task better.

As a result, metrics such as "box_loss" and "cls_loss" play an important role in evaluating and improving the model's performance in the object detection task. In the last part of the

training process, a graphical representation of these losses can help us see more clearly how the model's performance is changing. The graphics in Figures 13, 14, and 15 provide an important tool for monitoring the model's performance on the training dataset and adjusting the model as needed.
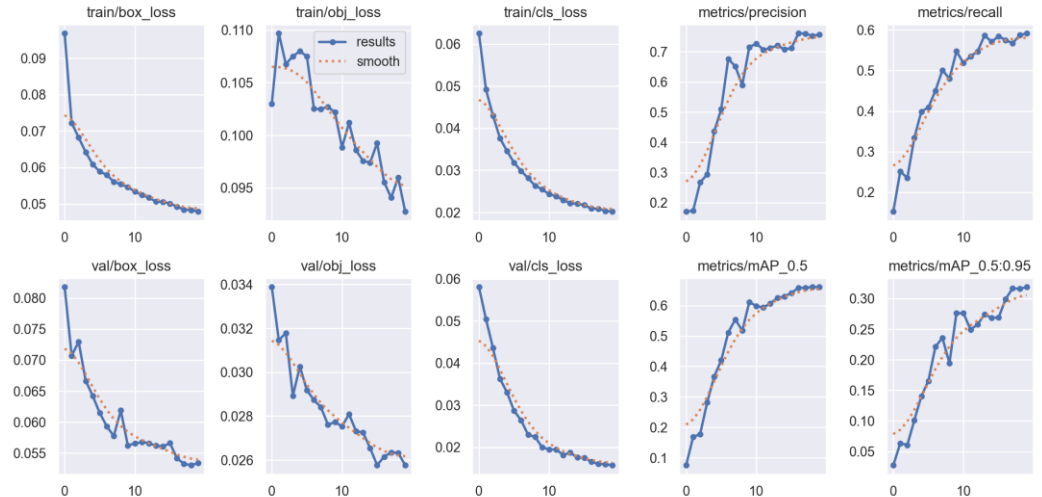

Figure 13. Training and Validation Results of YOLOv5 Model on Construction Safety Data Set
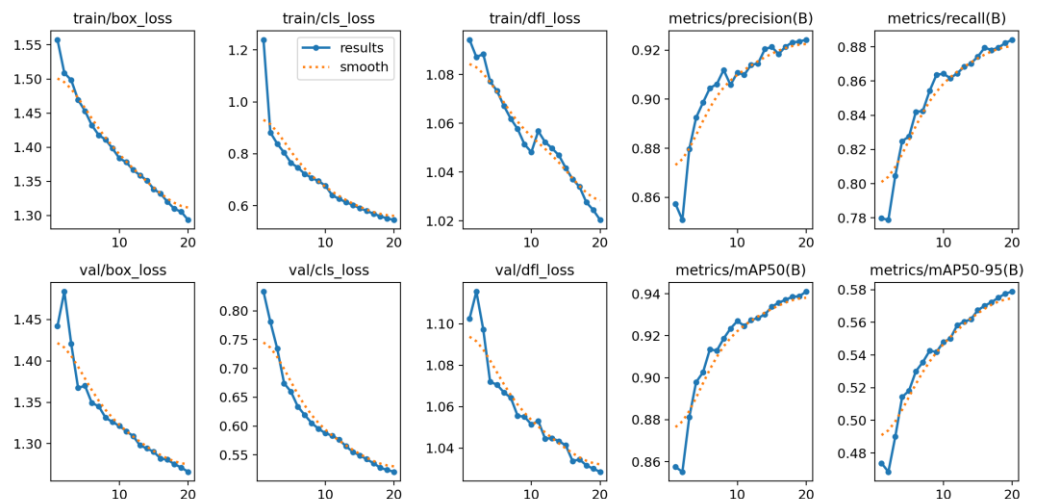

Figure 14. Training and Validation Results of YOLOv8 Model on YOLO Helmet/Head Data Set

As shown in Figures 13, 14, and 15, box_loss and cls_loss values are approaching zero. In this case, it can be said that the model accurately predicts the locations and classes of objects on the training data set, and the model's performance increases with the decrease in losses.

K-fold cross-validation is a widely used method for evaluating the performance of a deep learning model. We conducted K-fold cross-validation on YOLOv8. This method involves dividing the dataset into k subsets, or folds, and iteratively using k-1 folds for training the model and the remaining fold for testing. This process is repeated k times, with each fold used as the test set exactly once. In this study, we performed five iterations of cross-validation. This reduces the risk of overfitting. The results are shown in Table 3.
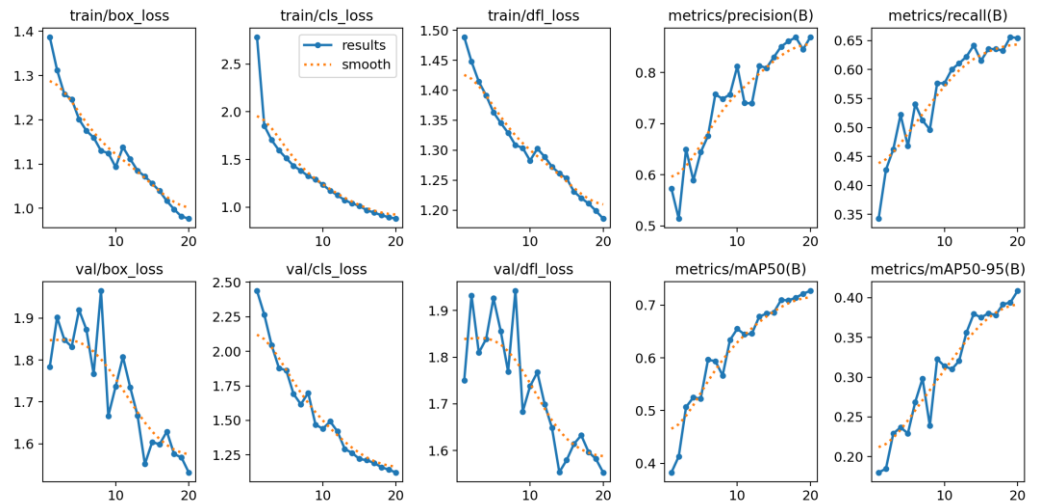
Figure 15. Training and Validation Results of YOLOv8 Model on Construction Safety Data Set

Table 3. The results of K-fold cross-validation.

Parameter	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Precision	0.819	0.861	0.891	0.898	0.900
Recall	0.628	0.681	0.699	0.723	0.769
mAP	0.711	0.764	0.794	0.811	0.824

Table 4. Comparison of performance metrics of experiments

	Precision	Recall	mAP50	F1 Score
YOLOv5 Model on Construction Safety Data Set	0.759	0.592	0.66	0.66
YOLOv8 Model on YOLO Helmet/Head Data Set	0.924	0.884	0.941	0.90
YOLOv8 Model on Construction Site Safety Data Set	0.871	0.654	0.727	0.74
YOLOv8 Model on Construction Site Safety Data Set using k-fold cross-validation	0.873	0.695	0.78	0.77

5. Conclusions

Equipment detection on construction sites is extremely important regarding occupational safety and efficiency. Research shows many benefits can be achieved by identifying the right equipment on construction sites, such as preventing potential hazards, reducing work accidents, and managing work processes more effectively.

Three different experiments on equipment detection at construction sites aimed to compare the performance of the YOLOv5 and YOLOv8 models. In the first experiment, a precision value of 75% was obtained from the YOLOv5 model in evaluating the construction site safety data set of 10 different classes. The YOLOv8 model obtained 87% precision value in the same data set experiment. These results indicate that YOLOv8 performs better than YOLOv5 when evaluated on the same data set.

In addition, an experiment was conducted on a 2-class YOLO Helmet/Head data set to evaluate YOLOv8's performance by class number. In this experiment, the YOLOv8 model obtained a precision value of 92% with the same parameters. When we look at the precision values, it is seen that the YOLOv8 model performs better in a 2-class data set than in a 10-class data set. In Tab.4, performance metrics for all experiments are presented.

These findings suggest that it is appropriate to choose the YOLOv8 model in the development and implementation of automatic equipment detection systems in construction sites. High accuracy values enable the accurate recognition of various equipment in the field, allowing occupational safety standards to be increased and business processes to be managed more

effectively. These experiments should be considered an important step toward meeting safety and efficiency requirements in the construction industry.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] A. Korkmaz and M. T. Ağdaş, "Deep Learning-Based Automatic Helmet Detection System in Construction Site Cameras," *Bitlis Eren Üniversitesi Fen Bilim. Derg.*, vol. 12, no. 3, pp. 773–782, Sep. 2023, doi: 10.17798/bitlisfen.1297952.

[2] S. Dündar, A. Bilim, and N. Bilim, "Türkiye'deki yaralanmalı iş kazalarının sektörel dağılımı ve analizi," *Ömer Halisdemir Üniversitesi Mühendislik Bilim. Derg.*, vol. 13, no. 1, pp. 28–33, Oct. 2023, doi: 10.28948/ngumuh.1292353.

[3] 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, p. 3478, May 2021, doi: 10.3390/s21103478.

[4] J. Bao, S. Li, G. Wang, J. Xiong, and S. Li, "Improved YOLOV8 Network and Application in Safety Helmet Detection," *J. Phys. Conf. Ser.*, vol. 2632, no. 1, p. 012012, Nov. 2023, doi: 10.1088/1742-6596/2632/1/012012.

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

[6] R. Daş, B. Polat, and G. Tuna, "Derin Öğrenme ile Resim ve Videolarda Nesnelerin Tanınması ve Takibi," *Fırat Üniversitesi Mühendislik Bilim. Derg.*, vol. 31, no. 2, pp. 571–581, Sep. 2019, doi: 10.35234/fumbd.608778.

[7] K. Hanbay and H. Üzen, "Nesne tespit ve takip metotları: Kapsamlı bir derleme," *Türk Doğa ve Fen Derg.*, vol. 6, no. 2, pp. 40–49, 2017, [Online]. Available: https://dergipark.org.tr/tr/pub/tdfd/issue/33022/367257

[8] O. Biçer and C. Zalluhoğlu, "Basketbol Hakemi El İşaretlerinin Nesne Tanıma Yöntemleri Kullanılarak Sınıflandırılması," *Int. Conf. Recent Acad. Stud.*, vol. 1, pp. 76–82, May 2023, doi: 10.59287/icras.675.

[9] S. Shinde, A. Kothari, and V. Gupta, "YOLO based Human Action Recognition and Localization," *Procedia Comput. Sci.*, vol. 133, pp. 831–838, 2018, doi: 10.1016/j.procs.2018.07.112.

[10] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, "A Review of Yolo Algorithm Developments," *Procedia Comput. Sci.*, vol. 199, pp. 1066–1073, 2022, doi: 10.1016/j.procs.2022.01.135.

[11] S. Li, H. Huang, X. Meng, M. Wang, Y. Li, and L. Xie, "A Glove-Wearing Detection Algorithm Based on Improved YOLOv8," *Sensors*, vol. 23, no. 24, p. 9906, Dec. 2023, doi: 10.3390/s23249906.

[12] M. M. Alateeq, F. R. P.P., and M. A. S. Ali, "Construction Site Hazards Identification Using Deep Learning and Computer Vision," *Sustainability*, vol. 15, no. 3, p. 2358, Jan. 2023, doi: 10.3390/su15032358.

[13] M. C. Raja, "Deep Learning-Based Head protector Detection for Construction Workers." Research Square, 2023. doi: 10.21203/rs.3.rs-3045767/v1.

[14] F. G. Tan, A. S. Yüksel, E. Aydemir, and M. Ersoy, "Derin öğrenme teknikleri ile nesne tespiti ve takibi üzerine bir inceleme. Avrupa Bilim ve Teknoloji Dergisi," *Eur. J. Sci. Technol.*, pp. 159–171, May 2021, doi: 10.31590/ejosat.878552.

[15] K. Öztürk and M. E. Şahin, "Yapay Sinir Ağları ve Yapay Zekâ'ya Genel Bir Bakış TT - A General View of Artificial Neural Networks and Artificial Intelligence," *Tak. Vekayi*, vol. 6, no. 2, pp. 25–36, 2018, [Online]. Available: https://dergipark.org.tr/tr/pub/takvim/issue/40063/427526

[16] M. A. Kizrak and B. Bolat, "Derin Öğrenme ile Kalabalık Analizi Üzerine Detaylı Bir Araştırma," *Bilişim Teknol. Derg.*, vol. 11, no. 3, pp. 263–286, Jul. 2018, doi: 10.17671/gazibtd.419205.

[17] M. Karakurt and İ. İşeri, "Patoloji Görüntülerinin Derin Öğrenme Yöntemleri İle Sınıflandırılması," *Eur. J. Sci. Technol.*, Jan. 2022, doi: 10.31590/ejosat.1011091.

[18] K. Yaman, A. Sarucan, M. Atak, and N. Aktürk, "Dinamik Çizelgeleme İçin Görüntü İşleme Ve Arıma Modelleri Yardımıyla Veri Hazırlama," *Gazi Üniversitesi Mühendislik Mimar. Fakültesi Derg.*, vol. 16, no. 1, pp. 19–40, 2013.

[19] E. Şimşek, B. Özyer, and G. Tümüklü Özyer, "Fotokapan Görüntülerinde Yerel Öznitelikler ile Nesne Tespiti," *Gümüşhane Üniversitesi Fen Bilim. Enstitüsü Derg.*, vol. 9, no. 4, pp. 633–644, Oct. 2019, doi: 10.17714/gumusfenbil.510717.

[20] F. Özbilgin and C. Tepe, "Robotik Uygulamalar İçin Derin Öğrenme Tabanlı Nesne Tespiti ve Sınıflandırması," *Karadeniz Fen Bilim. Derg.*, vol. 10, no. 1, pp. 205–213, Jun. 2020, doi: 10.31466/kfbd.734393.

[21] M. T. Ağdaş and S. Gülseçen, "Automatic Weapon and Knife Detection System on Security Cameras: Comparative YOLO Models," *Eur. J. Sci. Technol.*, pp. 16–22, Sep. 2022, doi: 10.31590/ejosat.1163675.

[22] A. Aouf, M. B. Mabedi, and H. Douadi, "Personal Protective (PPE) Detection Using Deep Learning," University of Ouargla, 2022. [Online]. Available: https://dspace.univ-ouargla.dz/jspui/handle/123456789/31771

[23] M. Nasrullah and A. Diker, "Derin Öğrenme ve YOLO Algoritmaları ile Nesne Tespiti ve Şerit Belirleme," in *2nd International Conference on Intelligent Transportation Systems*, 2021.

[24] A. Aktaş, Ö. Demir, and B. Doğan, "Derin öğrenme yöntemleri ile dokunsal parke yüzeyi tespiti," *Gazi Üniversitesi Mühendislik Mimar. Fakültesi Derg.*, vol. 35, no. 3, pp. 1685–1700, Apr. 2020, doi: 10.17341/gazimmfd.652101.

[25] S. Aydın, M. Taşyürek, and C. Öztürk, "MR Görüntülerinde Evrişimli Sinir Ağlar Kullanılarak Alzheimer Hastalık Tespiti," *Erciyes Üniversitesi Fen Bilim. Enstitüsü Fen Bilim. Derg.*, vol. 39, no. 3, pp. 357–368, 2023.

[26] E. Güçlü, İ. Aydin, K. Şahbaz, E. Akin, and M. Karaköse, "Demiryolu Bağlantı Elemanlarında Bulunan Kusurların YOLOv4 ve Bulanık Mantık Kullanarak Tespiti," *Demiryolu Mühendisliği*, no. 14, pp. 249–262, Jul. 2021, doi: 10.47072/demiryolu.939830.

[27] S. Yilmaz, "Beyin Tümörü Tanıları İçin YOLOv7 Algoritması Tabanlı Karar Destek Sistemi Tasarımı," *Kocaeli Üniversitesi Fen Bilim. Derg.*, vol. 6, no. 1, pp. 47–56, Jul. 2023, doi: 10.53410/koufbd.1236305.

[28] M. U. Türkdamar, M. Taşyürek, and C. Öztürk, "Transfer Öğrenmeli Derin Ağlar ile İnşaat Alanında Kask Tespiti," *Ömer Halisdemir Üniversitesi Mühendislik Bilim. Derg.*, vol. 12, no. 1, pp. 39–51, Nov. 2022, doi: 10.28948/ngumuh.1173944.

[29] M. Hussain, "YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection," *Machines*, vol. 11, no. 7, p. 677, Jun. 2023, doi: 10.3390/machines11070677.

[30] J. Terven, D.-M. Córdova-Esparza, and J.-A. Romero-González, "A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS," *Mach. Learn. Knowl. Extr.*, vol. 5, no. 4, pp. 1680–1716, Nov. 2023, doi: 10.3390/make5040083.

[31] I. Roboflow, "Construction Site Safety Computer Vision Project." https://universe.roboflow.com/roboflow-universe-projects/construction-site-safety

[32] Alexander, "YOLO helmet/head," *Kanggle.com*. https://www.kaggle.com/datasets/vodan37/yolo-helmethead

[33] D. Reis, J. Kupec, J. Hong, and A. Daoudi, "Real-Time Flying Object Detection with YOLOv8." May 17, 2023. [Online]. Available: http://arxiv.org/abs/2305.09972

[34] X. Jia, Y. Tong, H. Qiao, M. Li, J. Tong, and B. Liang, "Fast and accurate object detector for autonomous driving based on improved YOLOv5," *Sci. Rep.*, vol. 13, no. 1, p. 9711, Jun. 2023, doi: 10.1038/s41598-023-36868-w.

[35] J. Zhang, Z. Chen, G. Yan, Y. Wang, and B. Hu, "Faster and Lightweight: An Improved YOLOv5 Object Detector for Remote Sensing Images," *Remote Sens.*, vol. 15, no. 20, p. 4974, Oct. 2023, doi: 10.3390/rs15204974.

[36] R. Arifando, S. Eto, and C. Wada, "Improved YOLOv5-Based Lightweight Object Detection Algorithm for People with Visual Impairment to Detect Buses," *Appl. Sci.*, vol. 13, no. 9, p. 5802, May 2023, doi: 10.3390/app13095802.

[37] A. Dumitriu, F. Tatui, F. Miron, R. T. Ionescu, and R. Timofte, "Rip Current Segmentation: A Novel Benchmark and YOLOv8 Baseline Results," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2023, pp. 1261–1271. doi: 10.1109/CVPRW59228.2023.00133.

[38] B. Selcuk and T. Serif, "A Comparison of YOLOv5 and YOLOv8 in the Context of Mobile UI Detection," in *Mobile Web and Intelligent Information Systems*, Springer, Cham, 2023, pp. 161–174. doi: 10.1007/978-3-031-39764-6_11.

[39] I. Shamta, "İHA kullanılarak orman yangınlarının tespiti ve görüntülenmesi için derin öğrenme tabanlı gözetleme sisteminin geliştirilmesi," Karabük Üniversitesi, 2023.

[40] G. Oh and S. Lim, "One-Stage Brake Light Status Detection Based on YOLOv8," *Sensors*, vol. 23, no. 17, p. 7436, Aug. 2023, doi: 10.3390/s23177436.