SCIENTIFIC CULTURE
www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 2.1, (2026), pp. 4213-4224
Open Access. Online & Print

DOI: 10.5281/zenodo. 18930750

AUTOMATED HAZARD DETECTION AND SAFETY MONITORING FOR SMART CONSTRUCTION USING DEEP LEARNING ALGORITHM

Jothi Ganesan[1], Ahmad Taher Azar[2,3], Mona Alkanhal[2,3*], Nashwa Ahmad Kamal[4]

[1]Department of Computer Applications, Sona College of Arts and Science, Salem, Tamil Nadu, India
[2]College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia.
[3]Automated Systems and Computing Lab (ASCL), Prince Sultan University, Riyadh, Saudi Arabia
[4]Faculty of Engineering, Electrical Power and Machine Department, Cairo University, Giza, Egypt

Received: 27/12/2025
Accepted: 19/02/2026

Corresponding Author: Mona Alkanhal
(malkanhal@psu.edu.sa)

ABSTRACT

A construction site monitoring system combines several tools and technology to allow for remote monitoring of the site. The primary goal of this system is to maintain worker safety and to keep the project on schedule. This study proposes a deep learning-based surveillance system that detects safety compliance in real time and monitors worker behaviour. The present technique uses the YOLOv8 algorithm to train the deep learning model. If any abnormalities are found on the building site, the project engineer receives an alert immediately. The suggested solution allows constant monitoring of the facility, which improves worker safety. The model is trained using a construction site safety picture dataset obtained from the Roboflow universal datasets. The YOLOv8 model can recognize nine types of safety equipment. Image augmentation techniques are used to increase model accuracy. The suggested algorithm's performance is measured using a variety of measures, including accuracy, recall, the F1- measure, and mean Average accuracy (mAP). This deep learning- based monitoring system runs autonomously and requires no human intervention. The experimental outcomes were analysed using two distinct epochs: 50 and 100. The findings show that the YOLOv8 algorithm has the greatest accuracy of 0.94 when tested across 100 epochs.

KEYWORDS: Deep Learning, Object Detection, Safety Monitoring, YOLO Algorithm, Smart Construction.

1. INTRODUCTION

In construction sites, it is necessary to implement a safety monitoring system to protect workers and reduce the number of accidents (Kim et al., 2022). According to the United States Occupational Safety and Health Administration, the main cause of 4,764 workplace deaths per 100,000 workers in 2020 was the lack of proper safety equipment (Finkel et al., 2022). Insufficient safety equipment has led to numerous accidents on construction sites. The workers who lack personal protective equipment (PPE) have been a major contributing factor. PPE detection plays an important role in continuing safety monitoring systems. Due to heavy equipment, working at heights, and handling hazardous materials, the construction industry must push the importance of workplace safety (Nath et al., 2020). The real-time safety monitoring system is crucial for hazard prevention because it facilitates tracking safety measures, identify risks, and provide information that supports people making informed decisions to avoid misbehaviours. The monitoring system enables proactive control by monitoring worker safety, restricting unauthorized site access, and ensuring project execution adheres to the established plan. To improve safety on construction sites, researchers and practitioners have increasingly turned to advanced technologies such as the Internet of Things (IoT) and deep learning for monitoring purposes (Yu et al., 2020). Deep learning techniques, a subfield of artificial intelligence focused on image and video analysis, have found widespread application in the automated detection of safety equipment. Several prominent deep learning architectures, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Semantic Segmentation models, are commonly employed for this purpose within the context of construction site safety.

Over the past few decades, many researchers have focused on deep learning and computer vision models used to detect safety equipment on construction sites. Their studies have improved the construction scene and detected improper use of safety measures. The developed system can be used as an alarm source or display the results in a construction monitoring system. The main drawback of existing approaches is that the site cannot be monitored in the absence of the project engineer. To bridge the above gaps, this study introduces a real-time construction monitoring system to automatically detect safety measures and send an alert message to the relevant project engineer. The research contributions include:

1. Developing an automated construction monitoring system using a deep learning algorithm.
2. Using image augmentation technique to help improve algorithm performance.
3. Implementing the Yolov8 deep learning object detection algorithm to detect safety equipment on a smart construction site.
4. Incorporating the deep learning model experience into an interactive interface that sends a warning message to the project engineer for improper use of safety equipment.
5. The proposed integration reduces human intervention and project cost.
6. The proposed monitoring system facilitates remote site supervision, enabling construction engineers to provide real-time guidance to workers and ensure timely project completion.

The present research helps a real-time safety equipment detection system for construction sites, leveraging the YOLOv8 deep learning object detection algorithm. The system integrates continuous surveillance camera monitoring with this model (Raditya et al., 2024). The significance of the proposed system is that if any abnormalities are detected, it sends an alert message to the project engineer automatically. This helps the project engineer monitor the construction sites without their presence. The performance of the proposed system is assessed by using various measures such as overall model precision, recall, F1-measure, and mAP. The empirical results reveal that the YOLOv8 algorithm efficiently detects safety equipment on the construction site without human intervention. Key advantages include the identification and mitigation of risky behaviours, a reduction in accident-related costs, and the promotion of a safety-conscious and responsible work environment. Furthermore, the system's remote accessibility allows engineers to monitor multiple sites concurrently, minimizing project costs and manpower requirements.

The rest of the paper is structured as follows: Section 2 describes the relevant tasks for construction sites. Section 3 elaborates on the methodologies used. The experimental analysis is described in Section 4. The detailed results of the conducted experiments are discussed in Section 5. Finally, Section 6 presents the conclusions and outlines directions for future research.

2. RELATED WORKS

Shanti et al., (2022) developed a real-time monitoring system employing drones and deep learning algorithms to oversee worker activities at height. This approach utilizes drone-captured images as input for a deep-learning model. A trained model identified safety equipment, including harnesses, lifelines, and helmets. Furthermore, the system can regulate different weather and lighting conditions. The results demonstrate that the trained model achieves 90% accuracy, 97.2% precision, and a high recall rate. Biswas et al., (2023) developed a real-time construction site monitoring system using the YOLOv4 algorithm. This approach monitors the proper use of personal protective equipment among Bangladeshi construction workers, including boots, hard hats, gloves, face masks, and vests. The accuracy of YOLOv4 with the Darknet framework is 86.93% higher than that of the current method. The results show that the YOLOv4 algorithm performs better than YOLOv3 in terms of precision, recall, and mean average precision, with accuracy of 90%, 86%, and 96.93% respectively.

Arfan et al., (2023) implemented a PPE monitoring system that covers 132 different states within construction site environments. Experimental results demonstrate the system's reliability, achieving a mAP of 0.768, precision of 0.831, and recall of 0.693. The model encounters various difficulties in complex environmental conditions, including scaling and variations in image capture. Chen et al., (2024) developed the virtual safety query system integrating a head-mounted augmented reality (AR) device. The query system consists of three components: image acquisition, safety-focused visual question answering, and keyword-driven image-text retrieval. The results show that the proposed system greatly improved operational efficiency and safety awareness, achieving an accuracy of 89.7% and a recall of 80.1%. The automated system uses a visual question answering method to detect personal protective equipment in real-time (Wen et al., 2024). The YOLOv8 and double coordinate attention (DCA) modules are incorporated into this system to identify safety tools and provide safety assessment reports. The DNN model achieves 86% of mAP score, 95.60% of precision, 95.90% recall, and 91.20% of accuracy. Kim et al. developed deep neural network models to identify dangerous conditions at construction sites (Kim et al., 2024). In this approach, text-to image models creates artificial images and annotates them with segmentation masks.

Zhou et al., (2025) established a safety risk prediction system specifically for subway construction sites, utilizing deep reinforcement learning techniques. This approach leverages the double deep Q-network (DDQN), a sophisticated method within reinforcement learning, to enhance prediction accuracy. The proposed model was evaluated against existing machine learning frameworks, revealing that the DDQN effectively detects safety risks. Karatas et al., (2025) analyzed various deep learning algorithms aimed at assessing safety measures for work conducted at elevated heights. In this study, three distinct deep learning models, namely, Neural Networks (NN), Convolutional Neural Networks (CNN), and Long Short- Term Memory networks (LSTM) were examined across five different window sizes. The findings indicate that the CNN model achieved an impressive accuracy of 94.9% with a loss of 0.1696, outperforming other models in the analysis. Song et al., (2025) developed a drone-based safety monitoring system that employs the YOLOv10 model. This innovative model integrates IoT sensor data along with the GSConv module to facilitate lightweight feature extraction. Such an approach not only minimizes computational complexity but also enhances accuracy. The results demonstrate that the proposed model achieves a precision rate of 91.2% and a mean Average Precision (mAP) of 89.4%, surpassing the performance of existing models. Table 1 summarizes relevant research on safety monitoring systems for construction sites.

Table 1: Related Works for Safety Monitoring System in Construction Sites

Reference	Dataset	Total Images	Image Size	Algorithm Used	Accuracy %	Precision %	Recall %	mAP %
Shanti et al., (2022)	Real-Time Data	1000	224 x 224	YOLOv4 CNN	90.00	97.20	97.20	-
Biswas et al., (2023)	Real-Time Data	500	416 x 416	YOLOv4	86.93	90.00	86.00	0.869
Arfan et al., (2023)	Real-Time Data	132	256 x 256	SSD YOLO	90.00	83.10	69.30	0.768
Chen et al., (2024)	Real-Time Data	578	256 x 256	MLP	89.70	80.10	-	-
Wen et al., (2024)	MS COCO	4280	224 x 224	YOLOv8n-DCA	91.20	95.60	86.00	0.860
Kim et al., (2024)	MS COCO dataset	3585	640 x 640	YOLOv5 DNN	-	-	-	0.950
Karatas et al., (2025)	COCO & Kaggle	2801	640 x 640	YOLOv10	91.20	91.20	89.50	0.894
Song et al., (2025)	Real Time	-	-	CNN, DNN, LSTM	92.50	-	92.32	-

3. METHODOLOGY

The real-time safety surveillance system is designed to monitor construction sites, aiming to identify unsafe behaviors and ensure the precise usage of safety equipment (Lee et al., 2023). Object detection is a fundamental task in computer vision, enabling machines to identify, classify, and localize objects within images and videos (Saidani et al., 2024). YOLO (You Only Look Once) has emerged as a prominent approach among the numerous object detection algorithms developed recently. YOLO has been developed by Joseph Redmon in 2016, with several updates, evolving from YOLOv1 to YOLOv11. YOLOv8, the object detection model employed in this research, exhibits exceptional speed, achieving frame rates exceeding 45 FPS on a GPU. Table tab: compare presents the comparative study of architectural features for three different algorithms, namely YOLOv5, YOLOv8, and YOLOv10.

Tables II provide a systematic comparison of three distinct architectures: YOLOv5, YOLOv8, and YOLOv10. SPDarknet serves as the basis for the three algorithms. The most significant difference between these models is the activation function: Yolov5 has ReLU, Yolov8 has SiLU, and Yolov10 has the advanced activation function.

Table 2: Detailed Architectural Features Comparison.

Feature	YOLOv5	YOLOv8	YOLOv10
Backbone	CSPDarknet	CSPDarknet (enhanced)	CSPDarknet (optimized)
Neck	PANet	PANet (improved)	PANet (with efficiency optimizations)
Head	Anchor-based	Anchor-free	Anchor-free with dual assignments
NMS	Required	Required	NMS-free
Activation Function	Leaky ReLU	SiLU (Swish)	Advanced activation
Feature Pyramid	FPN	Modified FPN	Enhanced FPN
Loss Function	CIoU loss	Task-specific losses	Consistent Dual Assignment loss
Data Augmentation	Mosaic Cutout	Mosaic Mixup Cutout	Advanced augmentation techniques
Training Strategy	Single-stage	Single-stage	Two-stage with dual assignments

Based on the comparison table, it is considered that the YOLOv8 model is the most commonly utilized for construction site monitoring and identifying objects. In this study, the YOLOv8 model is utilized to detect the various safety measures that produce the greatest outcomes. This performance places among the fastest real-time object detection models currently available (Lou et al., 2023). YOLO employs a single neural network to simultaneously predict bounding boxes and class probabilities from the entire image, thereby operating as a one-stage object detector. Figure 1 visually represents the YOLOv8 model's architecture as applied in the safety surveillance system.

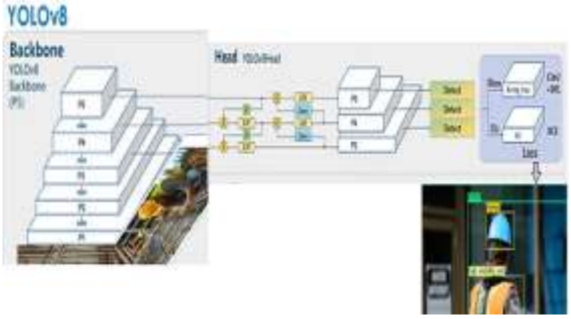
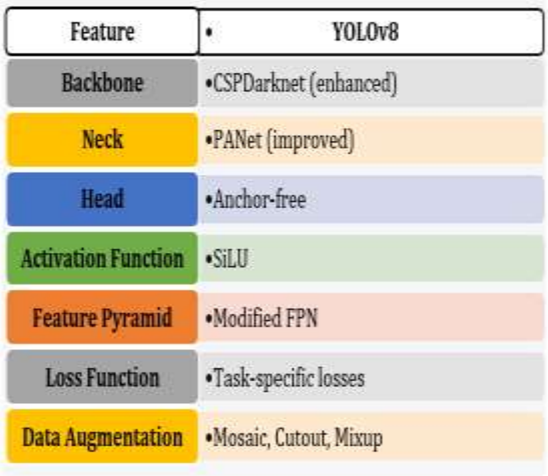


Figure 1: YOLOv8 Model Architecture for Safety Surveillance System.

This framework establishes a detailed methodology for object identification, for instance segmentation, and classification activities. The YOLOv8 architecture for object detection consists of three essential components: the backbone, the neck, and the head. The backbone serves as a feature extractor, deriving relevant features from the input image. The neck is responsible for feature fusion, which amalgamates various layers of data to enhance contextual awareness. The head is responsible for generating the final outputs such as bounding boxes and confidence levels for object detection. YOLOv8 leverages convolutional operations, applying filters to input images for feature extraction. Convolutional operations employ kernels, strides, and padding to preserve edge details during filtering. Following convolution, the Sigmoid Linear Unit (SiLU) activation function defined in equation 1 is applied.

$$SiLU(x) = x \times \sigma(x) \qquad (1)$$

where σ represents the sigmoid function. This activation function supports continuous and efficient learning across the network. Figure 2 denotes the various features of the YOLOv8 model architecture. YOLOv8's architecture leverages a CSPDarknet backbone, incorporating the SiLU activation function. A task-specific loss function guides training, while a modified Feature Pyramid Network (FPN) facilitates multi-scale feature fusion.



Feature	• YOLOv8
Backbone	•CSPDarknet (enhanced)
Neck	•PANet (improved)
Head	•Anchor-free
Activation Function	•SiLU
Feature Pyramid	•Modified FPN
Loss Function	•Task-specific losses
Data Augmentation	•Mosaic, Cutout, Mixup

Figure 2.: YOLOv8 Architectural Features.

4. EXPERIMENTAL ANALYSIS

The experiments were conducted using Google Colab Pro, leveraging Tesla T4 GPU (16GB VRAM) with a high-RAM runtime for model training and evaluation. The development environment included Python 3.8, PyTorch 1.13, Ultralytics YOLOv8, OpenCV 4.5.5, and CUDA 11.6. A dataset of 2,800 annotated construction site images, including workers, helmets, vests, and machinery, was used for training, with data augmentation techniques such as random cropping and brightness adjustments applied for improved generalization. The YOLOv8s model was trained using a batch size of 16, image size of 640×640, Adam optimizer, and cosine learning rate decay, for 50 epochs. Model evaluation was performed directly in Colab, and the trained model was exported for deployment. For real-time monitoring, the trained model was integrated with an IP camera system and deployed on a cloud-based server using Flask API, enabling live safety compliance detection and automated SMS notifications for safety violations.

4.1. Data Set

This study utilizes a publicly available dataset of construction site safety images sourced from Roboflow Universe (Raditya et al., 2024). A dataset of 2,800 images was compiled, featuring 10 distinct labels relevant to construction site safety. These labels include personal protective equipment such as helmets, safety vests, gloves, and masks, as well as machinery, scaffolding, and images of construction workers. The dataset comprises images of varying resolutions, which were preprocessed using Roboflow to meet specific training requirements (Roboflow et al., 2024). The image size of the YOLO model is 640 × 640 pixels. Figure 3 displays a sample image of the dataset.

Figure 3: Sample Input Images.

The dataset is split into three parts: 80% for training, 10% for validation, and 10% for testing. Figure 4 shows the percentage representation of each class across the training, validation, and testing sets. Figure 4 demonstrated that the classes/objects ″person,″ ″safety wear,″ and ″mask″ are the most frequently represented objects. This distribution indicates a main emphasis on worker safety, which aids in life-saving measures.

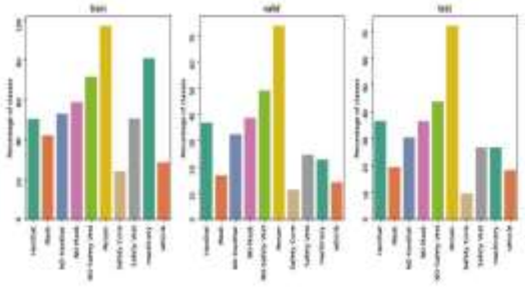


Figure 4: Dataset Distribution

4.2. Pre-processing

In pre-processing, input images are labelled using the YOLOv8 annotation and labelling tool. The collection includes annotated images of different safety standards and objects (Zolfagharian et al., 2024). For customized projects, additional building objects that were not in the original dataset can also be annotated. YOLOv8 uses the YOLOv5 PyTorch TXT annotation format, which is a modified version of the Darknet format. This study employs annotated data sourced from Roboflow Universe to train a YOLOv8 object detection model. The annotated images are depicted in Figure 5.

Figure 5: Annotated Images

4.3. Image Augmentation

Collecting a large number of imaging datasets is time-intensive and expensive. Image augmentation is a technique for artificially extending the size of a training dataset (Shorten et al., 2019). This improves the performance efficiency of the deep learning model. There is various image augmentation techniques accessible, including resize, rotate, flip, rescale, grayscale, brightness, cropping, and shifting. In this work, fifteen different augmentation approaches are used to improve the performance of the Yolov8 model. Figure 6 presents the outcomes of image augmentation techniques.

Figure 6: Image Augmentation Sample.

4.4. Results And Discussion

A YOLOv8 object detection model is trained on a construction safety monitoring dataset. Training employed an input image size of 640x640 pixels, a batch size of 16, and 50 or 100 epochs. Model performance is subsequently evaluated using mean average precision (mAP), precision, recall, and the F1-score. Precision is the accuracy of a model to make positive predictions (Lee et al., 2023). Similarly, recall is the ability of a model to detect all actual positive cases (Hayat et al., 2022). The mean average precision (mAP) is a popular statistical method used for the evaluation of object recognition tasks, including models such as YOLOv8. This scoring system integrates both precision and recall to produce a unified score that represents the proposed model's efficiency across various classes. These metrics are calculated with the following equations 2, 3, and 4.

$$Precision = \frac{Correctly\ Detected\ Safety\ Objects}{Correctly\ Detected\ Safety\ Objects + Incorrectly\ Detected\ Safety\ Objects} \quad (2)$$

$$Recall = \frac{Correctly\ Detected\ Safety\ Objects}{Correctly\ Detected\ Safety\ Objects + Missed\ Safety\ Objects} \quad (3)$$

$$mAP = \frac{\sum_{c=1}^{C} AP_{Category_c}}{C} \quad (4)$$

where:

- N is the number of classes.
- APi is the Average Precision for class i.

Experiments are conducted over two training durations: 50 and 100 epochs. The training and validation loss chart over 50 epochs is depicted in Figure 7. Optimal training and validation box losses of 0.724 and 1.278, respectively, are observed at epoch 50. The experimental results demonstrate that the proposed model became adept at recognizing objects in the training data while comparatively difficult with validation data. In the same way, the optical classification losses (CLS) were achieved at 0.425 for training at the 50th epoch and 1.247 for validation at the 47th epoch. Experimental results revealed superior object recognition performance on training data compared to validation data. In the same way, the optical classification losses (CLS) are achieved at 0.425 for training at the 50th epoch and 1.247 for validation at the 47th epoch, which means that the model efficiently learned to classify objects in the training dataset but is slightly challenging to generalize to the validation dataset. Regarding the distribution focal loss (DFL), a metric evaluating bounding box regression quality, the training loss reached a minimum of 0.964 at the 49th epoch. The validation loss was higher, with a minimum of 1.319 observed at the 44th epoch.

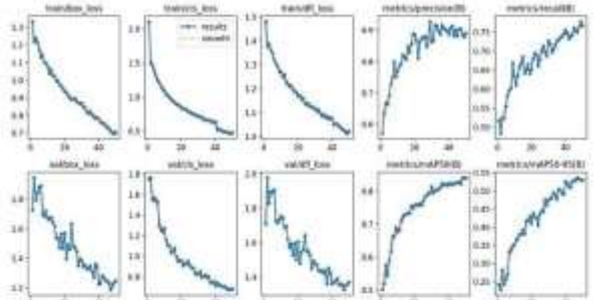


Figure 7: Training Loss chart - 50 epochs.

These inconsistencies between training and validation losses suggest some degree of overfitting, where the model performs significantly better on the training data than on test data. In this approach, the model achieved a precision of 0.872, recall of 0.754, and mean average precision (mAP) of 0.848. These scores replicate the overall performance of the model on the validation set. A precision of 0.872 indicates that the model has a high accuracy in predicting positive instances meaning most of its detections are correct. A recall of 0.754 shows that the model identified the actual positive instances, signifying some missed detections. The mAP score of 0.848 combines the model's ability to balance precision and recall across various confidence thresholds and intersection over union (IoU) thresholds.

Figure 8 depicts the training and validation loss charts for 100 epochs, as well as the model's performance metrics. At the 100th epoch, the best training box loss is 0.585, followed by the best validation box loss of 1.0945. The best training CLS loss of 0.3781 is achieved at the 100th epoch, while the lowest validation CLS loss of 0.5845 occurred at the 99th epoch. For DFL loss, the lowest training loss value 0.9501 is reached at the 99th epoch, while the best validation value 1.2834 is recorded at the 83rd epoch. The models have a precision of 0.856, recall of 0.8574, and mean average precision (mAP) of 0.894. Figure 8 illustrates the decreasing trend of both training and validation losses, indicative of effective model learning throughout the training epochs.

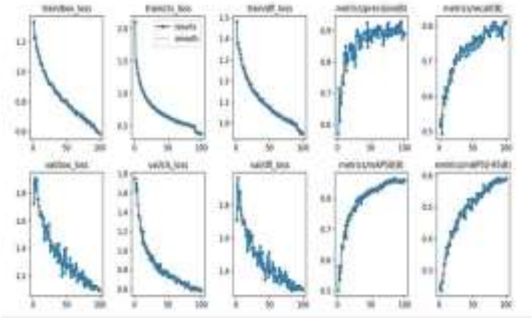


Figure 8: Training Loss chart - 100 epochs.

The observed discrepancy between training and validation losses, particularly for the box and DFL losses, suggests potential overfitting to the training data. The substantially higher validation box loss compared to the training box loss at the 100th epoch underscores the need for further investigation, potentially involving regularization techniques or hyperparameter fine-tuning. The model performs best in terms of CLS loss at the 99th and 100th epochs, indicating good generalization in classification. The DFL loss, optimal validation performance was observed at an earlier epoch (83rd), suggesting potential overfitting to the training data beyond this point. In terms of precision, a value of 0.856 indicating that a high level of accuracy in identifying relevant samples. A recall value of 0.8574 suggests strong performance in identifying relevant samples, indicating that the model captures a high proportion of true positives. The mAP of 0.894 indicates excellent performance in terms of balancing precision and recall across different thresholds, reflecting strong predictive capability.

4.5. Correlogram Analysis

The correlogram is a visualization tool that represents the correlation between various objects detected by the YOLOv8 algorithm (Granqvist et al., 2003). In this correlogram, the values range from 1to -1, and the number of each cell represents the correlation coefficient. The dark color, i.e., the value of 1, indicates that there is a positive correlation between two objects, and the light color, the value of -1, indicates a negative correlation, and 0 indicates no correlation. From this analysis, it is believed that objects, such as hardhats and masks, often appear together in the images, which means that they are safety tools that should be used concurrently. Similarly, the combination of a safety vest and a no-safety vest gives a negative correlation, which means that it is inversely related; if one is present, the other is probably absent. Figure 9 depicts the correlogram of the safety monitoring system and helps to understand how objects are related to each other in the data set based on the trained model.

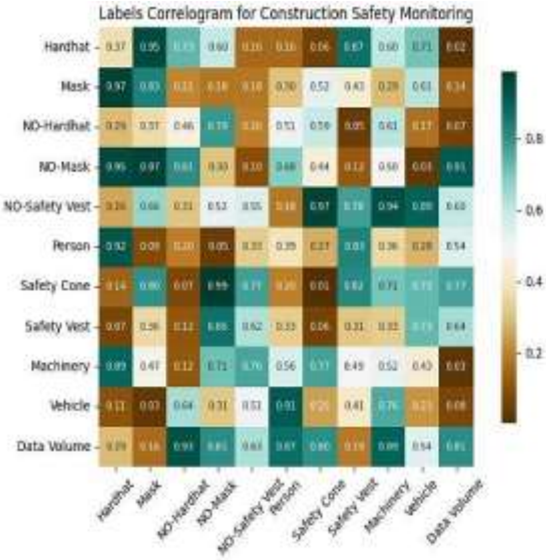
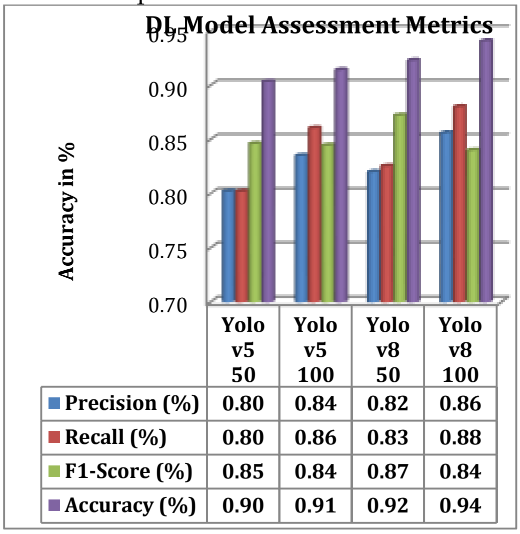


Figure 9: Correlation Map.

4.6. Deep Learning Model Assessment

The performance of the YOLOV5 and YOLOV8 models is evaluated using various evaluation measures such as precision, recall, F1-score, and accuracy. The experiments were conducted on two different epochs: 50th and 100th. The results of these measurements discuss the performance of the model on different epochs. From these results, it is possible to find out which model gives the best accuracy than the existing model. Figure 10 represents the experimental findings related to several assessment metrics, such as precision, recall, F1- score, accuracy, and mean Average Precision (mAP) for the Yolov5 and Yolov8 models. At the 100th epoch, the Yolov5 model records an accuracy of 91.42%, while the Yolov8 model demonstrates superior performance with an accuracy of 94.10% at the same epoch. Figure 10 illustrates that YOLOv8 provides better object detection performance than YOLOv5, and that an increase in training epochs generally correlates with improved results for both models. For both YOLOv5 and YOLOv8, extending the training epochs from 50 to 100 typically results in improvements across all metrics, particularly in mAP and accuracy, indicating that longer training periods facilitate the acquisition of more robust features and enhance their generalization capabilities.



	Yolo v5 50	Yolo v5 100	Yolo v8 50	Yolo v8 100
🟦 Precision (%)	0.80	0.84	0.82	0.86
🟥 Recall (%)	0.80	0.86	0.83	0.88
🟩 F1-Score (%)	0.85	0.84	0.87	0.84
🟪 Accuracy (%)	0.90	0.91	0.92	0.94

Figure 10: DL Model Assessment Metrics.

Figure 11 illustrates the performance of the Yolov8 model based on the distribution focal loss (DFL) plot. The DFL loss plot illustrates the training and validation loss trends over 50 epochs. From the figure, it is indicated that the YOLOv8 model learns to predict the bounding box coordinates very accurately during the training process.

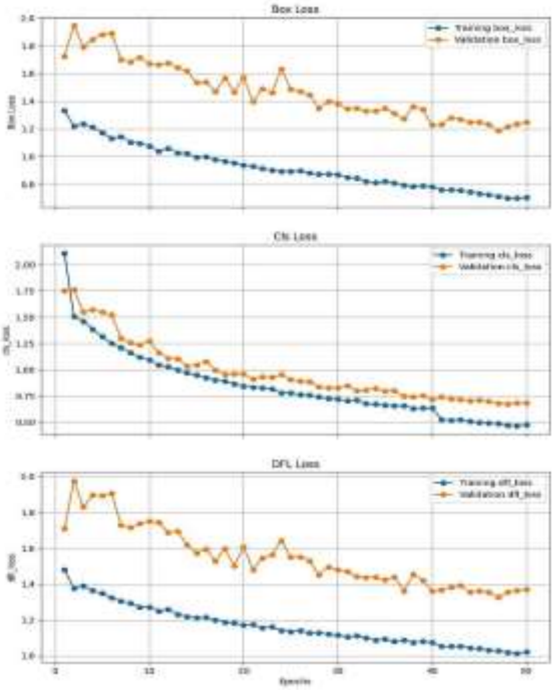


Figure 11: Training metrics vs epochs for Yolov8 50th Epoch.

Figure 12 represents the DFL loss chart for the Yolov8 model over 50 epochs. The training loss shows a steady decrease, indicating effective learning. However, the validation loss, while initially decreasing, levels off and shows variations after approximately 40–50 epochs. This variation between training and validation loss indicates potential overfitting. Although this model shows relatively low loss overall, further tuning may be required to improve its generalization capabilities.

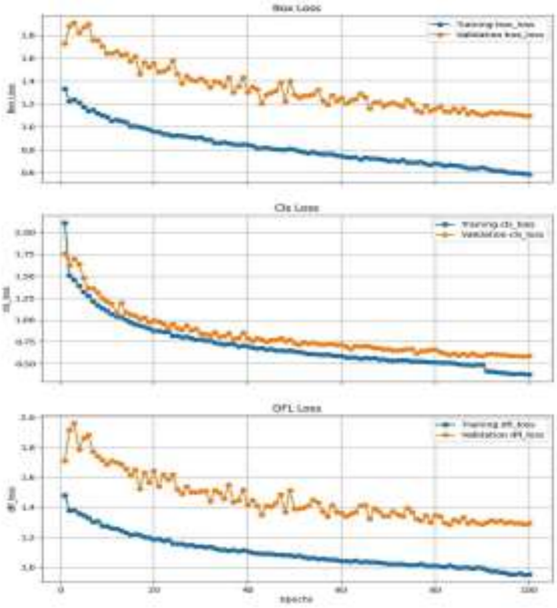


Figure 12: Training metrics vs epochs for Yolov8 100th Epoch.

Figure 13 depicts the mAP values for several models, namely Yolov5 and Yolov8, each evaluated at 50 and 100 epochs. Furthermore, Yolov8 achieves the highest mAP values among the models, with scores of 0.848 and 0.894 at the 50th and 100th epochs, respectively. The most significant advancement is noted in the mAP for YOLOv8, highlighting its enhanced ability to accurately identify objects across various classes and confidence levels.

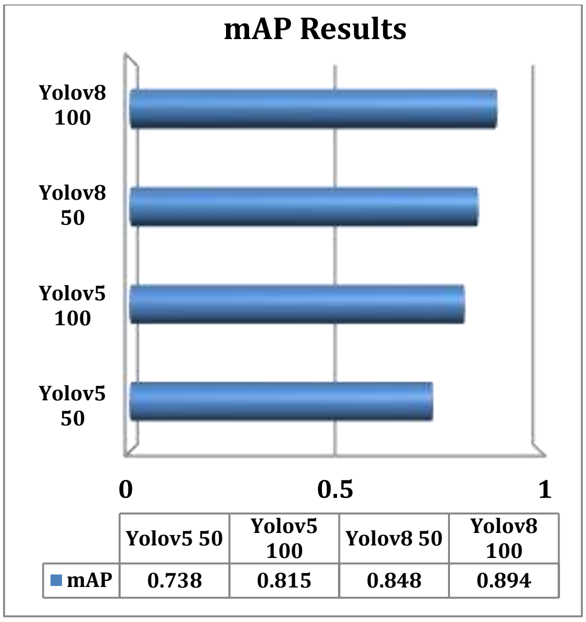


Figure 13: mAP Results.

A confusion matrix provides a visual representation of a classification model's performance by indicating the number of correct and incorrect predictions. The Figure 14 illustrates the confusion matrix for the YOLOv8 model during the 50th epoch.

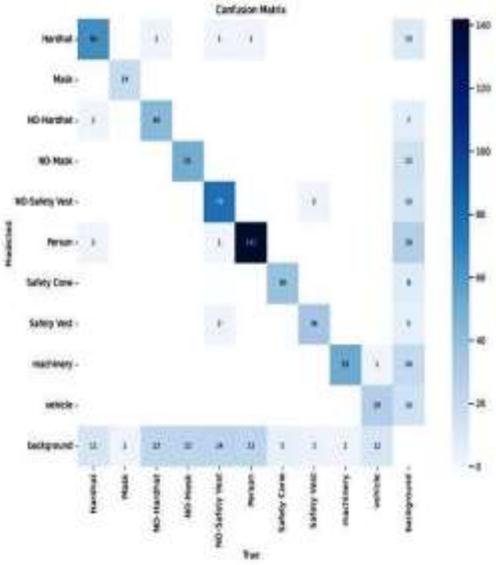


Figure 14: Confusion Matrix Yolov8 50th Epoch.

The model shows remarkable proficiency in classifying" Person" objects, as demonstrated by a high count of correct predictions totalling 342, with relatively few misclassifications. However, certain classes, including Hardhat, Safety Cone, and Person, are often misclassified as "Background," signifying difficulties in distinguishing these objects from the background. Furthermore, misclassifications occur between classes that may exhibit visual similarities, including "NO-Hardhat" and" NO-Mask." To assess the overall performance of the YOLOv8 model at the 50th epoch, one must analyze the distribution of counts across both the diagonal and off- diagonal cells. A proficient model will display high counts along the diagonal while maintaining low counts in the off-diagonal cells.

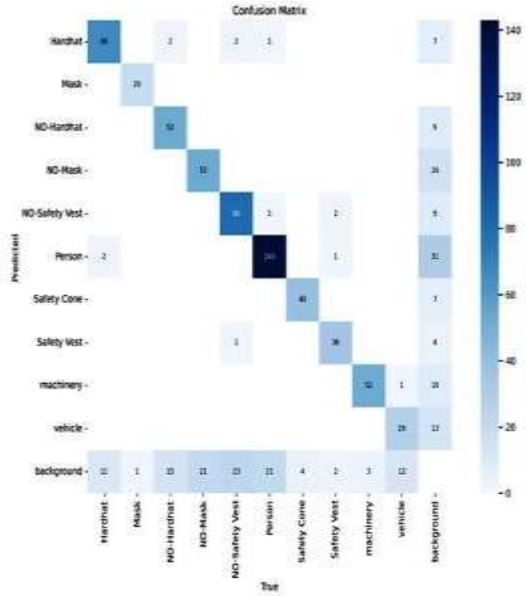
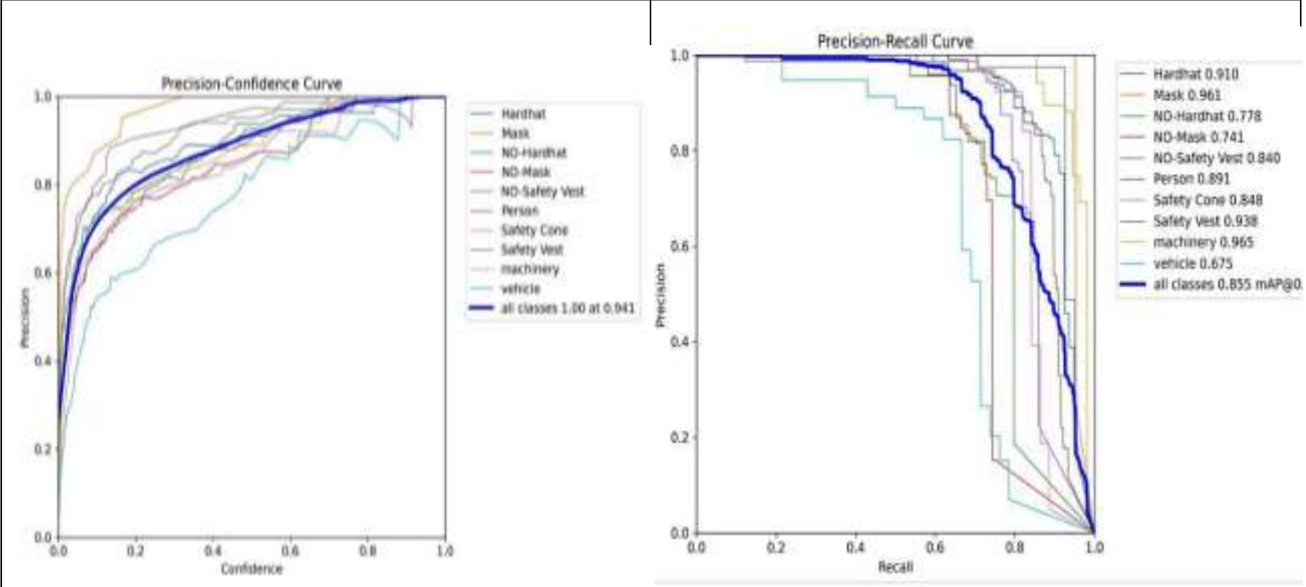


Figure 15: Confusion Matrix Yolov8 100 Epochs.

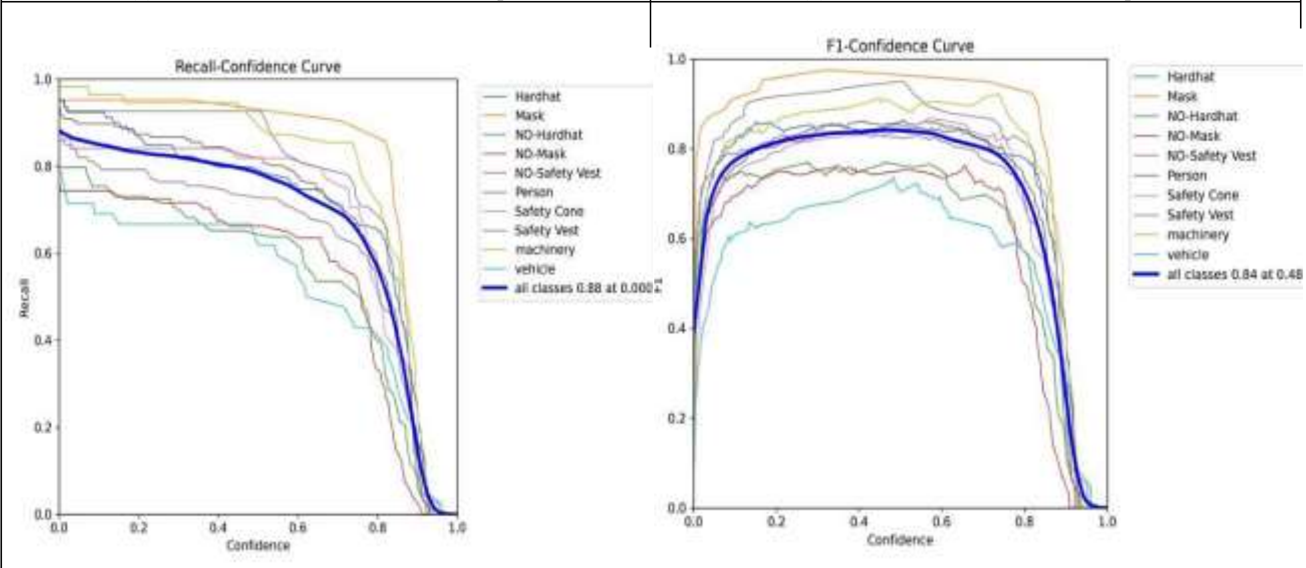
Figure 15 represents the confusion matrix for the Yolov8 model at the 100th epoch. From this analysis, it is clear that the model demonstrates strong performance in classifying" Person" objects, as indicated by the impressive count of correct predictions, which stands at 343. Additionally, it is noteworthy that several classes, including Hardhat, Safety Cone, and Person, are often misclassified as "Background."

This study used three different confidence curves to evaluate the efficacy of the surveillance monitoring system: the precision confidence curve, the recall confidence curve, and the F1 confidence curve. The x-axis of the presented curve represents the confidence threshold, varying from 0 to 1. The y-axis corresponds to the performance metric, which can be precision, recall, or F1-score, also ranging from 0 to 1. The precision-confidence curve is a valuable tool for estimating and improving model performance. Figure 16 (a), (b), (c), and (d) depicts various confidence curves for a 50-epoch assessment. In this Figures, the greatest precision value is estimated to be 0.923 with a confidence rate of 1.00 for all classes, indicating that the model accurately predicts safety compliance and worker actions. The model obtained maximum recall-confidence, F1-confidence, and precision-recall values of 0.87, 0.82, and 0.822, respectively, at different confidence threshold levels.



(a) Precision-Confidence Curve – 100 Epochs

(b) Precision - Recall Curve– 100 Epochs



(c) Recall-Confidence Curve– 100 Epochs

(d) F1-Confidence Curve– 100 Epochs

Figure 17. Confidence Curve – 100 Epochs.

The results of safety compliance and worker actions identification using the YOLOv8 algorithm are shown in Figure 18 and Figure 19. It represents the bounding boxes of various safety measures along with their precision values. It is also noted that the system generates an alert message if the worker is not wearing any safety tools.

ALERT: The following safety issues detected - No Mask, No Safety vest

Figure 18: YOLOv8 Model Prediction Output1

Figure 19 shows the results of No Hardhat detection in the given set of images. The proposed surveillance monitoring system sends an alert message to the construction engineer to facilitate them to take the required action promptly and potentially save their lives.

ALERT: The following safety issues detected - No Hardhat

Figure 19: YOLOv8 Model Prediction Output2

5. CONCLUSION

The increasing digitization of construction sites will likely drives further adoption of deep learning for safety monitoring. This trend promises to enhance workplace safety and efficiency through proactive, real-time interventions. This paper presents a deep learning-based safety surveillance system designed for automated monitoring of construction sites, enabling oversight without the constant presence of on-site engineers. The YOOv8 model utilizes surveillance camera imagery from construction sites as input. The YOLOv8 algorithm is subsequently employed to detect safety equipment and worker actions within these images. Algorithm performance was evaluated using several metrics, including precision, recall, the F1 confidence curve, and training and validation loss. The experimental results show that YOLOv8 achieves the highest model precision of 0.940 in a 100-epoch assessment when compared to a 50-epoch assessment. The proposed deep learning-based monitoring system offers several advantages. It enables remote site monitoring, allowing engineers to maintain oversight even when physically absent. Furthermore, the system facilitates simultaneous monitoring of multiple sites, resulting in significant time and labour savings for construction engineers.

Future work will expand the safety surveillance system's scope to encompass a wider range of equipment and machinery, thereby further enhancing construction site safety through automated monitoring. Additionally, incorporating optimization techniques for hyper parameter tuning will be explored to improve the performance of the detection models.

Acknowledgements: This paper is derived from a research grant funded by the Research, Development, and Innovation Authority (RDIA), Kingdom of Saudi Arabia, with grant number 13382-psu-2023- PSNU-R-3-1-EI-. The authors would like to acknowledge the support of Prince Sultan University, Riyadh, Saudi Arabia in paying the article processing charges of this publication. This research is supported by the Automated Systems and Computing Lab (ASCL), Prince Sultan University, Riyadh, Saudi Arabia. In addition, the authors wish to acknowledge the editor and anonymous reviewers for their insightful comments, which have improved the quality of this publication.

REFERENCES

Arfan, M., Sumardi, S., & Huboyo, H. (2023). Advancing workplace safety: A deep learning approach for personal protective equipment detection using single shot detector. 2023 International Workshop on Artificial Intelligence and Image Processing (IWAIIP), 127–132.

Biswas, S., Bhowmick, S., Islam, T., Ahmed, M. A., Islam, T., & Siddique, S. (2023). Real-time construction safety gear detection using yolov4 with darknet. 2023 IEEE 17th International Conference on Application of Information and Communication Technologies (AICT), 1–6.

Chen, H., Hou, L., Wu, S., Zhang, G., Zou, Y., Moon, S., & Bhuiyan, M. (2024). Augmented reality, deep learning and vision-language query system for construction worker safety. Automation in Construction, 157, 105158.

Finkel, A. M., Johns, D. O., & Whittaker, C. (2022). Occupational risk assessment. In Risk Assessment for Environmental Health (pp. 225–265). CRC Press.

Granqvist, S., & Hammarberg, B. (2003). The correlogram: A visual display of periodicity. The Journal of the Acoustical Society of America, 114(5), 2934-2945.

Hayat, A., & Morgado-Dias, F. (2022). Deep learning-based automatic safety helmet detection system for construction safety. Applied Sciences, 12(16), 8268.

Karatas, I. (2025). Deep learning-based system for prediction of work at height in construction site. Heliyon.

Kim, E., Ryu, H., Oh, H., & Kang, N. (2022). Safety monitoring system of personal mobility driving using deep learning. Journal of Computational Design and Engineering, 9(4), 1397–1409.

Kim, H., & Yi, J. S. (2024). Image generation of hazardous situations in construction sites using text-to-image generative model for training deep neural networks. Automation in Construction, 166, 105615.

Lee, K., & Jeon, C. (2023). Small tool image database and object detection approach for indoor construction site safety. KSCE Journal of Civil Engineering, 27(3), 930–939.

Lou, H., Duan, X., Guo, J., Liu, H., Gu, J., Bi, L., & Chen, H. (2023). DC-YOLOV8: Small-size object detection algorithm based on camera sensor. Electronics, 12(10), 2323.

Nath, N. D., Behzadan, A. H., & Paal, S. G. (2020). Deep learning for site safety: Real-time detection of personal protective equipment. Automation in Construction, 112, 103085.

Raditya, Y., Alunjati, F. A., Elviani, U., & Hidavat, F. (2024). Automated personal protective equipment (ppe) detection in construction environments: Enhancing safety through yolov8 object detection. 2024 International Conference on ICT for Smart Society (ICISS), 1–6.

Roboflow Universe Projects. (2024). Construction Site Safety Dataset. Roboflow. https://universe.roboflow.com/roboflow-universe-projects/construction-site-safety

Saidani, T., Ghodhbani, R., Alhomoud, A., Alshammari, A., Zayani, H., & Ben Ammar, M. (2024). Hardware acceleration for object detection using YOLOv5 deep learning algorithm on Xilinx Zynq FPGA platform. Engineering, Technology & Applied Science Research, 14(1), 13066–13071.

Shanti, M. Z., Cho, C.-S., de Soto, B. G., Byon, Y.-J., Yeun, C. Y., & Kim, T. Y. (2022). Real-time monitoring of work-at-height safety hazards in construction sites using drones and deep learning. Journal of Safety Research, 83, 364–370.

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6(1), 1–48.

Song, Y., Chen, Z., Yang, H., & Liao, J. (2025). GS-LinYOLOv10: A drone-based model for real-time construction site safety monitoring. Alexandria Engineering Journal, 120, 62–73.

Wen, S., Park, M., Lee, S., & Park, S. (2024). Automated construction safety reporting system integrating deep learning-based real-time advanced detection and visual question answering. Advances in Engineering Software, 198, 103779.

Yu, W. D., Liao, H. C., Hsiao, W. T., Chang, H. K., Tsai, C. K., & Lin, C. C. (2020, July). Automatic safety monitoring of construction hazard working zone: A semantic segmentation based deep learning approach. In Proceedings of the 2020 the 7th International Conference on Automation and Logistics (ICAL) (pp. 54-59).

Zhou, Z., Zhuo, W., Cui, J., Luan, H., Chen, Y., & Lin, D. (2025). Developing a deep reinforcement learning model for safety risk prediction at subway construction sites. Reliability Engineering & System Safety, 110885.

Zolfagharian, A., Abdellatif, M., & Briand, L. C. (2024). Smarla: A safety monitoring approach for deep reinforcement learning agents. IEEE Transactions on Software Engineering.