Uduwage, D.N.L.S., Sivanraj, S., Waidyasekara, K.G.A.S., Shiwakoti, R.K. and Tripathi. M., 2025. Detection of mobile phone use by labourers on construction sites using YOLOv11-s. In: Waidyasekara, K.G.A.S., Jayasena, H.S., Wimalaratne, P.L.I. and Tennakoon, G.A. (eds). *Proceedings of the 13th World Construction Symposium*, 15-16 August 2025, Sri Lanka. pp. 561-574. DOI: https://doi.org/10.31705/WCS.2025.42. Available from: https://ciobwcs.com/papers/

DETECTION OF MOBILE PHONE USE BY LABOURERS ON CONSTRUCTION SITES USING YOLOV11-S

D.N.L.S. Uduwage[1], S. Sivanraj[2], K.G.A.S. Waidyasekara[3], R.K. Shiwakoti[4] and M. Tripathi[5]

ABSTRACT

Mobile phone distractions among construction labourers pose significant productivity challenges. This study presents a YOLOv11-s-based model to detect and classify construction labourers who use mobile phones during work. The system was trained using a person dataset, a helmet dataset and a mobile phone dataset, obtained from an online database and custom images collected from Sri Lankan construction sites. The proposed system followed a four-stage approach, beginning with person detection, followed by helmet detection and classification. Then, through image preprocessing, the model analysed the helmet colour using histogram analysis and the Hue Saturation Value colour scale to detect labourers with yellow helmets. Subsequently, the performance evaluation metrics, such as precision-recall curve, mAP@0.5 and inference time, indicate that the trained model performs better on the testing data in detecting construction labourers who are using mobile phones during work. Finally, mobile phone detection is carried out. Images from Sri Lankan construction sites were used for deployment validation and to check for model overfitting. The system can be further developed by using motion detection through IoT to detect the continuous use of mobile phones through timeframe analysis. This study contributes to improving workplace productivity through the automated detection of distractions in construction.

Keywords: *Construction; Helmet Detection; Mobile Phone Detection; Productivity; Yolov11,*

1. INTRODUCTION

Construction labour productivity is defined as the amount of work completed per unit of labour input, typically measured in output per labour hour (Assaad et al., 2023). According to Hamza et al. (2022), it is a crucial factor in the construction industry, as it directly influences project efficiency and cost-effectiveness. However, distractions on-site can decrease productivity and lead to delays, inefficiencies, and increased operational costs. Distractions can arise from various sources (Ke et al., 2021), one of which is mobile phone usage on construction sites. Although mobile phones facilitate instant

[1] Senior Lecturer, Department of Building Economics, University of Moratuwa, Sri Lanka, nuwanthas@uom.lk
[2] Lecturer, Department of Building Economics, University of Moratuwa, Sri Lanka, sivanrajs.19@uom.lk
[3] Professor, Department of Building Economics, University of Moratuwa, Sri Lanka, anuradha@uom.lk
[4] Assistant Professor, Institute of Engineering, Tribhuvan University, Nepal, ranju.shiwakoti@ioe.edu.np
[5] PhD Student, School of ICT, Thammasat University, Thailand, d6722300024@g.siit.tu.ac.th

communication and coordination (Hasan et al., 2019), their unregulated use can become a major source of distraction due to the constant flow of notifications, messages and calls, which can easily become a source of distraction (Merchán et al., 2024).

To mitigate these issues, tighter supervision of labourers' mobile phone usage on construction sites is necessary. According to Hamza et al. (2022), supervision significantly enhances productivity. However, Raoofi et al. (2024) found that traditional supervision methods in construction can be time-consuming and costly. A case study conducted by Kim et al. (2023) in Korea highlighted that digital technologies can significantly enhance the efficiency of construction supervision. Automated systems utilising cameras and computer vision can detect, localise and classify mobile phones. The use of Artificial Intelligence (AI) in construction management can improve efficiency and reduce costs (Ruchit & Olivia, 2024).

One effective way to distinguish construction labourers from other personnel is through helmet colour detection. Studies have shown the feasibility of computer vision models in detecting labourers wearing construction helmets (Cheng, 2024; Hayat & Morgado-Dias, 2022; Li et al., 2023). Merchán et al. (2024) conducted a study to detect the use of mobile phones through an inertial measurement unit for safety monitoring. However, this approach relied on sensors embedded in helmets focusing on detections in indoor scenarios. Therefore, there is still a gap between the detection of mobile phone use in the external environment of the construction site.

This paper introduces a mobile phone detection model that uses the YOLOv11-s architecture to enhance productivity on construction sites. The model achieves this by (1) detecting persons, (2) detecting and classifying safety helmets (3) detecting the colour of helmets to differentiate labourers with yellow helmets, and (4) detecting their use of mobile phones. By integrating computer vision into construction supervision, this approach provides a solution to enhance productivity at construction sites. The paper is structured as follows. First, it provides a comprehensive review of labour productivity and computer vision for enhancing productivity. Next, the research method, including data collection and analysis techniques, is elaborated. This is followed by the findings and conclusions.

2. LITERATURE REVIEW

2.1 SIGNIFICANCE OF LABOUR PRODUCTIVITY ON CONSTRUCTION SITES

Construction productivity plays a vital role in the growth of a nation's economy (Naoum, 2016). Productivity in construction refers to the maximisation of output while optimising input utilisation, ensuring efficiency and cost-effectiveness (Hamza et al., 2022). According to Muqeem et al. (2011), construction labour is the most crucial resource in the industry, directly impacting overall construction productivity. However, low productivity levels remain one of the major challenges faced by the construction sector, leading to delays, increased costs, and inefficiencies. Dixit et al. (2019) further emphasised that low productivity can create inflationary pressure on a nation's economy, affecting overall economic stability.

One of the contributors to reduced labour productivity is mobile phone distractions among labourers (Sattineni & Schmidt, 2015). Labourers who divide their attention between critical tasks and mobile phone use are more likely to exhibit decreased efficiency, leading to lower work output and increased errors (Merchán et al., 2024).

Traditional supervision methods for maintaining productivity rely on manual monitoring, which is often time-consuming and costly (Raoofi et al., 2024). As a result, researchers are increasingly exploring technology-driven solutions to enhance productivity and automate site supervision (Adamu et al., 2024; Bai et al., 2023; Yuan et al., 2023).

2.2 COMPUTER VISION FOR ENHANCING PRODUCTIVITY AT CONSTRUCTION SITES

With the rapid growth of Artificial Intelligence (AI), computer vision has emerged as a significant tool for information perception and automated detection in various applications. In construction safety, helmet detection has been significantly enhanced using computer vision algorithms, enabling real-time monitoring of compliance on construction sites (Cheng, 2024). Numerous studies have utilised Convolutional Neural Networks (CNNs) to detect helmet usage, improving labour productivity and site supervision. Fang et al. (2018), utilised faster regional-based CNN to detect helmet use on construction sites, while Wu et al. (2019) and Han and Zeng (2022) developed deep learning-based systems that detected helmet usage and classified its colour.

CNN-based object detection models are categorised into single-stage and multi-stage detectors (Sumit et al., 2020) . Single-stage detectors, such as You Only Look Once (YOLO), perform object classification and localisation in a single step, making them significantly faster in object detection (Pham et al., 2020). In contrast, multi-stage detectors, such as Faster R-CNN and Mask R-CNN, use a two-step process, which results in higher accuracy but slower processing (Sumit et al., 2020)The primary advantage of single-stage detectors is their real-time capability, making them ideal for applications that require fast object detection. Therefore, YOLO models are an optimal choice for automated construction site supervision, which prioritises speed while maintaining a satisfactory level of accuracy.

The architecture of YOLO consists of three major components: (1) Backbone, (2) Neck, and (3) Head (He et al., 2024). The backbone is responsible for feature extraction within an image, identifying key visual patterns necessary for object detection. The neck enhances multi-scale feature fusion, enabling the model to detect objects of different sizes with improved accuracy. Finally, the head is responsible for final object classification, bounding box regression and refining detection accuracy (Alif, 2024).

2.3 YOLOV11 ARCHITECTURE

Among the YOLO series, YOLOv11 represents the latest iteration, offering significant improvements in real-time object detection (Khanam & Hussain, 2024). It incorporates a range of architectural enhancements designed to improve speed, accuracy, and adaptability. Lin (2024) proposed a YOLOv8-based algorithm designed to enhance helmet detection in complex backgrounds, including instances with small or distant objects. Furthermore, other studies have also used various YOLO models for safety helmet detection on construction sites (Wan et al., 2024; Yang et al., 2024). According to He et al. (2024), YOLOv11 demonstrated the highest accuracy and recall rate compared to previous YOLO versions. A comparison of key performance metrics of the YOLO series is provided in Table 1.

Table 1: Comparison of YOLO series
Source: (He et al. 2024)

Method	mAP/%	Precision/%	Recall/%
YOLOv5	54.4	64.5	62.6
YOLOv8	55.5	71.1	60.9
YOLOv9	43.8	55.2	50.7
YOLOv10	48.0	79.3	56.2
YOLOv11	57.2	66.4	64.8

The YOLOv11 series includes multiple variants designed for different application requirements. These variants, such as YOLOv11-n (nano), YOLOv11-s (small), YOLOv11-m (medium), YOLOv11-I (intermediate), and YOLOv11-x (extreme), offer a balance between accuracy, speed, and computational efficiency (Kishor, 2024). YOLOv11-s is particularly suitable for medium-scale applications, providing a balance between detection speed and accuracy.

2.4 SIGNIFICANCE OF THE STUDY

This study introduces a computer vision system using YOLOv11-s architecture to detect mobile phone use among construction labourers, which is a critical yet underexplored factor in construction productivity. Some studies have focused on helmet detection (Cheng, 2024; Hayat & Morgado-Dias, 2022) and indoor mobile use using sensors (Merchán et al., 2024). However, no studies have integrated person detection, helmet detection and mobile phone detection for alternative manual supervision for enhancing construction productivity. The system aims to offer a scalable, real-time alternative to manual supervision with implications for improving the productivity of labourers at construction sites.

3. METHODOLOGY

3.1 DATASET COLLECTION

Three datasets were used to train the model: (1) for person detection, (2) for "helmet" referred to as "hard hat" detection and (3) for mobile phone detection. These datasets were obtained from an online platform called RoboFlow, which provides a collection of images and associated annotations, specifically designed for training computer vision models. RoboFlow was chosen due to its high-quality and annotated datasets. The dataset consisted of construction images taken from different locations for the model development. Custom images from Sri Lankan construction sites were also incorporated for deployment validation and to identify overfitting. The images were standardised to a resolution of 640×640×3 (RGB scale) to ensure the consistency of the dataset throughout the model training, testing and validation processes.

The person dataset consisted of 5,483 images, with 80% allocated for training, 19% for validation, and 1% for testing. The hard hat dataset contained images of construction labourers with helmets (labelled as "hard hats") and no helmets (labelled as "no hard hats"). The hard hat dataset consisted of a total of 19,745 annotated images, with 70% allocated for training, 20% for validation, and 10% for testing. The mobile phone dataset consisted of 10,768 annotated images, capturing various instances of mobile phone usage by people. The dataset was divided into 82% for training, 14% for validation, and 4% for testing. These splits ensured that the model had a large enough dataset for learning, testing and validation. For this study, it was assumed that wearing a yellow hard hat is mandatory for construction labourers in Sri Lanka.

3.2 MODEL TRAINING

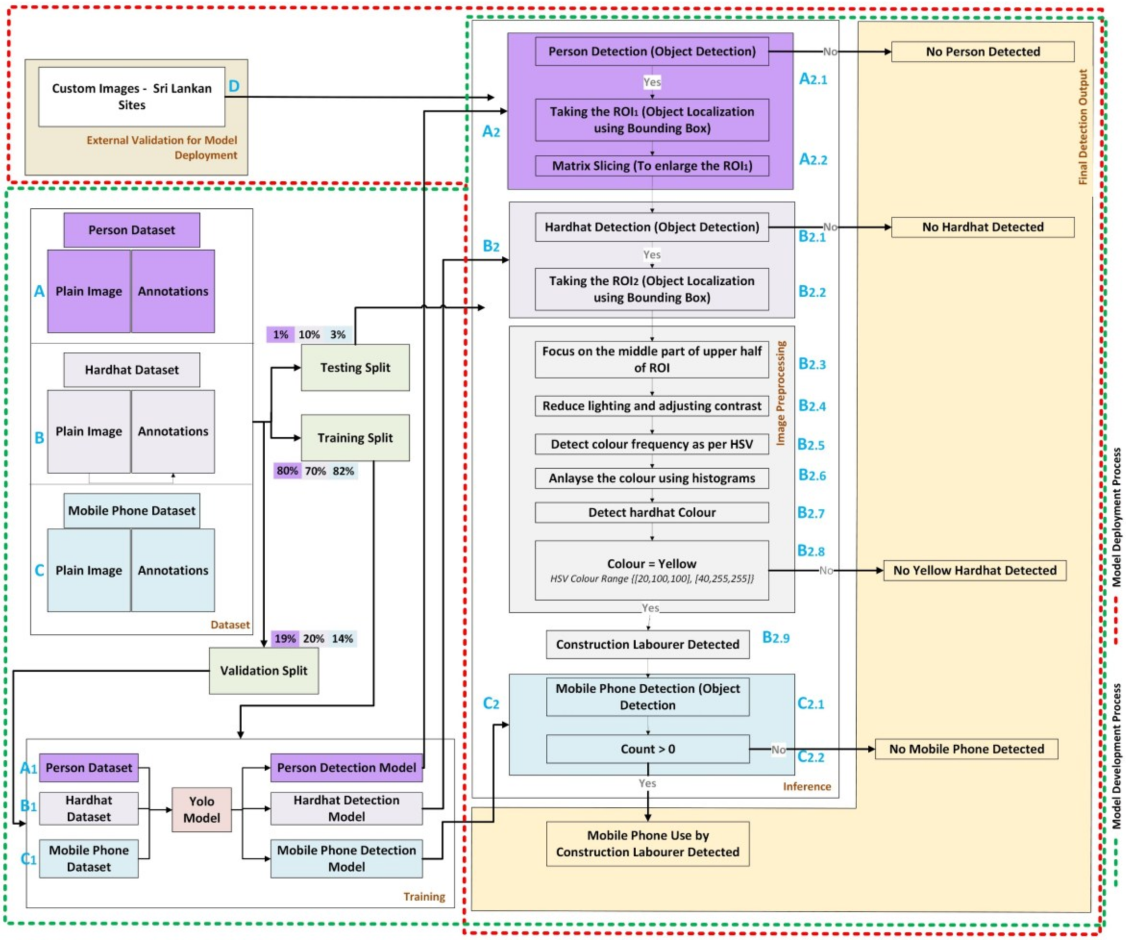

Figure 1: Overall architectural diagram of YOLOv11-s model for detection of mobile phone use by construction labourers

3.2.1 Hyperparameters

Before training, various hyperparameters are utilised to balance the model performance and computational efficiency. These hyperparameters are configuration variables that control the training process by determining the complexity, learning rate, batch size and other variables (Andonie, 2019). According to Alif (2024), key hyperparameters of the YOLOv11 models are learning rate, momentum, weight decay, batch size and the number of epochs. The learning rate of a model indicates how quickly a model learns based on the dataset, whereas momentum accelerates the training process by smoothing out updates of the model. Weight decay is a regularisation technique that prevents the model from overfitting. Batch size represents the number of training datasets processed at once. The number of epochs refers to the number of times the model processed the entire dataset.

In the model, all the hyperparameters, apart from the epochs number were preset with the values in Table 2. In this study, the model was trained for 100 epochs, allowing it to refine the internal parameters based on error calculations.

Table 2: Key parameters for YOLOv11-s training

Hyperparameter	Value
Learning Rate	0.01
Momentum	0.937
Weight decay	0.0005
Batch Size	16 images

3.2.2 Training Process

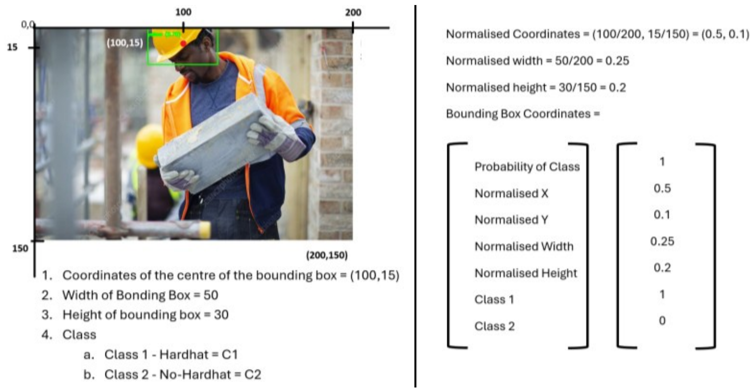

Figure 2: Bounding box coordinates

After splitting each dataset into training, validation, and testing sets, the YOLOv11-s architecture was trained in three phases, as illustrated in Figure 1. The steps have been labelled as "A" for person detection, "B" for hard hat detection, and "C" for mobile phone detection. Additionally, the model development and deployment processes are represented using two types of dotted boxes: the green-coloured dotted box indicates training, internal validation, and testing, while the red-coloured dotted box represents external validation using custom Sri Lankan images. The first phase (A1) involved identifying the region of interest (ROI) by detecting people in the image and applying matrix slicing to separate the ROI area while storing the original coordinates of the ROI within the image. During this stage, the model performed object localisation by generating bounding box coordinates, assigning a confidence score, and classifying detections as person or no-person. The bounding box coordinates were determined as shown in Figure 2. Here, if a person was detected in the image, the model was trained to process only the selected region of interest (ROI) to the next phases. The total training time of the person detection model was 2.416 hours.

Similarly, in the second phase (B1), the model was trained to detect construction hard hats and accurately localise their positions within the image, where detected objects were classified as either hard hat or no-hard hat. The total training time for the hard hat detection model was 6.827 hours. In the third phase (C1), the model was trained using the mobile phone dataset to detect and localise mobile phones. Unlike the person detection model and hard hat detection model, C1 focused solely on detection, where the model was trained to determine whether a mobile phone was present in an image, without further classification. The total training time for the mobile phone detection model was 5.192 hours.

3.3 MODEL VALIDATION (INTERNAL)

During each training epoch, the model underwent an internal validation process to assess its performance and generalisation ability. The validity of the model was determined using several metrics such as precision, recall, mean average precision (mAP), and inference time.

3.4 MODEL INFERENCE

After training, the inference process (A2 to C2) was conducted to evaluate each model's performance. As illustrated in Figure 1, the inference pipeline consists of four key components: to detect people, the presence of hard hats, the colour of hard hats and the use of mobile phones. The first component of the pipeline is the YOLOv11-s person detection model, which detects (A2.1), localises (A2.2), and slices the matrix (A2.3) to isolate ROI1 (the detected person) from the full image. This step provided the subsequent detections with enlarged ROI focusing only on relevant regions.

The sliced matrix is passed through the hard hat detection model, which detects (B2.1) and localises (B2.2) hard hats within the ROI. After localisation, the hard hat region undergoes image preprocessing (B2.3 to B2.8) to determine its colour accurately. Due to varying lighting conditions, detecting a fixed hard hat colour was challenging, as some areas of the hard hat appeared shinier or darker than others. The Red-Green-Blue (RGB) colour scale relies on fixed values, which are suitable for consistent lighting conditions. However, inconsistent lighting caused by environmental factors made the RGB-based colour scale unsuitable for this model. To overcome this limitation, this study used the Hue-Saturation-Value (HSV), which employs a range of values to define colours. By applying histogram analysis, the primary colour of the hard hat was determined based on the HSV range it belonged to. As shown in Figure 3, the central part of the upper half of the ROI was focused (B2.3), lighting and contrast were adjusted (B2.4), and its HSV values were extracted (B2.5). A histogram analysis (B2.6) was then performed to detect the colour (B2.7) and classify the hard hat as either "yellow hard hat" or "no yellow hard hat" (B2.8).

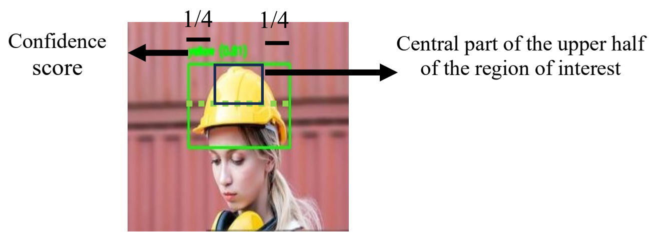

Figure 3:Region of interest

After the hard hat detection, the processed image was passed through the YOLOv11-s mobile phone detection model (C2). This model detects the presence or absence of a mobile phone (C2.1), highlighting whether a labourer is using a mobile phone in the image. Finally, the pixels in the original image were replaced with the processed image containing the detected elements. Here, the original image is displayed with the detections made.

3.5 MODEL DEPLOYMENT VALIDATION

The model was validated using images obtained from Sri Lankan construction sites (D) to assess its applicability within the Sri Lankan context and to check for potential overfitting.

4. RESULTS AND PERFORMANCE EVALUATION

4.1 CONFUSION MATRIX

A confusion matrix of 2x2 classifiers determines the predicted and actual classification. Table 3 provides the details of a 2x2 confusion matrix, which provides the relationship between true positives, false positives, true negatives and false negatives. As shown in Figure 2, the model provides a confidence score for the certainty of its detection, which plays a crucial role in identifying the false negative and false positive labels (Wenkel et al., 2021). In this study, a confusion matrix was generated only for the hard hat detection model, which performed detection and classification.

Table 3: Standard Confusion Matrix

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

Figure 4 provides the normalised confusion matrix of the hard hat detection model, which detects two classes: hard hat and no hard hat. The values are normalised to represent the proportion of the outcome instead of raw counts, which undergoe object detection and classification. The values are normalised to represent the proportion of the outcome instead of raw counts. The matrix indicates 95% accuracy for hard hat detection and 90% accuracy for no-hard hat detection, with significantly low false positive and false negative rates.

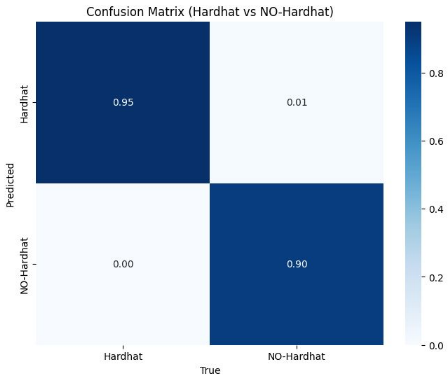

Figure 4: Normalised confusion matrix for hard hat classification

4.2 PRECISION-RECALL CURVE

Precision measures the ratio of true positive detections to the total positive detections made (see Equation 1). This reflects the ability of the model to accurately classify “hard hat” or “no-hard hat” without generating false detections. A high precision score indicates a lower rate of false detection. Meanwhile, recall measures the ratio of true positive detection to the total actual positive labels (see equation 2). It indicates how well positives are recalled. A high recall score indicates that the model is effective in detecting the relevant instances.

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

To achieve high prediction accuracy, it is crucial to maintain a balance between precision and recall. A model with high precision but low recall is very accurate in its positive predictions but misses a significant number of actual positive labels, while high recall but low precision may lead to excessive false positives. Therefore, a precision-recall (PR) curve for each detection task illustrates how well the model trades off these values.

The PR curve in Figure 5 shows the balance between precision and recall for the person detection model. Accordingly, the model achieved a high precision at a lower recall value, indicating a high probability of correct positive prediction. As the recall value increases, the false positive prediction also increases. The PR score for the person detection model was 0.826, indicating that the model correctly detects 82.6% of the actual instances. Similarly, the PR curve for hard hat detection in Figure 6 demonstrates high precision-recall scores, with 0.933 for hard hat detection and 0.927 for no-hard hat detection. The model maintains high precision across most recall values. This indicates the model can correctly classify hard hats while minimising false positives. Finally, Figure 7 presents the precision-recall curve for mobile phone detection, which also indicates a high PR value of 0.916, indicating that the model correctly detects 91.6% of the actual instances.

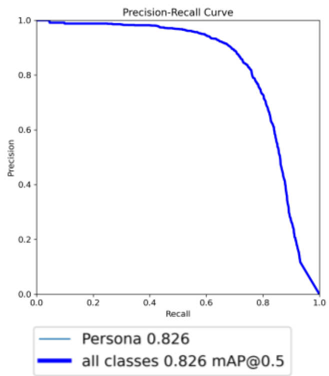

Figure 5: Precision-Recall curve for person detection

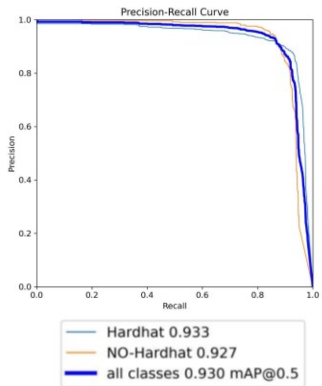

Figure 6: Precision-Recall curve for hard hat detection model

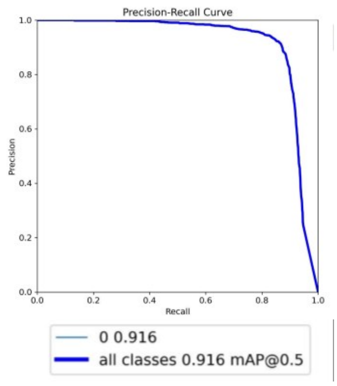

Figure 7: Precision-Recall curve for mobile phone detection

4.3 MEAN AVERAGE PRECISION (MAP)

The mAP calculates the average precision across multiple Intersections over Union (IoU) thresholds. Here, mAP@0.5 was calculated, where the average precision was calculated at a threshold of 0.5. This evaluated the accuracy of the classification.

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

According to Figures 5, 6, and 7, the mAP@0.5 of the person detection model, hard hat detection model, and mobile phone detection model were 0.826, 0.930, and 0.916, respectively. These results indicate that each model achieved a high detection accuracy, with scores exceeding the identified threshold values. The hard hat detection model showed the highest performance among the trained models. These results infer that the trained models can provide accurate detection in construction site environments.

4.4 INFERENCE TIME

Inference refers to the process of using a trained model to make predictions on new, unseen data by processing an image and generating detection results. The inference time of the model measures the average time taken by the YOLOv11-s model to process a single image and generate the detection results. The YOLOv11-s model demonstrated fast inference times, with the person detection model achieving the quickest processing of 4.9ms. The hard hat detection model took 8.5ms, which is slightly longer due to the added classification of hard hat and no-hard hat classes. Similarly, the mobile phone detection model had an inference time of 8.4ms, due to the additional processing required for detecting smaller objects. All three models demonstrated a fast-processing time which indicated that the model is suitable for real-time application.

4.5 TRAINING AND VALIDATION LOSS

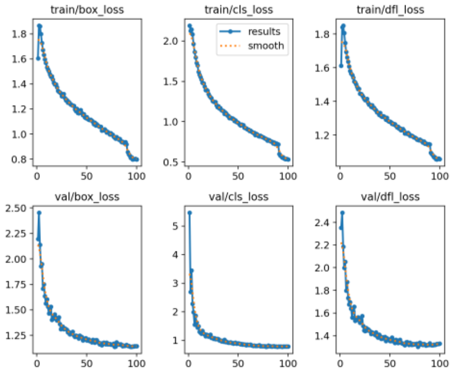

Figure 8: Training and validation loss curve for person detection

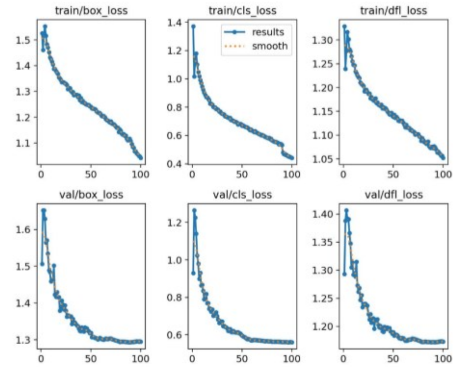

Figure 9: Training and validation loss curve for hard hat classification

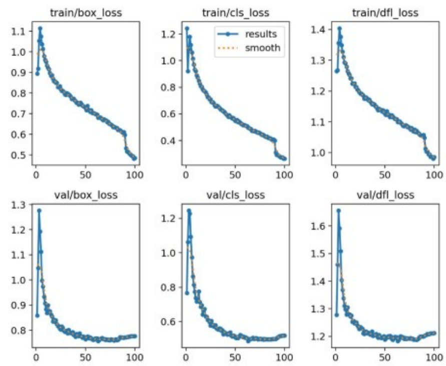

Figure 10: Training and validation loss curve for mobile phone detection

Furthermore, the loss function, also referred to as the error function, quantifies the error between the algorithm's predicted output and the actual output. The YOLOv11-s model uses a unified loss function, which is a combination of three loss components. This includes Complete Intersection over Union loss (CioU), Binary Cross Entropy Loss (BCE) and Distribution Focal Loss (DFL). CioU is used for the bounding box regression (box_loss) (Zheng et al., 2021), BCE is used for the classification (cls_loss) (Li et al., 2024)and DFL is used for better localisation accuracy by refining the predicted bounding box coordination (dfl_loss) (Wang et al., 2023).

Figures 8, 9 and 10 provide the training and validation loss curves recorded during the model's training in person detection, hard hat classification and mobile phone detection. As the number of epochs increased, the loss function steadily decreased indicating an effective learning and optimisation by the model. Similarly, the precision, recall and mAP metrics exhibited a steady increase over epochs, reinforcing the model's improving performance. The model's stable performance suggests high reliability in classifying hard hat and no-hard hat cases.

4.6 MODEL DETECTIONS

Figure 11 presents the detections performed by the model across various conditions during the testing phase. Here the detection is indicated with a bounding box and a confidence score which demonstrates the level of certainty of the model in making the detection.

4.7 MODEL DEPLOYMENT IN THE SRI LANKAN CONTEXT

The model was tested with custom images obtained from Sri Lankan construction sites, captured with cameras and mobile phones to validate its deployment performance and to check for model overfitting. The Sri Lankan construction site dataset included images taken from distant locations, where the person detection model and matrix slicing played a crucial role in enlarging the ROI for subsequent detections. After detecting the mobile phone, the processed image was refitted to the original image using its initial coordinates with detection results as shown in Figure 12 to Figure 14.

Figure 11: Construction labourers with mobile phones

Figure 12: Construction labourer with mobile phone

Figure 13: Construction labourer with blue helmet

Figure 14: Construction labourer without helmet and with mobile phone

5. CONCLUSIONS AND RECOMMENDATIONS

This study presented an analysis of the YOLOv11-s model in the detection of mobile phone use by construction labourers. The use of YOLOv11-s has demonstrated faster processing by structuring detection into four components: (1) detection of people, (2) detection and classification of hard hats, (3) image processing to detect hard hat colour and (4) mobile phone detection. Training, internal validation and testing of the model were done using datasets obtained from an online database. The model was externally validated for deployment using custom images obtained from Sri Lankan construction sites and checked for potential overfitting. With high inference speed and accuracy, the model indicated a satisfactory level of applicability to the Sri Lankan context and proved to be a solution for construction projects concerned with productivity challenges caused by mobile phone distractions among labourers. While the system was tested on Sri Lankan site images for the deployment, the methodology and model architecture are universally applicable across global construction environments. Even though the model demonstrated a satisfactory level of accuracy, the detection accuracy may vary depending on external factors such as dust levels in surroundings, weather conditions, detection distance, and the quality of captured photos, which is a limitation of this study. Therefore, future research can focus on enhancing the proposed model by training it with a large set of Sri Lankan images captured under diverse environmental conditions and lighting variations to improve its reliability. Additionally, the model can be developed to detect continuous mobile phone usage among site labourers through motion detection and timeframe analysis using the Internet of Things (IoT). This would enable the model to monitor frequent and prolonged mobile phone usage from CCTV footage, providing real-time alerts to supervisors to enhance productivity on construction sites.

6. REFERENCES

Adamu, I. I., Okanlawon, T. T., Oyewobi, L. O., Shittu, A. A., & Jimoh, R. A. (2024). Revolutionising construction safety: benefits of harnessing artificial intelligence tools for dynamic monitoring of safety compliance on construction projects in Nigeria. *International Journal of Building Pathology and Adaptation*, Vol. ahead-of-print, No. ahead-of-print. https://doi.org/10.1108/IJBPA-02-2024-0050

Alif, M. A. R. (2024). *YOLOv11 for vehicle detection: Advancements, performance, and applications in intelligent transportation systems* (arXiv:2410.22898). arXiv. https://doi.org/10.48550/arXiv.2410.22898

Andonie, R. (2019). Hyperparameter optimization in learning systems. *Journal of Membrane Computing*, *1*(4), 279–291. https://doi.org/10.1007/s41965-019-00023-0

Assaad, R. H., El-adaway, I. H., Hastak, M., & LaScola Needy, K. (2023). Key factors affecting labor productivity in offsite construction projects. *Journal of Construction Engineering and Management*, *149*(1), 4002158. https://doi.org/10.1061/jcemd4.coeng-12654

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

Cheng, L. (2024). A highly robust helmet detection algorithm based on YOLOv8 and transformer. *IEEE Access*, *12*, 130693–130705. https://doi.org/10.1109/ACCESS.2024.3459591

Dixit, S., Mandal, S. N., Thanikal, J. V., & Saurabh, K. (2019). Evolution of studies in construction productivity: A systematic literature review (2006–2017). *Ain Shams Engineering Journal*, *10*(3), 555–564. https://doi.org/10.1016/j.asej.2018.10.010

Fang, Q., Li, H., Luo, X., Ding, L., Luo, H., Rose, T. M., & An, W. (2018). Detecting non-hardhat-use by a deep learning method from far-field surveillance videos. *Automation in Construction*, *85*, 1–9. https://doi.org/10.1016/j.autcon.2017.09.018

Muqeem, S., Idrus, A., Khamidi, M.F., Ahmad, J.B., Zakaria, S.B. (2011). Construction labor production rates modeling using artificial neural network. *In Journal of Information Technology in Construction (ITcon)*, *16, 713-726.* http://www.itcon.org/2011/42

Hamza, M., Shahid, S., Bin Hainin, M. R., & Nashwan, M. S. (2022). Construction labour productivity: review of factors identified. *International Journal of Construction Management*, *22*(3), 413–425. https://doi.org/10.1080/15623599.2019.1627503

Han, K., & Zeng, X. (2022). Deep learning-based workers safety helmet wearing detection on construction sites using multi-scale features. *IEEE Access*, *10*, 718–729. https://doi.org/10.1109/ACCESS.2021.3138407

Hasan, A., Ahn, S., Rameezdeen, R., & Baroudi, B. (2019). Empirical study on implications of mobile ICT uses for construction project management. *Journal of Management in Engineering*, *35*(6), 04019029. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000721

Hayat, A., & Morgado-Dias, F. (2022). Deep learning-based automatic safety helmet detection system for construction safety. *Applied Sciences*, *12*(16), 8268. https://doi.org/10.3390/app12168268

He, Z., Kang, W., Fang, T., Su, L., Chen, R., & Fei, X. (2024). *Comprehensive performance evaluation of YOLOv11, YOLOv10, YOLOv9, YOLOv8 and YOLOv5 on object detection of power equipment. arXiv preprint. https://arxiv.org/abs/2411.18871*.

Ke, J., Zhang, M., Luo, X., & Chen, J. (2021). Monitoring distraction of construction workers caused by noise using a wearable electroencephalography (EEG) device. *Automation in Construction*, *125*, 103598. https://doi.org/10.1016/j.autcon.2021.103598

Khanam, R., & Hussain, M. (2024). *Yolov11: An overview of the key architectural enhancements*. *arXiv preprint.* https://arxiv.org/abs/2410.17725

Kim, C. W., Yoo, W. S., Seo, J., Kim, B. gun, & Lim, H. (2023). A roadmap for applying digital technology to improve the efficiency of construction supervision in building projects: Focusing on Korean cases. *Buildings*, *14*(1), 75. https://doi.org/10.3390/buildings14010075

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

Li, F., Chen, Y., Hu, M., Luo, M., & Wang, G. (2023). Helmet-wearing tracking detection based on strongsort. *Sensors*, *23*(3), 1682. https://doi.org/10.3390/s23031682

Li, Q., Jia, X., Zhou, J., Shen, L., & Duan, J. (2024). *Rediscovering BCE loss for uniform classification*. *arXiv preprint.* http://arxiv.org/abs/2403.07289

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

Merchán, M., Arquillos, A., Madrigal, J., & Gabriel, J. (2024). Use of smartphones in construction projects: Proposal for a worker monitoring system to avoid safety risks, *Safety management and human factors*, *151*, 109–114. https://doi.org/10.54941/ahfe1005307

Naoum, S. G. (2016). Factors influencing labor productivity on construction sites: A state-of-the-art literature review and a survey. *International Journal of Productivity and Performance Management*, *65*(3), 401–421. https://doi.org/10.1108/IJPPM-03-2015-0045

Pham, M.-T., Courtrai, L., Friguet, C., Lefèvre, S., & Baussard, A. (2020). YOLO-Fine: One-stage detector of small objects under various backgrounds in remote sensing images. *Remote Sensing*, *12*(15), 2501. https://doi.org/10.3390/rs12152501

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

Ruchit, P., & Olivia, M. (2024). Incorporating AI into construction management: Enhancing efficiency and cost savings. *International Journal of Science and Research Archive*, *13*(1), 1049–1058. https://doi.org/10.30574/ijsra.2024.13.1.1776

Sattineni, A., & Schmidt, T. (2015). Implementation of mobile devices on jobsites in the construction industry. *Procedia Engineering*, *123*, 488–495. https://doi.org/10.1016/j.proeng.2015.10.100

Sumit, S. S., Watada, J., Roy, A., & Rambli, D. R. A. (2020). In object detection deep learning methods, YOLO shows supremum to Mask R-CNN. *Journal of Physics: Conference Series*, *1529*(4), 042086. IOP Publishing. https://doi.org/10.1088/1742-6596/1529/4/042086

Wan, D., Deng, L., Dong, J., Liu, H., & Liu, L. (2024). An improved safety helmet detection algorithm based on YOLOv8. In *Proceedings of the 2024 IEEE 4th International Conference on Electronic Technology, Communication and Information (ICETCI)*, 811–815. IEEE. https://doi.org/10.1109/ICETCI61221.2024.10594057

Wang, Y., Jiang, F., Li, Y., Zhang, H., Wang, M., & Yan, S. (2023). Safety helmet detection algorithm for complex scenarios based on PConv-YOLOv8. In *Proceedings of the 2023 International Conference on the Cognitive Computing and Complex Data (ICCD)*, 90–94. IEEE. https://doi.org/10.1109/ICCD59681.2023.10420675

Wenkel, S., Alhazmi, K., Liiv, T., Alrshoud, S., & Simon, M. (2021). Confidence score: the forgotten dimension of object detection performance evaluation. *sensors*, *21*(13), 4350. https://doi.org/10.3390/s21134350

Wu, J., Cai, N., Chen, W., Wang, H., & Wang, G. (2019). Automatic detection of hardhats worn by construction personnel: A deep learning approach and benchmark dataset. *Automation in Construction*, *106*, 102894. https://doi.org/10.1016/j.autcon.2019.102894

Yang, M., Rao, F., Wang, L., Bai, M., & Zang, X. (2024). A safety helmet detection method using adjusted YOLOv8. *Advances in computer, signals and systems*, *8*(4), 43278. https://doi.org/10.23977/acss.2024.080402

Yuan, W., Zaixin, C., Qing, W., Huihui, J., Tao, D., & Li, C. (2023). Safety risk control system for electric power construction site based on artificial intelligence technology. In *Proceedings of the 2023 Panda Forum on Power and Energy (PandaFPE),* 1695–1699, IEEE. https://doi.org/10.1109/PandaFPE57779.2023.10140719

Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., & Ren, D. (2021). Enhancing geometric factors in model learning and inference for object detection and instance segmentation. *IEEE Transactions on cybernetics, 52*(8), 8574-8586. https://doi.org/10.1109/TCYB.2021.3095305