Automatic PPE Monitoring System for Construction Workers using YOLO Algorithm based on Deep Reinforcement Learning

Dominic Ocharo
Jomo Kenyatta University of Agriculture and Technology
P.O Box 2365-01000 Thika, Kenya

H. Chege Nganga
Jomo Kenyatta University of Agriculture and Technology
P.O Box 2365-01000 Thika, Kenya

Stephen Kiambi
Jomo Kenyatta University of Agriculture and Technology
P.O. Box 62000 – 00200, Nairobi

ABSTRACT
This paper describes an innovative system designed to detect and monitor the compliance of Personal Protective Equipment (PPE) usage within construction sites. A deep learning model is trained using the You Only Look Once (YOLO) algorithm and deployed on an IoT device to monitor employees on construction sites continuously. By seamlessly integrating advanced technologies such as deep learning methodologies, Bluetooth Low Energy (BLE) tags, Global Positioning System (GPS) modules, Fitbit devices, cameras, and sophisticated image processing algorithms, the system ensures the proper utilization of essential PPE items including safety helmets, vests, goggles, safety mask, and boots. Utilizing the different sensors affixed to PPE items and IoT devices, the system continuously emits signals for instant identification of compliance or non-compliance instances, while leveraging state-of-the-art image recognition techniques coupled with a convolutional neural network (CNN) to accurately discern PPE usage, ensuring compliance while on site. The system's efficacy was assessed based on precision, recall, and mean Average Precision (mAP) metrics, which confirmed its reliability and effectiveness in real-world operational environments. This comprehensive approach to PPE compliance monitoring signifies a significant advancement in ensuring workplace safety standards within construction sites, thereby contributing to the protection and well-being of workers in hazardous environments.

General Terms
Automatic Health and Safety System, Pattern Recognition, Algorithms, Security, Image Recognition, Artificial Intelligence, Deep Learning, IoT, and Machine Learning.

Keywords
Automatic Health and Safety System, PPE (Personal Protective Equipment), YOLO (You Only Look Once), Artificial Intelligence, Deep Learning, Image Recognition, BLE tags, Raspberry Pi, Harness, Fitbit, Convolutional Neural Network (CNN).

1. INTRODUCTION
Throughout its history, the construction industry has been plagued by significantly higher accident rates compared to other sectors [1]. This trend is primarily attributed to the high-risk nature of construction activities, exposing workers to various hazards in perilous environments. According to the United States Bureau of Labor Statistics, fatalities in the construction sector rose steadily from 985 in 2015 to 1,038 in 2018, reflecting an annual increase of 2% [2].

Despite the inherently hazardous nature of the industry, a significant number of injuries, illnesses, and fatalities can be prevented through consistent adherence to safety protocols, particularly the use of PPE such as helmets, safety glasses, gloves, and other necessary gear [3]. PPE represents the final layer of defence against accidents, with helmets playing a crucial role in protecting workers from head injuries caused by falling objects or impacts [4].

Traditionally, the manual supervision of workers wearing safety helmets and boots was the primary method of ensuring compliance. However, the extensive scope of construction activities rendered this approach inadequate for timely oversight of all workers. As a result, researchers [5] turned to machine learning and image processing technologies to automate PPE detection, aiming to enhance safety practices at construction sites.

Despite advancements in PPE detection techniques, challenges persist in achieving continuous and precise monitoring of PPE compliance among construction site staff. Conventional methods of supervision and intermittent inspections often fall short, creating opportunities for non-compliance and safety hazards. Sensor-based approaches [6], though valuable, have limitations such as the need for additional hardware and lack of visual context, which can impede their effectiveness.

The rationale for combining sensor-based and vision-based methods for monitoring PPE usage is to overcome these limitations and enhance the accuracy and completeness of assessments. Integrating sensor data with visual information offers real-time monitoring capabilities, allowing for immediate feedback and intervention if compliance issues arise. This proactive approach can prevent accidents and improve overall safety standards in the construction industry.

The main objective of this work is to develop an artificial intelligence-driven system that identifies, analyzes, and monitors PPE use and compliance in construction sites, with specific aims to develop an image recognition model, integrate BLE tags, GPS modules, Fitbit devices, cameras, and implement monitoring alerts for non-compliance. By addressing these objectives, this work seeks to contribute to the enhancement of health and safety measures for construction workers, ultimately reducing the incidence of workplace accidents and injuries.

2. LITERATURE REVIEW
2.1 Introduction
Advancements in computer science, particularly in information handling and processing, have greatly impacted the field of

construction management. Recent progress in computer vision and machine intelligence has paved the way for effective safety monitoring at construction sites [7]. The integration of Artificial Intelligence (AI) and Machine Learning (ML) into construction safety practices is gaining momentum, although research in these domains is still in its early stages. Machine learning techniques offer promising paths for analyzing data efficiently, providing timely business insights to managers. However, these techniques rely heavily on precise and extensive datasets, which can be challenging to obtain. The collection of such comprehensive datasets is crucial for accurate analysis but often poses logistical difficulties in terms of consistency and volume. In the context of PPE detection, acquiring sufficient quality data remains a significant hurdle. Among the different types of data collected from construction sites, video feeds, and images provide rich visual records of on-site activities, and computer vision-based methods assist in interpreting this visual data. Methods that have recently been proposed in the literature can be classified as sensor-based, vision-based, or ML-based.

2.2 Sensor-Based Approaches

Sensor-based approaches [6] have been deployed to detect hard hat use or Personal Protective Equipment (PPE) compliance. Zhang et al. [4] developed a system using pressure sensors to detect Non-Helmet Usage (NHU), but this method had drawbacks such as discomfort and susceptibility to false readings [4]. Sensor-based methods, like RFID portals [8] and Cyber-Physical Systems (CPS) [9], offer the advantage of personal identification and data storage, enabling recognition of workers and their behaviour. These systems vary in design, from mobile RFID portals [8] to pressure sensor-based helmet inspection systems like "Eye on Project" (EOP) [4]. While sensor-based methods have advantages, they may provide limited information compared to image-based systems and can be prone to false readings.

2.3 Vision-Based Methods

Vision-based methods, particularly those leveraging CNNs [10], have gained traction in construction safety. These methods utilize deep learning algorithms for object detection, categorizing specific objects within images or video frames. Faster algorithms like Faster R-CNN and Single Shot MultiBox Detector (SSD) offer real-time detection capabilities [11]. Despite their effectiveness, vision-based methods face challenges such as blurriness, hardware dependency, and limitations in identifying different colours of PPE items. Nonetheless, they offer significant advantages over sensor-based approaches in terms of accuracy and detail in identifying PPE usage.

2.4 ML Based Methods Used

Machine learning, particularly deep learning, plays a crucial role in safety management, enabling the development of advanced models for PPE compliance detection. CNNs are extensively used for image processing tasks, including PPE detection. Additionally, algorithms like SSD and YOLO offer efficient object detection capabilities [12], facilitating real-time monitoring of PPE usage. However, training deep learning models is computationally intensive and requires careful management of biases and variances in predictions.

2.5 Subsequent Pages

Despite the progress made in sensor-based and vision-based approaches, there are several research gaps and limitations to address. Sensor-based approaches face challenges related to sensor positioning, battery life, scalability, and limited information provided. Vision-based methods encounter issues with blurriness, hardware dependency, and the need for diverse datasets. Furthermore, both approaches lack comprehensive datasets for PPE detection, hindering the development of accurate and generalizable models. Additionally, privacy concerns and the slow adoption of new technologies pose challenges to the implementation of automated safety systems in construction settings. These research gaps underscore the need for further exploration and innovation in safety monitoring technologies for construction sites.

Recent research has suggested the use of machine learning (ML) techniques to monitor construction safety, emphasizing the requirement for more sophisticated and reliable solutions. CNNs have demonstrated potential in accurately detecting and classifying objects about PPE compliance. Techniques like Faster R-CNN, SSD, and YOLO [13] give real-time detection capabilities for dynamic construction contexts [14,15]. However, research gaps remain, such as the absence of comprehensive datasets for training and verifying ML models, and the difficulties in integrating ML models with existing safety management systems. Furthermore, it is imperative to acknowledge and resolve ethical and privacy issues that arise from the implementation of machine learning-powered monitoring systems in construction sites. Continued research is necessary to provide stronger, more precise, and morally acceptable solutions, despite the advances made so far [16].

3. METHODOLOGY

3.1 Introduction

This section presents the methodology for a real-time monitoring system designed to ensure PPE compliance at construction sites. The system utilizes a PPE image dataset, deep learning frameworks, and IoT devices equipped with a YOLOv8 model for object detection. BLE tags are employed for accurate tracking, and the MQTT protocol is used for efficient data transmission. The setup also incorporates Fitbit for health monitoring and a GPS module for real-time location tracking, enhancing worker safety monitoring.

3.2 Required Materials

a) PPE image dataset: A comprehensive collection of images depicting all pertinent PPEs (including safety hard hats, safety vests, goggles, safety gloves, and boots) applicable to construction site environments.
b) Software labelling tools: Essential tools facilitating the accurate labelling of images within the dataset.
c) Frameworks for deep-learning software: Including PyCharm, TensorFlow, and requisite libraries for image dataset labelling.
d) BLE tags: Utilized for precise detection of PPE compliance post-authorization for site access.
e) Image-processing platform: Capable of expeditiously and effectively evaluating PPE compliance, such as the DCP platform.
f) IoT device: IoT devices such as Raspberry Pi or Nvidia Jetson Nano deploy the trained model for detection.
g) Fitbit: Used to monitor health-related metrics, such as heart rate and blood pressure, to confirm the worker's fitness for the assigned tasks.
h) GPS Module: To provide real-time location details, including latitude and longitude.

3.3 Vision-Based Methods

The flowchart in Figure 1 illustrates the process of real-time detection of workers in the field to ensure they are adhering to safety measures by wearing the required Personal Protective

Equipment (PPE). The system leverages a trained YOLOv8 model deployed on an IoT device, such as a Raspberry Pi or Nvidia Jetson, and an attached camera to achieve this.

1. Start: The process begins with the system initialization.
2. Camera: The camera, attached to the IoT device, captures images of the workers at regular intervals.
3. IoT Device (Raspberry Pi/Nvidia Jetson):
   - Object Detection Model: The captured images are sent to the YOLOv8 object detection model deployed on the IoT device.
   - Real-Time Detection: The YOLOv8 model, known for its speed and accuracy, processes the images in real-time to detect whether workers are wearing the required PPE.
4. Detection Results: The results from the detection model are then forwarded to an MQTT broker.
5. MQTT Broker: Message Queuing Telemetry Transport (MQTT) is a lightweight messaging protocol for small sensors and mobile devices optimized for high-latency or unreliable networks. The MQTT broker facilitates the efficient and reliable transfer of detection results from the IoT device to the server.
6. Server:
   - The server receives the detection results from the MQTT broker.
   - PPE Violation Check: The server processes these results to determine if there is a PPE violation.
7. Decision Point - PPE Violation?
   - Yes: If a PPE violation is detected, the system generates notifications, such as alarms, to alert the authorities. This can also include access control measures to restrict entry or movement.
   - No: If no violation is detected, the system generates notifications confirming compliance.
8. Store Data for Analysis: Regardless of the PPE compliance status, the results are stored for further analysis. This data can be used for reporting, compliance monitoring, and improving workplace safety protocols.
9. Stop: The process ends here before it restarts, continuing the cycle of monitoring and detection.

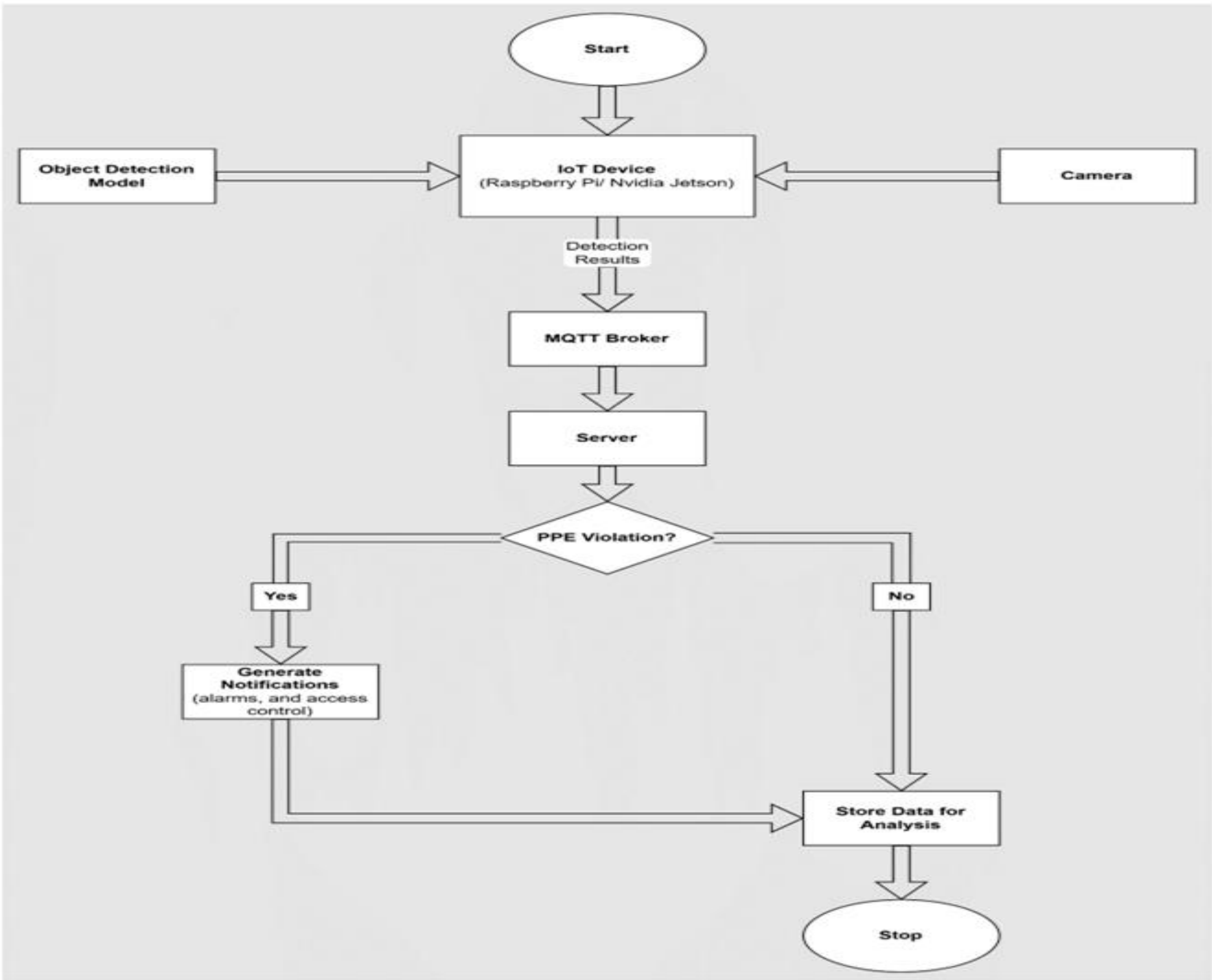

Fig 1: Flow chart of real-time detection

3.4 Dataset Preparation

The most important part of training the machine learning algorithm was the collection and preparation of data to facilitate the validation of the model. The preparation of the dataset is the most time-consuming and critical component, enabling efficient training and accurate detection by the algorithm. Data was collected from online image scraping and the Pavicon Kenya Limited database. Once the dataset was collected, the data was labelled using Roboflow, and annotations were saved as TXT files.

3.5 Dataset Partitioning

The dataset was partitioned into three subsets: training, testing, and validation datasets. The training dataset was used to train the CNN and refine its internal parameters. The testing dataset served the purpose of assessing the model's performance and making fine adjustments to its settings. Meanwhile, the validation data assisted in the selection of the most suitable model and prevented overfitting in the future.

4. RESULTS DISCUSSION

4.1 Introduction

This segment delves into crucial aspects of performance evaluation, providing a comprehensive understanding of precision, recall, mAP50, training and validation box losses, and F1-confidence results across various classes. Each metric plays a significant role in assessing the model's effectiveness as a whole, guiding potential adjustments to improve accuracy in practical personal protective equipment detection application scenarios such as construction sites.

4.2 Performance Evaluation

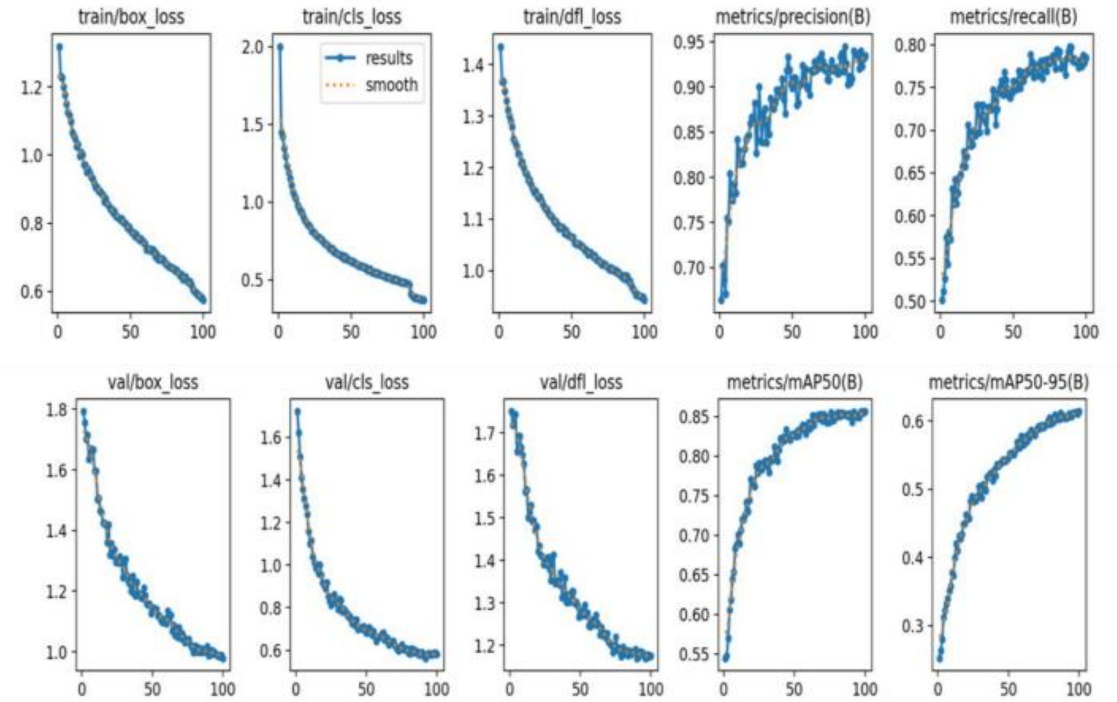

Fig 2: Performance Evaluation

The performance evaluation of the YOLO model for PPE compliance detection over 100 epochs reveals significant insights through various graphs representing training and validation losses and metrics. The training loss graphs, which include train/box_loss, train/cls_loss, and train/dfl_loss, show a consistent decrease across epochs as in Figure 2. The train/box_loss graph indicates that the model improves in predicting the bounding box coordinates for detected objects, as evidenced by the steadily declining loss values. Similarly, the train/cls_loss graph demonstrates a significant reduction in classification loss, reflecting the model's increasing accuracy in classifying detected objects like safety helmets and vests. The train/dfl_loss graph further supports this trend, with the focal loss decreasing steadily, indicating enhanced model confidence and precision in object detection.

The training metrics for precision and recall provide additional insights into the model's performance. The metrics/precision(B) graph shows an upward trend and stabilization at higher values, suggesting the model's effectiveness in minimizing false positives. Concurrently, the metrics/recall(B) graph displays an increase over the epochs, signifying the model's improved ability to capture actual positive instances. Together, these metrics indicate that the model is becoming more proficient at correctly identifying and classifying PPE items, with fewer instances missed or falsely detected.

Validation loss graphs, including val/box_loss, val/cls_loss, and val/dfl_loss, mirror the training loss trends, showcasing the model's generalization capabilities. The val/box_loss graph shows a decreasing trend, similar to the training phase, indicating accurate bounding box predictions on unseen data. The val/cls_loss graph, while showing some fluctuations, overall decreases, suggesting improvements in classification accuracy on validation data. The val/dfl_loss graph consistently decreases, reinforcing the model's growing confidence and accuracy in predictions.

Metrics for validation performance, particularly metrics/mAP50(B) and metrics/mAP50-95(B), highlight the model's detection capabilities. The metrics/mAP50(B) graph demonstrates an improvement and stabilization at higher values, reflecting enhanced detection performance with a good balance between precision and recall at a moderate overlap threshold. The metrics/mAP50-95(B) graph, which considers multiple IoU thresholds, also shows a steady increase, albeit lower than mAP50. This indicates that the model performs robustly across varying levels of localization strictness, improving overall detection accuracy and reliability.

Overall, the graphs collectively indicate that the YOLOv8 model effectively learns and enhances its performance across training and validation datasets. The steady decrease in loss values suggests better prediction accuracy, while the increasing precision and recall metrics affirm the model's reduced false positives and higher detection rates. The mAP metrics further confirm the model's reliability and robustness, demonstrating its potential to ensure PPE compliance on construction sites and contribute to improved safety and monitoring standards.

4.3 Results

The practical application of the YOLO model was tested in real-world scenarios, as shown in the figures below. In these images, workers on a construction site are monitored for PPE compliance using the trained model. In Figure 3, the uploaded image shows workers at a construction site, and the detected image highlights the model's ability to accurately identify hard hats and safety vests, with confidence scores provided for each detection. Prohibited items such as the absence of masks and safety vests are flagged, demonstrating the model's capability to detect non-compliance effectively.

Fig 3: Example of PPE compliance detection in a telecommunication site using YOLOv8

In Figure 4, another set of workers is analyzed. The uploaded image is compared with the detected image, where the model successfully identifies hard hats and safety vests. The model's detections are annotated with confidence scores, showing high accuracy. Again, prohibited items such as the absence of masks are flagged, ensuring that safety protocols are monitored and enforced.

Fig 3: Example showcasing the model's accuracy in identifying PPE items and flagging non-compliance

These results illustrate the practical utility of the YOLOv8 model in real-time PPE compliance monitoring, providing reliable detections and actionable insights to improve workplace safety.

5. CONCLUSIONS

In conclusion, the thorough evaluation of the YOLOv8 object detection model for PPE detection yields substantial insights into its effectiveness and potential implications. The observed positive trend in precision metrics, demonstrating commendable improvement throughout training and stabilization emphasizes the model's heightened accuracy and precision. This improved precision is crucial for real-world applications, ensuring a reliable system capable of accurately identifying issues related to personal protective equipment removal and non-compliance.

The model's proficiency in distinguishing between properly worn PPE and those with potential concerns (those with a good probability of not wearing proper PPE) implies its capability to provide precise recommendations for targeted corrective actions. Furthermore, the consistent improvement in recall performance, as more epochs are used in model training, signifies the model's evolving competence in capturing a higher percentage of actual positive instances related to PPE compliance detection.

These findings collectively depict the YOLOv8 model progressing towards higher accuracy and reliability, establishing a robust foundation for delivering effective and actionable recommendations in the context of personal protective equipment. The successful deployment of the YOLOv8 object detection model further underscores its adaptability and practical applicability. This successful deployment highlights the YOLOv8 model's immediate practical utility in addressing challenges in PPE compliance, making strides toward revolutionizing safety practices.

Practically, these conclusions hold significant implications for safety technology and compliance monitoring. The heightened accuracy and reliability of the YOLOv8 model empower users with a potent tool for early detection of issues related to personal protective equipment. This capability facilitates prompt and targeted interventions, contributing to improved workplace safety and overall adherence to safety protocols. The findings underscore the potential of deep learning-powered solutions in addressing challenges in PPE monitoring, providing a scalable and efficient approach for ensuring safety in various industries.

In industrial settings, the model can swiftly detect and alert operators to potential hazards, ensuring a proactive response to prevent accidents. In the context of personal protective equipment (PPE), the YOLOv8 model excels in recognizing compliance issues, ensuring that safety protocols are adhered to with precision. Whether in construction sites, manufacturing facilities, or healthcare institutions, the YOLOv8 model serves as an invaluable guardian, continually scanning its surroundings to identify and address safety concerns promptly.

The model's ability to operate in real-time and adapt to diverse scenarios positions it as a key player in fostering a safer and more secure environment for workers and stakeholders alike. As technology in safety monitoring continues to advance, the success of the YOLOv8 model represents a notable stride towards leveraging technology for enhanced efficiency and safety in global workplace environments.

6. ACKNOWLEDGMENTS

Gratitude is extended to Pavicon Kenya Limited Company for providing the resources and funding necessary for this project. Their support was instrumental in facilitating this research. Appreciation is also conveyed to Jomo Kenyatta University of Agriculture and Technology for granting the right to work on this project and for the invaluable guidance provided by H. Chege Nganga and Dr Stephen Kiambi, whose mentorship was crucial to the successful completion of this work.

7. REFERENCES

[1] 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, pp. 3478, (2021).

[2] Y. Shim, “Comparative Analysis of Fatalities Per 100,000 Workers in the Construction Industry Among Developed Countriesusing Multi-Criteria Evaluation”, (2024).

[3] A. Hayat and F. Morgado-Dias, “Deep Learning-Based Automatic Safety Helmet Detection System for Construction Safety”, Applied Sciences, vol. 12, no. 16, pp. 8268, (2022).

[4] H. Zhang, X. Yan, H. Li, R. Jin, and H. Fu, “Real-Time Alarming, Monitoring, and Locating for Non-Hard-Hat Use in Construction”, Journal of Construction Engineering and Management, vol. 145, no. 3, pp. 04019006, (2019).

[5] N. D. Nath and A. H. Behzadan, “Deep Learning Detection of Personal Protective Equipment to Maintain Safety Compliance on Construction Sites”, Nov. 2020.

[6] S. Márquez-Sánchez, I. Campero-Jurado, J. Herrera-Santos, S. Rodríguez, and J. M. Corchado, “Intelligent Platform Based on Smart PPE for Safety in Workplaces”, Sensors, vol. 21, no. 14, pp. 4652, (2021).

[7] J. Seo, S. Han, S. Lee, and H. Kim, “Computer vision techniques for construction safety and health monitoring”, Advanced Engineering Informatics, vol. 29, no. 2, pp. 239–251, (2015).

[8] A. Kelm et al., “Mobile passive Radio Frequency Identification (RFID) portal for automated and rapid control of Personal Protective Equipment (PPE) on construction sites”, Automation in Construction, vol. 36, (2013).

[9] K.-J. Park, R. Zheng, and X. Liu, “Cyber-physical systems: Milestones and research challenges,” Computer Communications, vol. 36, no. 1, pp. 1–7, (2012).

[10] Z. Jiang and J. I. Messner, “Computer Vision-Based Methods Applied to Construction Processes: A Literature Review”, Construction Research Congress 2020, (2020).

[11] A. Kumar, Z. J. Zhang, and H. Lyu, “Object detection in real-time based on improved single shot multi-box detector algorithm”, EURASIP Journal on Wireless Communications and Networking, vol. 2020, no. 1, (2020).

[12] Y. Yin, H. Li, and W. Fu, “Faster-YOLO: An accurate and faster object detection method”, Digital Signal Processing, vol. 102, pp. 102756, (2020).

[13] A. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection”, Cv-foundation.org, pp. 779–788, (2016).

[14] A. S. Chen and K. Demachi, “A Vision-Based Approach for Ensuring Proper Use of Personal Protective Equipment (PPE) in Decommissioning of Fukushima Daiichi Nuclear Power Station”, Applied Sciences, vol. 10, no. 15, pp. 5129, (2020).

[15] A. J. Wu, N. Cai, W. Chen, H. Wang, and G. Wang, “Automatic detection of hardhats worn by construction personnel: A deep learning approach and benchmark dataset”, Automation in Construction, vol. 106, pp. 102894, (2019).

[16] A. S. Chi, C. H. Caldas, and D. Y. Kim, “A Methodology for Object Identification and Tracking in Construction Based on Spatial Modeling and Image Matching Techniques,” Computer-Aided Civil and Infrastructure Engineering, vol. 24, no. 3, pp. 199–211, (2009).

[17] A. S. Zhang, J. Teizer, N. Pradhananga, and C. M. Eastman, “Workforce location tracking to model, visualize and analyze workspace requirements in building information models for construction safety planning,” Automation in Construction, vol. 60, pp. 74–86, (2015).