CCC 2025 **Proceedings of the Creative Construction Conference** (2025)
Edited by: Miroslaw J. Skibniewski & Miklós Hajdu & Žiga Turk
https://doi.org/10.22260/CCC2025/0031

COMPUTER VISION AND LARGE LANGUAGE MODEL-BASED SAFETY MANAGEMENT FOR CONSTRUCTION PROJECT SITES

Bin Tang[1,2,3], Hanbin Luo[1,2]
1 School of Civil and Hydraulic Engineering, Huazhong University of Science and Technology, Wuhan, China
2 National Center of Technology Innovation for Digital Construction, Huazhong University of Science and Technology, Wuhan, China
3 Institute of Artificial Intelligence, Huazhong University of Science and Technology, Wuhan, China

Abstract

Construction sites are inherently hazardous environments where accidents and fatalities constitute a significant global concern. Despite continuous efforts to enhance safety, the incidence of workplace accidents within the construction industry remains elevated. This study introduces HUST AI Box, an AI and computer vision-based safety management system, enhanced by the integration of DeepSeek, a large language model (LLM), to provide advanced natural language processing (NLP) capabilities for real-time decision support, knowledge management, and safety optimization. The integration of DeepSeek enables the system to process unstructured text data and provide actionable insights, significantly enhancing the safety management framework. We demonstrate the localized deployment of DeepSeek within the HUST AI Box system, ensuring robust performance and real-time responsiveness in dynamic construction environments. The results suggest that the combined system significantly enhances safety management by automating the detection of unsafe behaviours, improving the efficiency and precision of face recognition, and leveraging DeepSeek for real-time decision support and knowledge management. This system is versatile and can be applied in various other contexts beyond construction sites.

Keywords: Artificial intelligence (AI), Computer vision, Safety management, Construction sites, Large language model (LLM), Localized deployment

Peer-review under responsibility of the scientific committee of the Creative Construction Conference 2025.

1. Introduction

Construction sites remain one of the most dangerous work environments globally, contributing to a high incidence of workplace accidents and fatalities [1]. Accidents and fatalities have always been a worldwide problem on construction sites. Ensuring the safety of personnel during construction is still a pervasive challenge. However, despite ongoing efforts to improve safety standards, the construction industry continues to account for a significant portion of workplace injuries and deaths. [2,3]. According to the International Labour Organization (ILO), construction workers represent approximately 7% of the global workforce but account for 30–40% of all workplace fatalities [4]. In the European Union alone, more than 20% of fatal industrial accidents occur within the construction sector [5]. These statistics highlight the urgency of improving safety management practices to prevent further loss of life.

A key factor contributing to these accidents is unsafe worker behaviour and poor working conditions. Research indicates that more than 90% of construction site accidents are caused by human error, such as non-compliance with safety protocols, lack of proper personal protective equipment (PPE), or unsafe working conditions. Given the complexity and dynamic nature of construction sites, it is often difficult for safety managers to oversee all aspects of site safety manually, particularly as sites grow in size and personnel numbers [6-8]. However, despite the promise of these technologies, several challenges

Corresponding author email address: thomastang@hust.edu.cn

remain in their practical application. Many existing systems focus on narrow, single-task safety monitoring, such as helmet detection or PPE compliance, without integrating multiple safety parameters into a unified framework. Furthermore, the high variability of construction site environments—such as changes in lighting, weather conditions, and worker movements—poses significant challenges for ensuring reliable performance of computer vision algorithms. In particular, many face recognition systems require workers to follow specific actions, such as taking off glasses or blinking, which is impractical in fast-paced construction settings.

The existing literature also highlights several gaps in the field of automated construction safety management. First, while many studies have proposed systems for detecting specific unsafe behaviours, few have developed integrated systems that combine multiple safety checks (e.g., helmet, reflective clothing, safety belt) alongside environmental hazard detection (e.g., open flame detection). Second, most existing systems rely on highly controlled environments or require substantial setup, such as the installation of multiple cameras or extensive calibration, making them difficult to deploy on large, dynamic construction sites. Third, face recognition technologies in construction safety management often require workers to adhere to rigid pose requirements or specific actions, limiting their effectiveness in uncontrolled environments.

In this context, we present a new safety management system "HUST AI box" for construction sites, enhanced with DeepSeek, a large language model (LLM), to provide advanced NLP capabilities for real-time decision support, knowledge management, and safety optimization.. This study covers the recognition of safety helmets, reflective clothing, safety belts, open flames, and personnel presence during construction activities. Personnel detection presents challenges, particularly given the high frequency with which workers enter and exit the site. With multiple workers involved, safety management methods must ensure smooth and efficient site operations. To address this, we employ face recognition technology to verify the identity of individuals. This method is cost-effective, difficult to counterfeit, and widely accepted as an ideal solution for safety management. Moreover, this system is versatile and can be applied in various other contexts beyond construction sites.

2. Safety management system architecture

2.1. Framework of the system

The construction industry is notorious for its high rates of workplace accidents, injuries, and fatalities, making effective safety management crucial. Traditionally, safety management was carried out through paper-based documentation, a system that is prone to inefficiencies and errors, such as loss of records, negligence, and miscommunication. Due to these limitations, the construction industry has increasingly turned to digital safety management systems that rely on technologies such as cameras and artificial intelligence (AI) algorithms for enhanced oversight and more effective management.

The proposed safety management system focuses on automating the identification and monitoring of key safety violations on construction sites. This includes detecting whether workers are wearing safety helmets, reflective clothing, safety belts, and monitoring for other hazards like open flames or personnel presence. The system also tracks behaviours such as unauthorized solo operation, unsafe gatherings, personnel falls, and prolonged inactivity (which could indicate fatigue or safety risks). Table 1 summarizes the specific functions of the system, which collectively aim to reduce human error and improve the reliability and efficiency of safety management.

Table 1. Safety management system functions.

Function	Description
Safety helmet recognition	Automatically identify whether a person is wearing a helmet
Reflective clothing recognition	Automatically identify whether a person is wearing reflective clothing
Safety belt recognition	Automatically identify whether a person is wearing a safety belt
Smoke recognition	Automatically identify whether a person is smoking
Open flame recognition	Automatically identify the presence of open flames

Personnel detection	Automatically identify whether a person belongs to the specified construction site
Lone worker detection	Automatically monitor whether workers are operating alone in restricted areas
Personnel gathering monitoring	Automatically monitor the presence of multiple people in a designated area (number of people can be configured)
Personnel fall recognition	Automatically identify the fall behaviour of the person and generates an alarm when the fall state lasts for a specified period of time
Sleep monitoring	Automatically detect if a worker is still or in a potentially dangerous posture for long

By integrating computer vision with AI, the system moves beyond traditional methods and provides real-time, automated monitoring. This greatly reduces the risk of human error and enhances the speed and accuracy of safety oversight on construction sites.

2.2. Face recognition algorithm

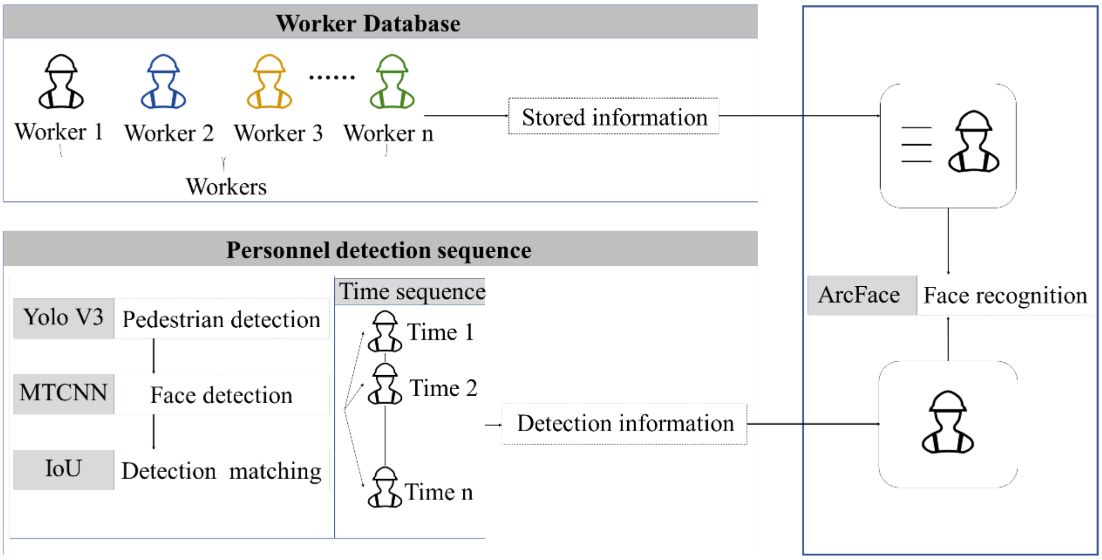

Fig. 1. Process of face recognition for safety management on construction sites.

This study proposes a safety management system for construction sites based on face recognition, and the step of face recognition mainly includes the following parts, i.e. personnel detection sequence, face recognition & personnel judgment and multi-person tracking. The designed process of the system of face recognition for safety management on construction sites is shown in Fig. 1.

The generation of personnel detection sequence includes three steps, i.e. pedestrian detection, face detection and detection matching. Pedestrian detection is to reduce the amount of undetected workers and determine whether pedestrians exist in the images and to give accurate positioning, which helps to implement face detection precisely. The result of pedestrian detection provides a restrictive area for face detection. The purpose of face detection is to acquire the image of the workers in order to prepare the material for face recognition. Detection matching can match the information pedestrian detection and face detection to reduce the rate of error detection, which is based on IoU (intersection-over-union). The possibility of mistaken identification will be improved without the application of detection matching, on account of detecting other objects which has similar features with faces, such as pictures with faces, wheels, etc. The personnel detection sequence is generated, which can prevent repeated alarm so that improve the efficiency and availability of the proposed system. The image that best completes the requirements and the information in the database of workers' identity are compared in the process of face recognition, by which personnel judgment can be implemented. The application of multi-person tracking can apply to multi-person scenes for improving the pass-ability. The system can display whether each worker can pass the access control or not, respectively.

The ROI monitoring area is drawn at random according to the construction site. Pedestrian detection is proceeded in the ROI area, which uses YOLOv3 to train. Because of the camera images for the twodimensional information, we cannot determine the position of accurately by the boxes of pedestrians. Therefore, the positions of pedestrians' heads in the beginning and the end image frames are regarded as a symbol of people going in and out of the ROI area. YOLOv3 was changed from YOLOv2, which is an effective method in deep learning and shows a better performance. Successive 3×3 and 1×1 convolutional layers are used in the network of YOLOv3. As shown in Fig. 2, the network has 53 convolutional layers, which is called Darknet-53.

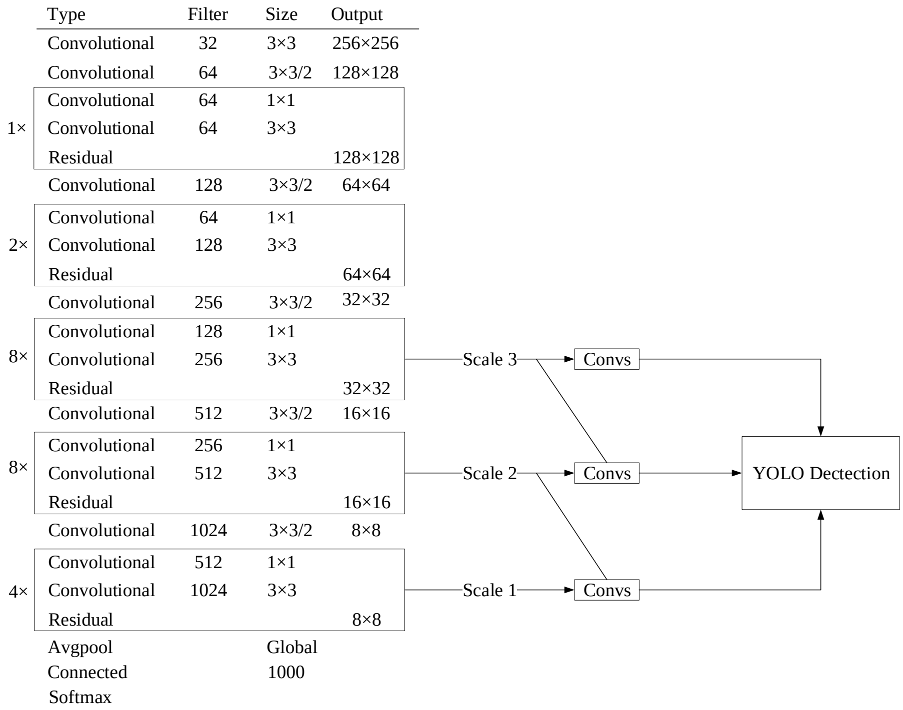

Fig. 2. Architecture of Darknet-53.

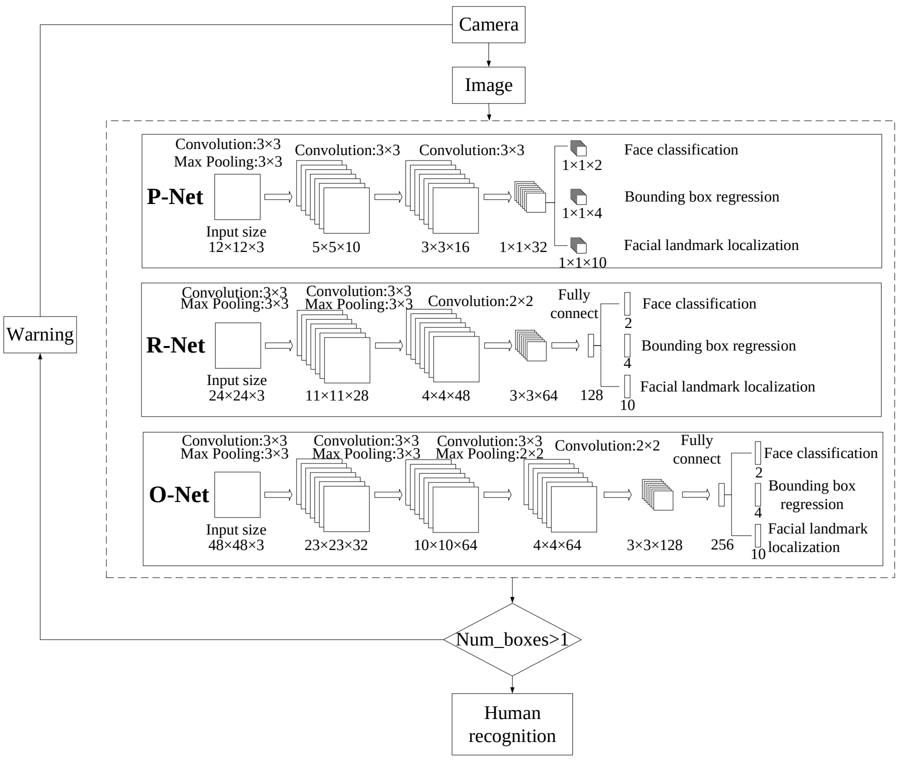

Fig. 3. Architectures of P-Net, R-Net, and O-Net. The step size in convolution and pooling is 1 and 2, respectively.

Since the scene of face detection at the gate is of low complexity and the person to be recognized is usually facing the camera center, the Multi-task Cascaded Convolutional Network (MTCNN) can be adopted in this study. MTCNN is composed of three sub-networks, respectively P-Net, R-Net and O-Net. The basic functions of the three networks are as follows: face classification, bounding box regression and facial landmark localization. The network structure is shown in Fig. 3. The three tasks of face classification, bounding box regression and facial landmark localization are used to train the CNN detectors.

Detection matching is to match the information of pedestrian detection and face detection. The process of detection matching is based on IoU (intersection-over-union). IoU is a standard that measures the accuracy of detecting an object in a particular dataset. IoU is defined as:

$$\mathrm{IoU} = \frac{Area\ of\ Overlap}{Area\ of\ Union}$$

IoU describes the correlation between a ground-truth bounding box and predicted bounding box. We use IoU to judge if the results of pedestrian detection and face detection correspond to each other. In general, IoU>0.5 is considered to be a good performance. Then we can get personnel detection sequence according to the time sequence. The best image of the time sequence will be selected, which is used to be the final detection image.

3. System development and application

3.1. Project overview

The illustrative case study is application for Wuhan Metro Line 12. With the rapid growth of the modern economy, the number of personnel employed in metro enterprises—critical to driving social and economic development—has significantly increased. However, due to the inherently hazardous nature of construction processes, which involve sedimentation, water gushing, shield collapse, the demand for robust safety management has grown.

This growing complexity in workforce management and the critical need for safety have made tasks such as safety helmet recognition, reflective clothing detection, safety belt monitoring, open flame detection, and personnel tracking urgent priorities. Given the large scale of the project and the increasing difficulty in manual safety oversight, the need for an automated, intelligent safety management system has become apparent. An essential component of this safety management system is face recognition. The project enforces strict access control procedures, and the use of face recognition enables automatic verification of personnel qualifications before they are allowed access to the site. This ensures that only certified and trained workers are allowed entry, enhancing both security and safety management efficiency. The prototype of HUST AI box is shown in Fig. 4. The application of the safety system is shown in Fig. 5.

Fig. 4. Prototype of HUST AI box.

Fig. 5. Application of the safety system for Wuhan Metro Line 12 project.

3.2. System functions and modules

3.2.1 Personnel detection

Personnel detection serves as a preliminary step for other modules within the system. The process of personnel detection, based on face recognition, involves the following steps. Initially, images are extracted from video footage. Subsequently, pedestrian detection, face detection, and face alignment are performed. Once the personnel time sequence is generated, the optimal image is selected for face recognition. Additionally, multi-person tracking is utilized to allow several workers to pass simultaneously. The system employs a deep learning algorithm for face recognition on construction sites. The images from pedestrian detection are input into YOLOv3 to extract pedestrian bounding boxes, which are then used as input for MTCNN. This allows for the acquisition of face detection and face alignment results. Finally, face recognition is conducted using ArcFace to compare the face image with pre-recorded worker photos. Based on the results, the system determines whether a worker is authorized to pass the access control. If unauthorized personnel enter the construction site, the system triggers an alarm to alert the manager to intervene promptly. Illegal entry attempts are recorded, while authorized personnel are granted access. This component is the core of the developed system, incorporating the central algorithm of this study. Statistical analysis of face recognition enables effective assessment and management. The interface for face recognition is shown in Fig. 6. Pedestrian detection results are displayed as green or red boxes, with green boxes indicating authorized workers and red boxes for others. Face detection boxes are blue, and facial landmark localization is represented by five pink points.

Fig. 6. Interface of face recognition.

3.2.2 Personnel detection

In addition to personnel detection, the safety management system includes a variety of other functions and modules designed to enhance overall site safety. These modules work in conjunction with the personnel detection system to provide comprehensive safety monitoring.

Safety Helmet Recognition: This module automatically detects whether workers are wearing safety helmets. By analyzing the video feed, the system can identify the presence or absence of helmets and alert supervisors if any workers are not in compliance with safety regulations.

Reflective Clothing Recognition: Similar to helmet recognition, this module checks for the presence of reflective clothing. Reflective clothing is crucial for visibility in low-light conditions, and the system ensures that all workers are wearing it as required.

Safety Belt Recognition: The system also monitors the use of safety belts, particularly in areas where workers are at risk of falls. It can detect whether workers are wearing safety belts and alert supervisors if any are missing.

Smoke and Open Flame Detection: These modules are designed to detect potential fire hazards. By analyzing the video feed for signs of smoke or open flames, the system can alert personnel to take immediate action to prevent fires.

Personnel Gathering Monitoring: This module monitors the number of people in designated areas to prevent overcrowding, which can lead to safety hazards. It can be configured to alert supervisors if the number of people exceeds a set threshold.

Lone Worker Detection: In some situations, it is unsafe for workers to operate alone. This module detects when a worker is operating in a restricted area without others present and alerts supervisors to ensure their safety.

Personnel Fall Detection: The system can identify fall behaviours and generate an alarm if a worker remains in a fallen state for a specified period of time. This helps in quickly responding to potential injuries.

Sleep Monitoring: This module detects if a worker is still or in a potentially dangerous posture for an extended period, which could indicate fatigue or other safety risks.

Fig. 7 displays the interfaces of some of these modules, which were captured in a laboratory environment. These interfaces provide a visual representation of the system's capabilities and how it integrates various safety monitoring functions into a cohesive system. By combining these modules with the personnel detection system, the safety management system offers a comprehensive solution for enhancing safety on construction sites. It not only monitors individual worker compliance but also detects environmental hazards and ensures that all safety protocols are being followed. This integrated approach significantly reduces the risk of accidents and improves overall site safety.

Fig. 7. Interface of part of modules. (a) safety belt recognition, (b) personnel gathering monitoring, (c) safety helmet recognition, (d) reflective clothing recognition.

3.2.3 DeepSeek localization deployment

To enhance the system's capabilities, we integrate DeepSeek, a large language model, for the following functions:

Safety Log Analysis: DeepSeek processes unstructured text data from safety logs and incident reports to identify patterns and provide actionable insights.

Real-Time Decision Support: DeepSeek provides real-time recommendations based on detected safety violations, historical data, and environmental conditions.

Knowledge Management: DeepSeek manages safety protocols, training materials, and regulatory compliance documents, ensuring that all personnel have access to up-to-date information.

To deploy DeepSeek locally within the HUST AI Box system, the following steps are recommended:

(1) Hardware Requirements: High-performance GPUs (e.g., NVIDIA A100) for model inference, sufficient storage capacity for large datasets and model weights, high-speed network infrastructure for real-time data processing.

(2) Software Requirements: Python 3.8 or higher with relevant libraries (e.g., TensorFlow, PyTorch), docker for containerized deployment of DeepSeek, Kubernetes for orchestration and scaling of DeepSeek instances.

(3) Deployment Steps:

Model Download: Download the pre-trained DeepSeek model weights and configuration files.

Containerization: Package DeepSeek into a Docker container for easy deployment.

Orchestration: Use Kubernetes to manage and scale DeepSeek instances across multiple nodes.

Integration: Integrate DeepSeek with the existing HUST AI Box system via RESTful APIs.

Testing: Conduct extensive testing to ensure seamless integration and optimal performance.

(4) Maintenance and Updates: Regularly update DeepSeek with the latest model weights and configurations, monitor system performance and optimize resource allocation as needed.

By following these steps, DeepSeek can be effectively deployed and integrated into the HUST AI Box system, providing advanced NLP capabilities for enhanced safety management on construction sites.

4. Conclusion

The development and implementation of HUST AI Box safety management system for construction sites have demonstrated a significant advancement in the field of construction safety, enhanced by the integration of DeepSeek, a large language model (LLM), to provide advanced natural language processing (NLP) capabilities for real-time decision support, knowledge management, and safety optimization. This study has successfully addressed the research question by introducing an AI and computer vision-based system capable of monitoring a range of safety-critical behaviours and hazards in dynamic and uncontrolled environments without requiring workers to perform specific actions or use additional equipment.

Innovation in Algorithm Integration: HUST AI Box represents a novel approach by combining YOLOv3 for pedestrian detection, MTCNN for face detection and alignment, and ArcFace for face recognition. This integration not only enhances the robustness of the system in varying environmental conditions but also addresses the challenge of face recognition in dynamic construction settings. Unlike traditional systems that require workers to perform specific actions or use additional equipment, our system operates seamlessly in real-time, without imposing any such constraints.

Advancement in Safety Monitoring: The system's ability to monitor multiple safety parameters simultaneously, including helmet, reflective clothing, safety belt usage, open flame detection, and personnel tracking, sets it apart from single-task safety monitoring systems. This comprehensive approach to safety management provides a unified framework that can dynamically adapt to the changing conditions of a construction site, thereby significantly reducing the risk of accidents and improving overall safety standards.

Innovative Assessment Management: The assessment management module of HUST AI Box is a significant innovation. It provides a quantitative measure of safety performance, allowing for the identification of trends and areas of concern. This data-driven approach to safety management is a departure from traditional qualitative assessments and provides a more objective and comprehensive view of safety on the construction site.

Real-time Monitoring and Response: The real-time monitoring capabilities of HUST AI Box allow for immediate action to be taken in response to safety violations. This is a significant advancement over manual oversight methods, which are often delayed and can result in missed opportunities to prevent accidents. The system's automated alerts ensure that managers are promptly notified of potential safety issues, enabling them to intervene before an incident occurs.

Acknowledgements

The authors want to acknowledge the financial support of the National Natural Science Foundation of China (U21A20151).

References

[1] W. Liu, A. Li, and C. Liu, “Multi-objective optimization control for tunnel boring machine performance improvement under uncertainty,” Automation Construction, vol. 139, p. 104310, Jul. 2022, doi: 10.1016/j.autcon.2022.104310.

[2] S.Y. Guo, L.Y. Ding, H.B. Luo and X.Y. Jiang, “A big-data-based platform of workers’ behaviour: observations from the field,” Accident Analysis & Prevention, vol. 93, pp. 299–309, Aug. 2016, doi: 10.1016/j.aap.2015.09.024.

[3] S.Y. Guo, L.Y. Ding, Y. Zhang, M.J. Skibniewski and K. Liang, “Hybrid recommendation approach for behaviour modification in the Chinese construction industry,” Journal of Construction Engineering and Management, vol. 145, no. 6, p. 04019035, Apr. 2019, doi: 10.1061/(ASCE)CO.1943-7862.0001665.

[4] J. Tompson, R. Goroshin, A. Jain, Y. LeCun and C. Bregler, “Efficient object localization using Convolutional Networks,” 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2015, pp. 648-656, doi: 10.1109/CVPR.2015.7298664.

[5] M. Vignoli, K. Nielsen, D. Guglielmi, M.G. Mariani, L. Patras and J.M. Peir`o, “Design of a safety training package for migrant workers in the construction industry,” Safety Science, vol. 136, p. 105124, Apr. 2021, doi: 10.1016/j.ssci.2020.105124.

[6] Y.-Y. Chen, M.-L. Tsai and J.-K. Jan, “The design of RFID access control protocol using the strategy of indefinite-index and challenge-response,” Computer Communications, vol. 34, no. 3, pp. 250-256, Mar. 2011, doi: 10.1109/ICGEC.2011.10.

[7] M.R. M. Isa, Y. Yahaya, M. Hazali Mohamed Halip, M.A. Khairuddin and K. Maskat, “The design of fingerprint biometric authentication on smart card for PULAPOT main entrance system,” 2010 International Symposium on Information Technology, 2010, pp. 1-4, doi: 10.1109/ITSIM.2010.5561969.

[8] Q. Fang, H. Li, X. Luo, L. Ding, T.M. Rose, W. An and Y. Yu, “A deep learning-based method for detecting non-certified work on construction sites,” Advanced Engineering Informatics, vol. 35, pp. 56-68, Jan. 2018, doi: 10.1016/j.aei.2018.01.001.