Intelligent Vision-Based Framework for Automated Detection of Safety Helmet Usage in Dynamic Workspaces

Nagma Arshin[1], *Omer*[2], Mohammed Fateen[3], *Mohammed Asad Uddin*[4], Mohammed Ahsan Aali[5]
[1]Assistant Professor, Department of CSE (Data Science),
Lords Institute of Engineering and Technology, Hyderabad, Telangana, India.
[2,3,4,5]UG Students, Department of CSE (Data Science),
Lords Institute of Engineering and Technology, Hyderabad, Telangana, India.

***Abstract*— This paper presents an automated system for detecting safety helmet usage in industrial environments using deep learning–based object detection techniques. Ensuring that workers wear helmets is essential for preventing head injuries, yet maintaining compliance remains a significant challenge. To address this, the proposed system utilizes an enhanced model based on the YOLO architecture, referred to as YOLO-ESCA, which is designed to improve detection accuracy under complex conditions such as varying lighting, occlusions, and diverse worker poses. The system provides a complete pipeline including dataset upload, preprocessing, model training, and evaluation. Performance is measured using metrics such as precision, recall, and mean Average Precision (mAP), along with graphical analysis of training progress. In addition, an extended YOLOv11 model is implemented to further enhance detection capability. Comparative results show that the extended model achieves better accuracy and robustness than the baseline YOLOv5 approach. The system also supports real-time helmet detection from both images and videos, making it suitable for practical deployment in surveillance systems. By automating helmet compliance monitoring, the proposed solution reduces the need for manual supervision and improves workplace safety. Overall, this work demonstrates the effectiveness of deep learning in developing scalable and efficient safety monitoring systems for hazardous environments.**

***Keywords*— Safety Helmet Detection, YOLO-ESCA, YOLOv5, YOLOv11, Object Detection, Deep Learning, Computer Vision, Industrial Safety, PPE Compliance, Real-Time Monitoring, Image Processing, Video Surveillance, mAP Evaluation, Workplace Safety**.

I. INTRODUCTION

In recent years, worker safety—particularly in high-risk environments like construction sites, manufacturing units, and mining areas—has gained significant attention. Among the essential protective measures, the use of personal protective equipment (PPE), especially safety helmets, plays a vital role in preventing head injuries. Despite established safety regulations [12], consistent helmet usage remains an issue, often resulting in avoidable accidents and fatalities. With the rapid advancement of technologies such as deep learning and computer vision, new opportunities have emerged to enhance safety practices in industrial settings. One effective approach is the use of automated monitoring systems that can identify whether workers are wearing safety helmets correctly. These systems reduce dependency on manual supervision while enabling continuous, real-time compliance tracking. However, developing a reliable detection model is challenging due to factors like different helmet styles, varied worker postures, changing lighting conditions, and visual obstructions.

To overcome these limitations, this work introduces YOLO-ESCA, an enhanced model designed for detecting proper helmet usage. It is built upon YOLOv5, a well-known object detection framework recognized for its balance between speed and accuracy [4]. Although YOLOv5 performs effectively in general scenarios, its performance can be further improved for complex industrial environments. The proposed YOLO-ESCA model extends the capabilities of YOLOv5 by incorporating targeted improvements aimed at handling real-world challenges. These include enhanced feature extraction mechanisms, better adaptability to occluded objects, and improved performance under varying illumination [3]. The architecture is refined to recognize different worker positions and helmet appearances more effectively, while maintaining real-time processing capabilities required for practical deployment. The main goal of YOLO-ESCA is to deliver a dependable and efficient system for monitoring helmet compliance in real time. By applying advanced deep learning techniques, the model supports automated and scalable safety supervision, helping reduce the likelihood of head injuries in workplaces [9].

In addition, the model is designed to be flexible and applicable across different industrial environments without requiring extensive adjustments. Its ability to function reliably under diverse conditions makes it suitable for integration into surveillance systems that require continuous monitoring. This paper presents the design and structure of the YOLO-ESCA model, highlights its improvements and evaluates its performance through experimental analysis [5]. The results demonstrate that YOLO-ESCA achieves higher accuracy, faster detection, and greater robustness compared to conventional approaches, making it a valuable contribution to industrial safety solutions. Overall, YOLO-ESCA represents an advancement in the use of artificial intelligence for workplace safety [9]. By automating helmet detection, it helps organizations improve compliance with safety standards, minimize human oversight errors, and create safer working conditions.

The main contribution of this work focuses on enhancing the performance of the standard YOLOv5 model for safety helmet detection. The improved model, referred to as YOLO-ESCA, is designed to better identify helmets in challenging conditions such as partial occlusion, varied lighting, and different worker orientations. The enhancement mainly targets stronger feature extraction so that even small or unclear objects can be detected more accurately. In addition, training strategies such as data augmentation and parameter tuning are applied to improve the model's ability to generalize across different environments [5]. The study also includes a comparison with an extended YOLOv11 model to examine performance improvements. These modifications aim to provide a more reliable and efficient solution for real-time helmet detection in industrial scenarios.

II. LITERATURE SURVEY

Redmon et al., (2016) [1] introduced the YOLO framework, which revolutionized object detection by treating it as a single regression problem instead of a multi-stage process. The model processes the entire image in one pass, making it significantly faster than traditional methods. It divides the image into a grid and predicts bounding boxes along with class probabilities simultaneously. This unified approach reduces computational complexity and improves real-time performance. The authors demonstrated that YOLO achieves high speed while maintaining competitive accuracy. However, early versions showed limitations in detecting small objects and handling complex scenes. Despite these challenges, the model laid the foundation for subsequent improvements in the YOLO family. The study highlighted the importance of balancing speed and accuracy in detection systems. It also opened new possibilities for real-time applications such as surveillance and safety monitoring. Overall, this work played a key role in advancing modern object detection techniques.

Redmon et al., (2017) [2] proposed YOLO9000, an improved version of the original YOLO model with enhanced accuracy and detection capabilities. The model introduced joint training on detection and classification datasets, allowing it to recognize a larger number of object categories. It utilized a more efficient architecture and improved feature representation to enhance performance. The authors also incorporated hierarchical classification, which enabled the model to generalize better across different classes. Compared to its predecessor, YOLO9000 achieved better localization accuracy and faster detection speed. The study addressed some limitations of the original YOLO, particularly in terms of generalization and scalability. It also demonstrated the effectiveness of combining multiple datasets for training. The improved model was capable of detecting over 9000 object categories. This work contributed significantly to the development of scalable object detection systems. It further strengthened the applicability of YOLO in real-time environments.

Bochkovskiy et al., (2020) [3] presented YOLOv4, which focused on optimizing both speed and accuracy for object detection tasks. The model introduced several improvements, including advanced data augmentation techniques and enhanced feature extraction methods. It incorporated the concept of "bag of freebies" and "bag of specials" to improve training efficiency without increasing inference cost. The authors utilized CSPDarknet as the backbone network to enhance feature representation. YOLOv4 also improved detection performance for small objects and complex scenes. The study demonstrated that the model achieves state-of-the-art results on benchmark datasets. It provided a practical solution that can be trained and deployed on conventional hardware. The model showed strong performance in real-time detection scenarios. This work made object detection more accessible and efficient for real-world applications. Overall, YOLOv4 significantly improved the robustness and reliability of detection systems.

Fang et al., (2018) [9] developed a real-time safety helmet detection system using deep learning techniques. The study focused on improving workplace safety by automatically identifying whether workers were wearing helmets. The authors used a convolutional neural network to detect head regions and classify helmet usage. The system was designed to handle real-world challenges such as varying lighting conditions and complex backgrounds. Experimental results showed that the model achieved high detection accuracy in practical scenarios. The study also demonstrated the feasibility of deploying such systems in surveillance environments. It highlighted the importance of automated monitoring in reducing workplace accidents. The model performed well in detecting multiple individuals simultaneously. However, the study noted limitations in handling heavy occlusions. Despite this, it provided a strong foundation for further research in PPE detection. Overall, the work contributed to the development of intelligent safety systems.

Nath et al., (2020) [12] explored the use of deep learning for detecting personal protective equipment in construction environments. The study focused on real-time identification of safety gear such as helmets using image-based analysis. The authors applied advanced object detection models to improve detection accuracy and reliability. The system was capable of processing live video streams, making it suitable for continuous monitoring. The study emphasized the importance of safety compliance in reducing workplace risks. It also highlighted the challenges associated with dynamic environments and worker movement. Experimental results showed that deep learning models can effectively detect PPE with high precision. The authors discussed the scalability of the system for large construction sites. The study also suggested integrating such systems with existing surveillance infrastructure. Overall, this work demonstrated the practical applicability of deep learning in enhancing industrial safety and monitoring systems.

III. DATASET DESCRIPTION

The dataset used in this study is based on the Safety Helmet Wearing Dataset (SHWD), which contains images of individuals in industrial environments with and without safety helmets. The dataset includes a diverse collection of images captured under different real-world conditions such as varying lighting, complex backgrounds, and multiple viewing angles. This diversity helps improve the model’s ability to generalize across different scenarios. The dataset is organized into labelled categories representing helmet-wearing and non-helmet-wearing classes. Each image is annotated with bounding boxes that indicate the location of the head region and the presence or absence of a helmet. These annotations are essential for training object detection models, as they provide spatial information required for accurate localization and classification. Before training, the dataset undergoes a preprocessing phase that includes data cleaning, normalization, and formatting to match the input requirements of the YOLO-based models. The dataset is then divided into training and testing subsets to evaluate model performance effectively. The training set is used to learn patterns, while the testing set is used to validate the model’s accuracy and generalization capability. To further improve model robustness, the dataset supports variations such as occlusions, different helmet colors, and worker poses. This ensures that the trained model can perform reliably in real-time environments. The dataset plays a crucial role in enabling the system to detect helmet usage accurately in both images and video streams.

IV. IMPLEMENTATION DETAILS

The proposed system is implemented as an integrated application that enables dataset handling, model training, performance evaluation, and real-time detection through a user-friendly interface. The execution begins by running the main application file, which launches the system interface for interaction. The first step involves uploading the Safety Helmet Wearing Dataset (SHWD), where the dataset folder containing annotated images is loaded into the system. Once the dataset is loaded, a preprocessing stage is performed to prepare the data for model training.

This stage includes organizing image-label pairs, resizing images to match model input dimensions, and converting annotations into the required format for YOLO-based training. The dataset is then split into training and validation sets to ensure proper performance evaluation. The system implements the YOLO-ESCA model, which is an enhanced version of the YOLOv5 architecture. During training, the model learns to detect helmet and non-helmet classes using annotated bounding boxes. Key training parameters such as epochs, learning rate, and batch size are configured to optimize performance.

Throughout the training process, important metrics such as precision, recall, and loss are continuously monitored. These metrics are visualized using graphs to analyze model learning behavior, where increasing precision and recall along with decreasing loss indicate improved performance. In addition to YOLO-ESCA, an extended YOLOv11 model is also implemented to further improve detection accuracy. The training process for YOLOv11 follows a similar procedure, with enhanced feature extraction and optimization techniques. The system generates training and validation graphs, including precision, recall, mAP50, and mAP50–95 curves, to provide a comprehensive view of model performance. A comparison module is included to evaluate the performance of YOLOv5 and the extended YOLOv11 model. The comparison is based on accuracy metrics, where the improved model demonstrates better detection capability. Finally, the system supports real-time detection from both images and video inputs. Users can upload images or video files, and the trained model processes them to detect helmet usage. The output displays bounding boxes around detected objects along with classification labels, including cases with uncertainty. This end-to-end implementation ensures an efficient and practical solution for automated helmet detection in real-world environments. In this implementation, YOLO-ESCA is treated as an enhanced version of the YOLOv5 model, and its performance is analysed during training, while final comparison is carried out between YOLOv5 and the extended YOLOv11 model

Figure 1: Workflow of Proposed Safety Helmet Detection System

V. EXPERIMENTAL SETUP

The evaluation of the proposed system was carried out using a standard computing setup suitable for deep learning applications. The training process was performed using a GPU-supported environment to ensure faster computation and efficient model learning. The Safety Helmet Wearing Dataset was divided into training and validation sets to assess the model's performance on unseen data.

During training, several parameters were selected carefully to achieve stable and effective learning. The models were trained over multiple epochs to allow gradual improvement in detection capability. A balanced batch size was used to maintain computational efficiency while ensuring proper convergence. The learning rate was adjusted to avoid instability during training. To enhance robustness, various data augmentation techniques such as image scaling, flipping, and brightness adjustments were applied. The effectiveness of the models was measured using commonly accepted evaluation metrics, including precision, recall, mAP50, and mAP50–95. These metrics help in understanding both the classification accuracy and localization performance of the detection models.

VI. RESULT AND DISCUSSION

The performance of the proposed system was evaluated using both the YOLO-ESCA model and the extended YOLOv11 model on the Safety Helmet Wearing Dataset. During training, key evaluation metrics such as precision, recall, and loss were monitored to assess the effectiveness of the models. The training graphs showed a steady increase in precision and recall values with the progression of epochs, while the loss values decreased consistently. This indicates that the models were able to learn meaningful features from the dataset and improve their prediction capability over time.

The YOLO-ESCA model demonstrated good detection performance in identifying both helmet and non-helmet classes. It was able to detect objects accurately under different conditions, including variations in lighting, background complexity, and worker poses. However, in some cases involving occlusions or unclear visibility, the model showed slight uncertainty in detection, which is expected in real-world scenarios. The extended YOLOv11 model further improved the detection performance.

The training and validation graphs for YOLOv11 showed more stable learning behaviour, with better convergence and higher metric values compared to YOLO-ESCA. The mAP50 and mAP50–95 scores indicated improved accuracy in object localization and classification. This suggests that the enhancements in the YOLOv11 model contributed to better feature extraction and generalization. A comparative analysis between YOLOv5 (base model) and the extended YOLOv11 model clearly showed that YOLOv11 achieved higher accuracy. The improved model was more effective in detecting helmets with fewer errors and better consistency across different test samples.

The system was also tested on real-time inputs, including images and video streams. The detection results confirmed that the model can successfully identify helmet usage and highlight detected regions with bounding boxes. The system performed efficiently in real-time scenarios, making it suitable for practical deployment in surveillance applications. Overall, the results demonstrate that the proposed approach provides a reliable and effective solution for safety helmet detection. The improvements introduced in the extended model significantly enhance detection accuracy and robustness, contributing to better monitoring of safety compliance in industrial environments. The comparison results indicate that the extended YOLOv11 model achieves better overall accuracy compared to the baseline YOLOv5 model, as illustrated in Figure [2]. The improvement is reflected in better localization performance and reduced detection errors across different test samples.

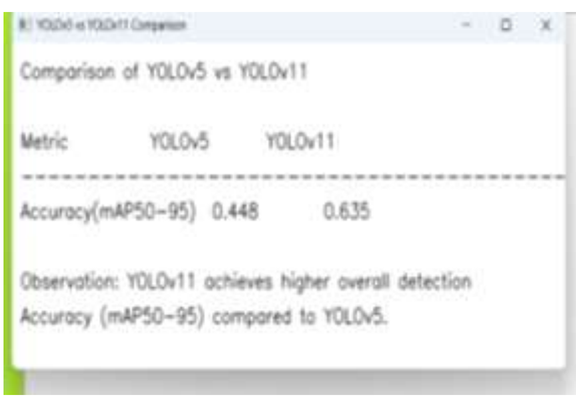

Figure 2: comparison results of yolov5 (yolo-esca) and yolov11

VII. CONCLUSION

This paper presented an automated system for detecting safety helmet usage using deep learning-based object detection models. The proposed approach focused on improving workplace safety by ensuring compliance with helmet-wearing regulations in industrial environments. The YOLO-ESCA model was developed as an enhanced version of the existing YOLO framework to handle challenges such as varying lighting conditions, occlusions, and diverse worker poses. The experimental results demonstrated that the proposed model achieved reliable performance in terms of precision, recall, and overall detection accuracy. The training analysis showed consistent improvement in learning, with reduced loss and increased accuracy across epochs. Furthermore, the implementation of the extended YOLOv11 model provided additional improvements, achieving higher accuracy and better generalization compared to the baseline YOLOv5 model. The system was also validated on real-time image and video inputs, where it successfully detected helmet usage and identified cases of non-compliance. This confirms the practical applicability of the system in real-world monitoring scenarios. The comparison results highlighted that the improved model can significantly enhance detection performance and reduce errors.

Overall, the proposed solution offers an efficient, scalable, and automated method for monitoring safety compliance. It reduces the need for manual supervision and supports continuous surveillance in hazardous environments. The work demonstrates the potential of deep learning techniques in developing intelligent safety systems and contributes to improving worker protection and reducing workplace accidents.

REFERENCES

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788.

[2] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 7263–7271.

[3] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *arXiv preprint arXiv:2004.10934*, 2020.

[4] G. Jocher et al., "YOLOv5," *Ultralytics GitHub Repository*, 2020. [Online].

[5] C. Y. Wang, A. Bochkovskiy, and H. Y. M. Liao, "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors," *arXiv preprint arXiv:2207.02696*, 2022.

[6] W. Liu, D. Anguelov, D. Erhan, et al., "SSD: Single Shot MultiBox Detector," *European Conference on Computer Vision (ECCV)*, 2016, pp. 21–37.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[8] T. Y. Lin, P. Dollár, R. Girshick, et al., "Feature Pyramid Networks for Object Detection," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2117–2125.

[9] W. Fang, L. Ding, B. Zhong, et al., "Real-Time Safety Helmet Detection Based on Deep Learning," *IEEE Access*, vol. 6, pp. 20162–20172, 2018.

[10] J. Wu, Z. Cai, X. Chen, and Y. Yu, "A Real-Time Helmet Detection System Based on Improved YOLO," *Proceedings of the IEEE International Conference Workshops*, 2019.

[11] H. Li, Z. Zhao, and L. Wang, "Safety Helmet Detection Based on Improved YOLOv3," *Journal of Physics: Conference Series*, vol. 1624, 2020.

[12] N. D. Nath, A. H. Behzadan, and S. G. Paal, "Deep Learning for Site Safety: Real-Time Detection of Personal Protective Equipment," *Automation in Construction*, vol. 112, 2020.

[13] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015.

[14] M. Everingham, L. Van Gool, C. K. I. Williams, et al., "The Pascal Visual Object Classes (VOC) Challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010.

[15] T. Y. Lin, M. Maire, S. Belongie, et al., "Microsoft COCO: Common Objects in Context," *European Conference on Computer Vision (ECCV)*, 2014, pp. 740–755.

Babburi, S. Privacy-Preserving Collaborative Framework with Auditable Federated Learning.

Gaddam, S. Integrating Analytics into the Development Process: Bridging the Gap between Data Insights and Design Execution.

Immadi, S. K. (2025). Optimizing ERP for Human Capital Management. Applied Research for Growth, Innovation and Sustainable Impact, 377–384. https://doi.org/10.1201/9781003684657-63

Reddy, S. K. R. Developing a Modular AI Framework to Enhance Scalability and Personalization in Next-Generation Reward Platforms.

Poojari, R. INTELLIGENT SYSTEMS+B108 AND APPLICATIONS IN ENGINEERING.

Mahimalur, R. K., Vasgam, M., & Manoharan, D. Devops Lifecycle Management And Cloud Migration Assessments: A Security-Driven CICD Perspective.

Viswanathan, V. (2023). AI-Augmented Decision Intelligence for Enterprise Systems: Integrating Cognitive Analytics for Resource and Talent Optimization.

Agrawal, A. M., Gajula, S., Shinde, R. P., Shah, H., & Ghosh, H. (2025, July). Machine Translation for Long Sequences with Enhanced Attention Mechanisms. In 2025 5th International Conference on Electrical, Computer and Energy Technologies (ICECET) (pp. 1-6). IEEE.

Maturi, S. Y. (2021). Blockbond hardening: Securing pooled-hash protocols against traffic tampering, MITM hash-rate hijacking, and template coercion. International Journal of Communication Networks and Information Security, 13(3), 718–728.

Adabala, P. K. (2024). Utilizing predictive analytics to improve efficiency and decision-making in ERP-connected supply chains. International Journal of Intelligent Systems and Applications in Engineering, 12(22s), 2465

Kavuri, S. (2026). An Explainable Machine Learning Framework for Predicting Software Defects in Large-Scale Software Systems. 2026 IEEE 5th International Conference on AI in Cybersecurity (ICAIC), 1–6. https://doi.org/10.1109/icaic67076.2026.11395777

Gajula, S., & Kandula, S. T. R. (2026). Securing Financial Data in Multi-Tenant Clouds Through AI, Blockchain, and Attribute-Based Encryption. Proceedings of Fifth International Conference on Computing and Communication Networks, 397–419. https://doi.org/10.1007/978-3-032-21499-7_33

Shashank, A. (2025). Self-Healing Data Pipelines for Enhanced Reliability: A Paradigm Shift in Enterprise Data Management. Journal of Computer Science and Technology Studies, 7(8), 1097-1104.

Susarla, R. S., Boyapati, P. K., & Kandula, S. T. R. (2025, July). Cloud-Based Secure Data Storage in Smart Cities Using Central-Smoothing Hypergraph Neural Networks. In 2025 IEEE 4th World Conference on Applied Intelligence and Computing (AIC) (pp. 279-284). IEEE.

Boyapati, P. K. Building a centralized data operations hub for healthcare enterprise integration. IJSAT-Int. J. Sci. Technol. 16 (2). https://doi.org/10.71097/IJSAT.v16.i2.3219