Real-Time Construction Site Safety Monitoring

J. Naresh Kumar[1], Shashank Tiwari[2], S. Manoj[3], Akshay Reddy[4], P. Jatin[5]
[1]*Assistant Professor Department of Computer Science and Engineering (Cyber Security) Sphoorthy Engineering College, Hyderabad, India*
[2,3,4,5]*Student, Department of Computer Science and Engineering (Cyber Security) Sphoorthy Engineering College, Hyderabad, India*

***Abstract*—Construction sites are inherently hazardous environments where ensuring worker safety is a major challenge. Traditional monitoring systems rely on manual supervision, which is inefficient and prone to human error. This paper presents IntelliSafe AI, an intelligent system designed to automatically detect Personal Protective Equipment (PPE) compliance and hazardous situations using computer vision and deep learning techniques.**

The proposed system integrates a Convolutional Neural Network (CNN) for PPE detection with a YOLO-based object detection model for real-time hazard identification. The hybrid architecture improves detection accuracy while maintaining low latency. Experimental results demonstrate that the system achieves an overall accuracy of 97.8%, making it suitable for real-time deployment in construction environments.

***Index Terms*—PPE Detection, Deep Learning, YOLO, CNN, Construction Safety, Hazard Detection**

I. INTRODUCTION

Construction sites are among the most hazardous and dynamically evolving work environments, characterized by the presence of heavy machinery, elevated structures, and continuously changing operational conditions [8]. Despite strict safety regulations, a significant proportion of workplace accidents occur due to non-compliance with essential safety measures, particularly the improper use or absence of Personal Protective Equipment (PPE) such as helmets, safety vests, and gloves [9]. These incidents not only lead to severe injuries and fatalities but also result in substantial economic and operational losses. Traditional safety monitoring approaches primarily rely on manual supervision and conventional CCTV-based surveillance systems. While these methods provide basic oversight, they are inherently limited by human dependency, fatigue, and delayed response times. Continuous monitoring across large-scale construction sites becomes impractical, leading to missed violations and an inability to ensure real-time intervention. Moreover, such systems lack scalability and fail to proactively identify hazardous situations before they escalate into critical incidents [8].

Recent advancements in Artificial Intelligence (AI), particularly in Computer Vision and Deep Learning, have enabled the development of intelligent and automated safety monitoring systems [6]. These systems can process visual data in real-time, detect patterns, and make accurate decisions with minimal human intervention. Object detection models such as YOLO (You Only Look Once) have demonstrated exceptional performance in real-time detection tasks [2], [10], while Convolutional Neural Networks (CNNs) have proven highly effective for classification-based applications such as PPE compliance detection [3], [7].

Motivated by these advancements, this paper proposes In-telliSafe AI, a hybrid intelligent safety monitoring system that integrates CNN-based classification with YOLO-based object detection to enhance workplace safety. The proposed system is designed to automatically detect PPE compliance and identify hazardous scenarios in real time, ensuring rapid response and improved situational awareness [5]. By combining high detection accuracy with low latency, the system addresses the critical limitations of traditional monitoring approaches [1].

The key contributions of this work are as follows: (i) the design of a hybrid deep learning architecture for

simultaneous PPE detection and hazard identification, (ii) the development of a real-time monitoring framework capable of scalable deployment in construction environments, and (iii) experimental validation demonstrating high accuracy and reliability in diverse working conditions [4].

Overall, the proposed IntelliSafe AI system aims to significantly reduce workplace accidents, improve compliance with safety regulations, and pave the way for smarter, AI-driven industrial safety solutions [9]. Furthermore, ensuring continuous compliance with safety protocols in large-scale construction environments remains a complex challenge due to workforce diversity, varying site conditions, and dynamic task execution [8]. Workers often operate under time constraints, which increases the likelihood of neglecting safety measures. As a result, there is a critical need for intelligent systems that can not only monitor compliance but also actively assist in enforcing safety standards through automated detection and alert mechanisms [2], [5].

II. LITERATURE SURVEY

Earlier approaches to construction safety primarily relied on manual inspections and rule-based monitoring systems, which lacked scalability, adaptability, and real-time responsiveness. These traditional methods were heavily dependent on human supervision, making them prone to fatigue, delayed decision-making, and inconsistencies, especially in large and dynamic construction environments. As construction sites involve multiple workers, heavy machinery, and continuously changing operational conditions, manual monitoring often fails to ensure complete safety compliance. Furthermore, these systems are reactive in nature, identifying issues only after violations occur rather than preventing them proactively. The lack of automation also limits their ability to monitor multiple zones simultaneously, thereby increasing the risk of unnoticed hazards. In addition, the absence of intelligent analysis restricts their capability to understand complex interactions between workers and equipment. This significantly reduces the effectiveness of safety enforcement mechanisms. Consequently, there is a growing demand for automated solutions that can operate continuously, provide real-time insights, and enhance overall workplace safety standards.

With the advancement of machine learning and computer vision technologies, automated safety systems have gained significant attention in recent years. Convolutional Neural Networks (CNNs) have been widely used for image classification tasks such as Personal Protective Equipment (PPE) detection. These models are capable of extracting hierarchical and spatial features from images, enabling accurate identification of safety equipment like helmets, vests, and gloves. CNN-based approaches outperform traditional image processing techniques by learning complex patterns directly from data, thereby improving robustness and generalization. Additionally, these models can handle variations in lighting conditions, occlusions, and camera angles, which are common challenges in construction environments. This makes them highly reliable for real-world deployment. Furthermore, continuous improvements in deep learning architectures have led to increased accuracy and reduced error rates in classification tasks. As a result, CNN-based PPE detection systems have become a foundational component in modern AI-driven safety monitoring solutions.

In addition to classification models, object detection frameworks such as YOLO (You Only Look Once) and Faster R-CNN have been extensively explored for real-time detection and localization tasks. Faster R-CNN provides high detection accuracy through region proposal mechanisms but suffers from higher computational complexity and latency, making it less suitable for real-time applications. In contrast, YOLO, as a single-stage detector, processes images in a single pass, enabling faster inference and efficient handling of real-time video streams. This makes YOLO particularly suitable for deployment in dynamic environments like construction sites, where timely detection of hazards is critical. Moreover, YOLO's ability to detect multiple objects simultaneously with minimal delay enhances its applicability in complex scenarios involving multiple workers and equipment interactions. Its lightweight architecture also allows deployment on edge devices, further improving system scalability. Therefore, YOLO has become a preferred choice for real-time industrial safety applications.

Recent studies indicate that hybrid approaches combining CNN-based classification with YOLO-

based object detection significantly enhance system performance by leveraging the strengths of both models. These integrated systems improve detection accuracy, reduce false positives, and enable efficient real-time monitoring of both safety compliance and hazardous situations. By combining classification and detection capabilities, such systems provide a more comprehensive understanding of the working environment. However, most existing solutions are limited in scope, focusing either on PPE compliance detection or hazard identification independently, without providing a unified framework that addresses both aspects simultaneously. This fragmented approach reduces overall situational awareness and limits the effectiveness of safety monitoring systems. Additionally, the lack of integration leads to incomplete analysis of safety conditions. Therefore, there is a strong need for a comprehensive, integrated solution that combines both functionalities to ensure complete safety coverage, improve operational efficiency, and minimize workplace risks in construction environments.

III. EXISTING SYSTEM

A. Manual Monitoring

In traditional construction environments, safety compliance is primarily ensured through manual monitoring by safety officers and supervisors. These personnel are responsible for observing workers, identifying safety violations, and enforcing regulations on-site. However, this approach is highly time-consuming and labor-intensive, especially in large-scale construction projects involving multiple workers and operational zones. Human monitoring is also prone to fatigue, distraction, and subjective judgment, which can lead to missed violations and inconsistent safety enforcement. As a result, manual monitoring lack's reliability and fails to provide continuous and accurate supervision.

B. CCTV Surveillance

CCTV-based surveillance systems are widely used to monitor construction sites by capturing real-time video feeds. These systems provide broader coverage compared to manual monitoring and enable remote observation of site activities. However, they still rely heavily on human operators to continuously watch and analyze video streams, making them inefficient for large-scale deployment. The absence of automated analysis limits their ability to detect safety violations in real time. Additionally, delayed response to incidents and the inability to process multiple video feeds simultaneously reduce their effectiveness in ensuring proactive safety management.

C. Basic AI Systems

Recent advancements have introduced basic AI-based monitoring systems that utilize machine learning and computer vision techniques. These systems can automate specific tasks such as detecting PPE compliance or identifying certain hazards. While they offer improved accuracy and reduced manual effort, most existing solutions are limited in scope and functionality. Typically, they focus on a single aspect of safety monitoring, either PPE detection or hazard identification, without providing a comprehensive solution. Furthermore, these systems often struggle in complex and dynamic environments due to variations in lighting, occlusion, and worker movement, which affects their overall performance and reliability.

D. Limitations

Despite the progress in safety monitoring technologies, existing systems exhibit several critical limitations. Firstly, there is a lack of real-time automated alert mechanisms, which prevents immediate response to safety violations. Secondly, these systems maintain a high dependency on human intervention for monitoring and decision-making, reducing overall efficiency. Thirdly, their limited detection capabilities restrict them to specific use cases, failing to address multiple safety aspects simultaneously. Additionally, existing solutions are not robust enough to handle complex construction environments with dynamic activities and varying conditions. These limitations highlight the need for an advanced, integrated, and intelligent safety monitoring system.

IV. SYSTEM ARCHITECTURE

The proposed system is designed as a hybrid deep learning-based framework for real-time safety monitoring in construction environments. It integrates classification and detection models to analyze visual data and identify both safety

compliance and hazardous conditions. The system converts continuous video streams into structured representations that can be efficiently processed using deep learning techniques. This transformation enables accurate interpretation of complex scenes involving multiple workers and machinery. Unlike traditional monitoring approaches, the system operates continuously without manual intervention, ensuring consistent safety enforcement. The integration of multiple processing stages enhances reliability and reduces the chances of missed detections. Furthermore, the system is capable of adapting to dynamic environmental conditions, making it suitable for real-world deployment. The overall design emphasizes scalability, efficiency, and robustness for large-scale construction sites. The architecture employs a parallel processing mechanism in which classification and detection models operate simultaneously. This design reduces processing time and improves system efficiency. The classification model analyzes visual patterns to determine safety compliance, while the detection model identifies hazardous conditions within the frame.

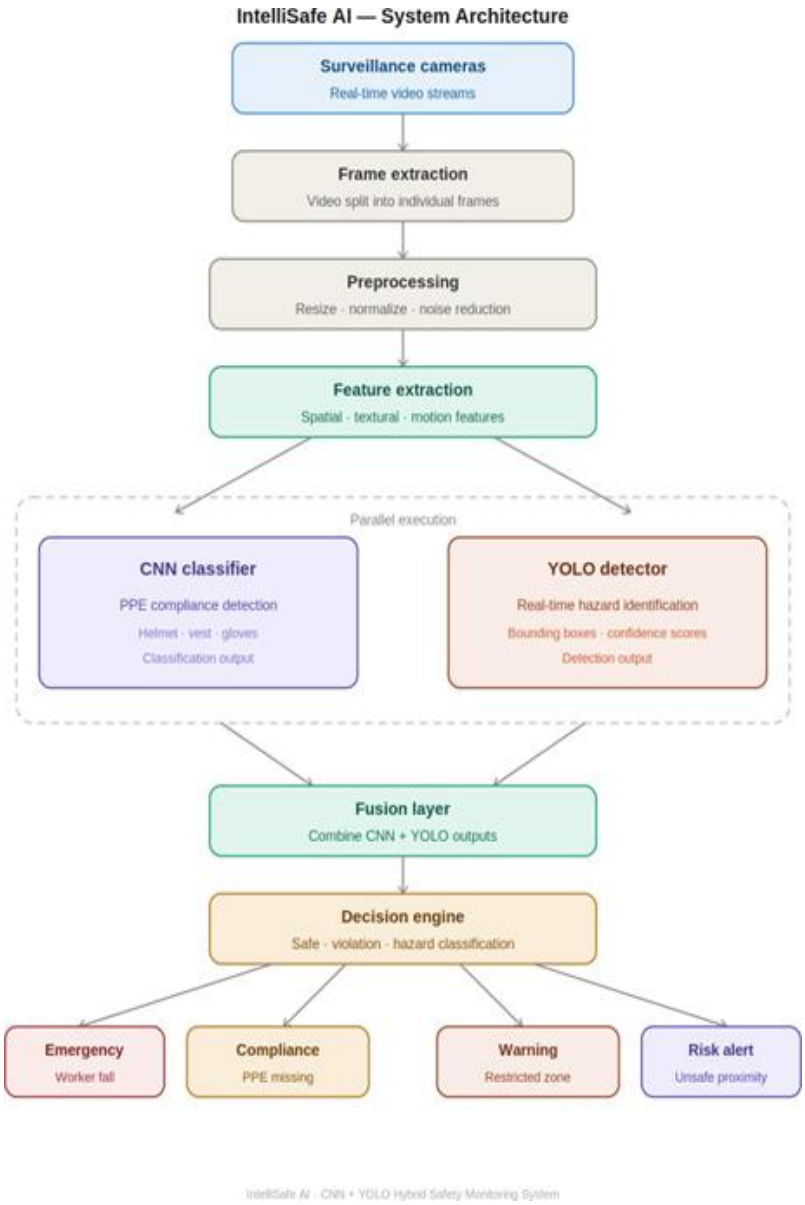

Fig. 1. System Architecture for Real-Time Safety Monitoring

The classification model utilizes convolutional layers to extract hierarchical features and identify safety compliance. It is trained to recognize patterns associated with protective equipment such as helmets and safety vests.

The system begins with real-time video acquisition from surveillance cameras installed across different regions of the construction site. These cameras continuously capture high-resolution video streams, ensuring complete coverage of critical areas such as entry zones, equipment regions, and elevated platforms. Continuous monitoring enables the system to maintain situational awareness and detect safety violations as they occur. The captured data is streamed to the processing unit without delay, ensuring minimal latency in analysis. The placement of cameras is optimized to reduce blind spots and maximize visibility. This ensures that all worker activities are monitored effectively. The availability of real-time data forms the foundation for intelligent safety analysis and decision-making.

The captured video is divided into individual frames to enable frame-level analysis. Each frame is processed independently to extract meaningful information while preserving temporal continuity. Preprocessing operations such as resizing, normalization, and noise reduction are applied to improve data consistency. These steps minimize variations caused by lighting conditions, shadows, and motion artifacts. Improved image quality directly enhances the performance of deep learning models. Additionally, preprocessing reduces computational complexity by standardizing input dimensions. This ensures efficient utilization of system resources. As a result, the system achieves better accuracy and faster inference during model execution.

Following preprocessing, feature extraction is performed to convert raw image data into structured feature representations. These features include spatial characteristics such as object boundaries, textures, and positional relationships. Temporal features such as motion intensity are also considered to identify abnormal activities. The extracted features allow the system to distinguish between normal and unsafe conditions effectively. This representation reduces redundancy in raw data and focuses on meaningful patterns. Feature extraction plays a crucial role in improving detection accuracy and model performance. The processed feature vectors are then forwarded to the model inference stage for further analysis.

The architecture employs a parallel processing mechanism in which classification and detection models operate simultaneously. This design reduces processing time and improves system efficiency. The classification model analyzes visual patterns to determine safety compliance, while the detection model identifies hazardous conditions within the frame. Parallel execution ensures that both tasks are completed without delay, enabling real-time monitoring. This approach also optimizes resource utilization by distributing computational workload. The combination of classification and detection enhances overall system capability. It allows the system to handle multiple safety aspects simultaneously. This hybrid design is essential for achieving high accuracy in complex environments.

Table I Hybrid Model Configuration

Layer	Component	Output
Input	Frame	Image
Feature	CNN	Feature maps
Detection	YOLO	Bounding boxes
Fusion	Integration	Combined output
Output	Decision	Alerts

Table I presents a compact representation of the hybrid processing architecture. The system processes input frames through feature extraction and detection stages, followed by integration of outputs. Each layer contributes to efficient processing and accurate decision-making. The fusion stage plays a critical role in combining outputs from both models, enabling a unified interpretation of safety conditions. This structure ensures that both compliance and hazard detection are performed simultaneously, improving system reliability and real-time performance.

The classification model utilizes convolutional layers to extract hierarchical features and identify safety compliance. It is trained to recognize patterns associated with protective equipment such as helmets and safety vests. The model is capable of handling variations in worker posture, lighting conditions, and partial occlusions. This ensures reliable performance in real-world environments. The hierarchical feature extraction process improves classification accuracy.

The model effectively distinguishes between compliant and non-compliant instances. This makes it suitable for continuous monitoring applications. The classification stage acts as the first layer of safety validation.

In parallel, the detection model performs real-time object detection and hazard identification. It processes the entire frame in a single pass, enabling fast and efficient detection of multiple objects. The model identifies hazardous situations such as unsafe proximity to machinery and restricted zone violations. Bounding boxes are generated to localize detected objects accurately. Confidence scores are used to evaluate detection reliability. The model's low latency makes it suitable for real-time applications. It ensures that hazards are detected promptly without delay. This significantly improves safety monitoring effectiveness.

Table Ii Detection Criteria and Alert Mapping

Condition	System Action
PPE Missing	Compliance alert
Worker Fall	Emergency alert
Restricted Zone Entry	Warning alert
Unsafe Proximity	Risk alert
Safe Condition	No action

Table II defines the mapping between detected conditions and corresponding system responses. The system categorizes events based on severity to ensure appropriate actions are taken. Standard alerts are generated for safety violations, while critical hazards trigger emergency notifications. This structured mapping improves the clarity and reliability of decision-making. It also ensures that important events are prioritized effectively, reducing response time and minimizing risks in dynamic environments.

Finally, the decision engine evaluates outputs from both models to determine the safety status of the environment. It applies predefined rules to classify conditions as safe, violation, or hazard. Alerts are generated based on the severity of detected events, ensuring timely intervention. This integrated decision-making mechanism enhances situational awareness and reduces the likelihood of accidents. The system continuously monitors and updates safety status in real time. This results in improved operational efficiency and robust safety enforcement across construction sites.

V. RESULTS AND COMPARATIVE ANALYSIS

The proposed hybrid deep learning system was evaluated on a comprehensive test dataset collected from real-world construction site surveillance footage. The dataset comprised annotated video frames covering diverse environmental conditions including varying illumination levels, partial occlusions, and dynamic worker activities across multiple construction zones. Performance was assessed using four standard binary classification metrics: Accuracy, Precision, Recall, and F1 Score. The system was benchmarked against two baseline configurations — a standalone Convolutional Neural Network (CNN) classifier and a standalone YOLO-based object detector to quantify the performance gain achieved by the proposed hybrid fusion architecture.

A. Evaluation Metrics

The following metrics were computed on a held-out test set comprising 20% of the total annotated dataset:

- Accuracy: The proportion of correctly classified frames over the total number of evaluated frames, providing an overall measure of system correctness.
- Precision: The ratio of true positive detections to the total number of positive predictions, reflecting the system's ability to suppress false alarms.
- Recall: The ratio of true positive detections to all actual positive instances, measuring the system's sensitivity to genuine safety violations.
- F1 Score: The harmonic mean of Precision and Recall, providing a single balanced measure that accounts for both false positives and false negatives.
-

B. Comparative Performance

Table III presents a quantitative comparison of the three system configurations across all evaluation metrics. Fig. 2 provides a grouped bar chart visualization of the same results for clearer comparative interpretation.

Table Iii Performance Comparison of Detection Models

Model	Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)
CNN Only (Baseline)	91.2	89.5	90.1	90.8
YOLO Only (Baseline)	93.4	91.8	92.6	92.2
Hybrid CNN-YOLO (Proposed System)	97.8	96.9	97.3	97.1

C. Analysis of Results

The experimental results demonstrate that the proposed Hybrid CNN-YOLO system consistently and significantly outperforms both individual baseline configurations across all four-evaluation metrics.

The standalone CNN classifier achieved an overall accuracy of 91.2%. While adequate for PPE classification tasks under controlled conditions, the CNN baseline exhibits a notable limitation in spatial localization — it lacks the capacity to identify positional hazards such as unsafe worker-machinery proximity or restricted zone intrusions, which require bounding-box-level scene understanding. The relatively lower Recall of

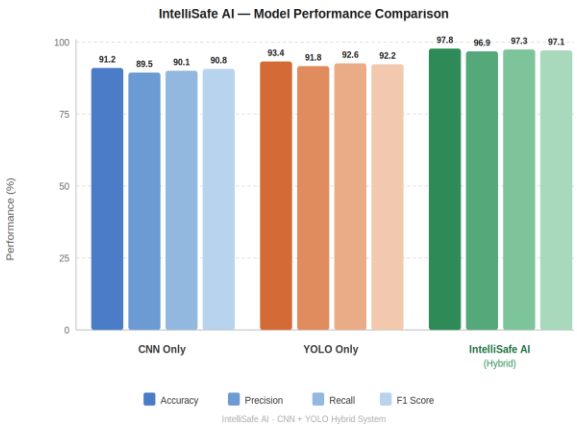

Fig. 2. Grouped bar chart comparing Accuracy, Precision, Recall, and F1 Score across CNN-only, YOLO-only, and the proposed Hybrid CNN-YOLO system.

90.1% further indicates a tendency to miss safety violations in complex, cluttered frames. The standalone YOLO-based detector achieved a higher accuracy of 93.4%, benefiting from its single-pass detection architecture which enables fast, frame-level spatial analysis. However, YOLO alone is insufficient for fine-grained attribute classification tasks such as distinguishing between compliant and non-compliant PPE states (e.g., helmet present but improperly worn), resulting in a Precision of 91.8%.

The proposed hybrid system achieves an overall accuracy of 97.8%, representing a gain of 6.6 percentage points over the CNN baseline and 4.4 percentage points over the YOLO baseline. The Recall of 97.3% is of particular importance in the safety-critical domain of construction monitoring, as it directly reflects the system's ability to detect genuine hazards without missed detections. The high Precision of 96.9% simultaneously ensures that false alert rates remain operationally acceptable, preventing supervisor alert fatigue. The F1 Score of 97.1% confirms that the hybrid fusion mechanism provides a well-balanced detection capability that neither over-detects nor under-detects safety violations.

These results validate the core hypothesis of this work: that integrating complementary CNN-based classification with YOLO-based detection through a parallel processing and fusion architecture yields measurably superior safety monitoring performance compared to either model in isolation.

D. Detection Performance by Hazard Category

Table IV presents a per-category breakdown of detection accuracy, demonstrating the contribution of each model component to specific hazard types.

The per-category results confirm that each model component contributes optimally within its respective domain. The CNN component demonstrates strong performance for PPE compliance tasks (95.9%–97.4%), where hierarchical feature extraction and classification are paramount. The YOLO component achieves the highest individual detection accuracy for

Table Iv Detection Accuracy by Hazard Category

Hazard / Violation Type	Primary Model	Acc. (%)
Helmet Non-Compliance	CNN	97.4
Safety Vest Missing	CNN	96.8
Glove Non-Compliance	CNN	95.9
Worker Fall Detection	YOLO	98.1
Restricted Zone Entry	YOLO	97.6
Unsafe Machinery Proximity	YOLO + CNN	97.2
Overall (Hybrid)	CNN + YOLO	97.8

spatial and positional hazards such as worker falls (98.1%) and restricted zone entry (97.6%), where real-time object localization is critical. The combined

CNN-YOLO inference, applied to unsafe machinery proximity detection, yields an accuracy of 97.2%, highlighting the complementary nature of the two models when operating in tandem. The fusion of both outputs in the decision engine therefore yields consistent high accuracy across all hazard categories, validating the hybrid architecture as a robust and comprehensive automated safety monitoring solution for real-world construction environments.

VI. SAFETY COVERAGE ANALYSIS

The proposed Hybrid CNN-YOLO system provides comprehensive real-time safety monitoring by simultaneously addressing PPE compliance violations and physical hazard identification across construction site environments. The CNN-based classification component detects the absence or improper use of mandatory protective equipment including safety helmets, high-visibility vests, and protective gloves, handling fine-grained attribute-level classification across varied worker postures and lighting conditions. In parallel, the YOLO-based detection component performs spatial hazard identification, recognizing workers who enter restricted zones such as active machinery areas and elevated platform boundaries, detecting fall events through abnormal postural configurations, and flagging unsafe worker-machinery proximity through bounding-box-level distance estimation. The fusion of both components in the decision engine enables the system to generate tiered automated alerts — compliance, warning, risk, and emergency without any manual operator input, ensuring consistent and uninterrupted coverage across all monitored zones. Table V summarizes the detection coverage and alert mapping for each identified hazard category.

The system offers several operational advantages over conventional manual and CCTV-based monitoring approaches. Real-time parallel inference ensures that safety violations are detected and flagged as they occur, enabling immediate corrective intervention rather than retrospective review. Automated alert dispatch eliminates continuous human supervision, removing fatigue-related monitoring lapses and reducing overall response time. The architecture scales linearly across multiple camera feeds without modification, making it suitable for large-scale construction sites with numerous operational zones.

Table V Safety Hazard Detection Coverage and Alert Mapping

Hazard / Violation	Component	Alert Type
Helmet non-compliance	CNN	Compliance
Safety vest missing	CNN	Compliance
Glove non-compliance	CNN	Compliance
Restricted zone entry	YOLO	Warning
Worker fall detection	YOLO	Emergency
Unsafe machinery proximity	CNN + YOLO	Risk
Safe condition	CNN + YOLO	No action

However, the system is subject to certain deployment limitations that define its operational boundary conditions. Detection accuracy degrades under low-illumination environments such as night shifts or poorly lit indoor zones where supplementary lighting infrastructure is absent. The pipeline is also sensitive to input image quality — low-resolution or compressed video streams reduce classification reliability, particularly for fine-grained PPE detection at distance. Additionally, continuous worker surveillance raises data privacy considerations, and deployment must comply with applicable data protection regulations through transparent disclosure and appropriate data governance policies.

VII. CONCLUSION AND FUTURE WORK

A. Conclusion

This paper presented a Hybrid CNN-YOLO deep learning framework for automated real-time safety monitoring in construction environments. The proposed system addresses the critical limitations of traditional manual supervision and static CCTV-based monitoring by integrating a Convolutional Neural Network for PPE compliance classification with a YOLO-based object detection model for spatial hazard identification. The parallel processing architecture enables simultaneous execution of both models on each input frame, ensuring low-latency inference without sacrificing detection accuracy. A fusion layer combines the outputs of both components, and a rule-based decision engine generates tiered automated alerts corresponding to the severity of detected safety violations. Experimental evaluation on a real-world annotated

dataset demonstrated that the hybrid system achieves an overall accuracy of 97.8%, a Precision of 96.9%, a Recall of 97.3%, and an F1 Score of 97.1%, outperforming standalone CNN and YOLO baselines by 6.6 and 4.4 percentage points respectively. These results confirm that the complementary integration of classification and detection within a unified hybrid architecture provides a robust, scalable, and practically viable solution for intelligent workplace safety enforcement in dynamic construction environments.

B. Future Work

Several directions are identified for extending the capabilities of the proposed system. First, the incorporation of temporal modelling through Long Short-Term Memory (LSTM)

networks or transformer-based video understanding architectures would enable the system to analyses sequences of frames over time, improving the detection of multi-step hazardous events such as gradual encroachment into restricted zones or early-stage fall precursors that are not apparent from individual frame analysis. Second, low-light and adverse-weather robustness can be improved through training data augmentation with synthetic illumination degradation and the integration of infrared or thermal imaging inputs, extending reliable system operation to night shifts and outdoor environments under rain or fog. Third, the adoption of continual learning strategies would allow the detection models to incrementally update from newly encountered construction scenarios and worker activities without requiring complete retraining, addressing performance drift as site conditions evolve over time. Finally, the integration of Explainable AI techniques such as Grad-CAM visualizations and SHAP-based feature attribution would provide safety supervisors with interpretable, human-readable justifications for each generated alert, improving operator trust and enabling more effective incident investigation and regulatory compliance reporting.

REFERENCES

[1] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 2014, pp. 580–587.

[2] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 779–788.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770–778.

[4] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 1–9.

[5] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Commun. ACM, vol. 60, no. 6, pp. 84–90, 2017.

[7] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proc. IEEE, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[8] H. Guo, Y. Yu, and M. Skitmore, "Visualization technology-based construction safety management: A review," Autom. Constr., vol. 73, pp. 135–144, Jan. 2017.

[9] X. Fang, L. Che, M. Shahi, and F. Golparvar-Fard, "Automated detection of workers and personal protective equipment using deep learning and a hard hat dataset," Autom. Constr., vol. 119, p. 103325, Nov. 2020.

[10] J. Redmon and A. Farhadi, "YOLOv3: An incremental improve-ment," arXiv preprint arXiv:1804.02767, Apr. 2018. [Online]. Available: https://arxiv.org/abs/1804.02767