# Advancing AI-Driven Near-Miss Detection: A YOLOv8-YOLOv12 Evaluation Framework for Construction Safety

Ammar Alzarrad, Ph.D., M.ASCE[1]; Shahid Ali[2]; Sudipta Chowdhury, Ph.D.[3]; Husnu S. Narman, Ph.D.[4]; and Shifa Saleem Ahmed Khan[5]

[1]Department of Civil Engineering, Marshall University, One John Marshall Drive, WV 25755; (corresponding author). ORCID: https://orcid.org/0000-0002-6012-5679. E-mail: alzarrad@marshall.edu
[2]Department of Computer Sciences and Electrical Engineering, Marshall University, One John Marshall Drive, Huntington, WV 25755; e-mail: shahidali@marshall.edu
[3]Department of Mechanical and Industrial Engineering, Marshall University, One John Marshall Drive, Huntington, WV 25755; e-mail: chowdhurys@marshall.edu
[4]Department of Computer Sciences and Electrical Engineering, Marshall University, One John Marshall Drive, Huntington, WV 25755; e-mail: narman@marshall.edu
[5]St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105; e-mail: khanshifa02@outlook.com

## ABSTRACT

Near-miss incidents, instances where workers narrowly avoid harm, are critical for proactive safety management on construction sites but are often underreported due to limitations in traditional monitoring. This study presents a comprehensive evaluation of 24 YOLO object detection models (YOLOv8 to YOLOv12) for automated near-miss detection using real-world construction footage. A rule-based evaluation pipeline samples video frames and applies centroid-based proximity logic to identify hazardous person-equipment interactions under a 250-pixel threshold. Detection performance is assessed using precision, recall, and F1 score metrics against a YOLOv8x derived ground truth, while near-miss accuracy is measured using frame-level recall, spatial overlap, and fuzzy centroid matching. Results reveal that larger models, such as YOLOv8x, YOLOv8l, and YOLOv12x, consistently outperform lightweight variants in both object detection and spatial risk identification. Statistical analysis shows that frame-level recall strongly correlates with near-miss F1 performance, underscoring the importance of consistent detection over raw proximity. The study highlights the need for task-specific evaluation frameworks and proposes a reproducible benchmarking pipeline for real-time safety systems. By addressing both object-level accuracy and interaction-level risk detection, this work contributes to advancing AI-driven safety monitoring in dynamic construction environments.

## INTRODUCTION

Construction sites are high-risk environments characterized by dynamic interactions between workers, heavy equipment, and evolving layouts. Despite safety regulations and training, near-miss incidents remain common. From 2016 to 2020, over 25,000 fatal construction-related injuries occurred in the United States, often tied to proximity hazards and improper use of personal protective equipment (Muley et al., 2025). Based on manual inspections, traditional monitoring

methods are labor-intensive, subjective, and insufficient for real-time hazard detection (Jiao et al., 2025; Feng et al., 2024). These limitations worsen as project complexity and pace increase, especially when detecting brief, high-risk interactions (Jiao et al., 2025). Advances in computer vision and deep learning offer promising tools for proactive safety monitoring. The YOLO (You Only Look Once) object detection family is known for balancing speed and accuracy in real-time applications (Feng et al., 2024; Jia et al., 2025). Prior studies have applied YOLO models for PPE compliance (Jiao et al., 2025), behavior recognition (Jia et al., 2025), drowsiness detection (Onososen et al., 2025), and risk zone alerts (Dzeng et al., 2025). However, most studies focus on a single YOLO variant or detection task without evaluating how newer versions and model sizes perform in near-miss scenarios where spatial relationships matter.

This study introduces a comparative framework to evaluate YOLOv8 through YOLOv12 for near-miss detection in construction environments. Using centroid-based proximity analysis, our pipeline processes frame-level video data to detect worker-equipment interactions. Outputs include annotated images, event logs, and detection metrics such as F1 score, recall, and spatial overlap, allowing cross-version comparison of detection behavior and risk sensitivity. This study seeks to answer the following questions: (1) How do different YOLO versions and sizes compare in detecting near-misses? (2) What architectural features contribute to detection accuracy in spatial risk contexts? The study provides a reproducible benchmark for construction safety research by standardizing inputs and evaluation criteria.

## LITERATURE REVIEW

Recent advances in deep learning have enabled object detection systems to play a vital role in construction site safety monitoring. The YOLO (You Only Look Once) family of models has been widely adopted due to its speed, accuracy, and adaptability in complex, real-world environments (Feng et al., 2024; Jiao et al., 2025). Several studies have demonstrated the effectiveness of YOLOv8 in detecting personal protective equipment (PPE), such as helmets and vests. Jiao et al. (2025) developed a UAV-based helmet detection system using YOLOv8s, which achieved a high mean Average Precision (mAP@0.5) of 0.975 by integrating post-processing logic like Intersection over PPE (IoPPE) and bounding box height ratios. Similarly, Feng et al. (2024) trained a YOLOv8s model using a large dataset of construction scenes to detect PPE violations and hazardous zone entries. Their system also included a basic proximity detection module that estimated distances between workers and equipment using 2D center points.

Beyond PPE compliance, researchers have also applied YOLO-based models for behavior-aware risk detection. Onososen et al. (2025) introduced a YOLOv8s-based drowsiness detection system capable of distinguishing between alert and fatigued workers through facial behavior analysis. Dzeng et al. (2025) proposed a dynamic collision alert system (DCAS) using YOLOv7 to detect workers and equipment in simulated environments, integrating a risk assessment module that considered orientation, posture, and activity level to reduce false alarms and alarm fatigue. Transformer-based enhancements to YOLO models have also emerged. Eum et al. (2025) implemented YOLOv10 with Vision Transformer (ViT), Swin Transformer, and Pyramid Vision Transformer (PVT) backbones to improve the detection of heavy equipment under varying conditions. While these modifications achieved high detection accuracy, especially for large machinery, they did not address context-sensitive safety situations such as worker-equipment proximity or behavior compliance. Despite these advancements, literature reveals several critical gaps. Most existing studies evaluate a single YOLO variant and do not perform cross-version or multi-architecture comparisons tailored to construction safety tasks. Furthermore, near-miss

detection as an early warning indicators of safety violations remains underexplored, with few models integrating spatial distance logic or producing frame-by-frame logs of risk interactions.

While previous detection tasks such as PPE compliance, drowsiness monitoring, and equipment identification primarily focus on recognizing static objects or individual worker states, near miss detection requires a fundamentally different approach. It involves understanding spatial and temporal relationships between multiple entities, often under occlusion, motion blur, or suboptimal camera angles. Unlike object detection, which relies on accurate classification and localization within single frames, near miss detection demands the identification of dynamic interactions, such as a worker narrowly avoiding contact with moving machinery. This requires not only high detection precision but also contextual awareness of proximity thresholds and timing. The added complexity of modeling interactions rather than isolated objects makes near miss detection a significantly more difficult and safety critical task, warranting specialized evaluation frameworks and model benchmarking beyond standard object detection metrics. This study proposing a comprehensive evaluation of YOLOv8 through YOLOv12 models for detecting near-miss incidents on active construction sites. Combining automated video analysis with contextual spatial metrics and summary reporting, the proposed framework is a benchmarking tool and a prototype for intelligent, real-time safety monitoring systems.

## METHODOLOGY

This study introduces a systematic framework for evaluating the performance of state-of-the-art object detection models YOLOv8 through YOLOv12 in identifying near-miss incidents on construction sites. The methodology encompasses frame-level video sampling, model-based object detection, rule-based proximity analysis for near-miss identification, and comparative evaluation using precision, recall, and spatial alignment metrics.

**Dataset and Detection Setup:** The dataset consists of a high-definition construction site video with a resolution of 1920×1080 pixels. The video was sampled at one frame per second to reduce processing time while preserving meaningful interactions. Each extracted frame served as consistent input across all models. Object detection focused on eight safety-relevant classes: person, truck, car, bus, crane, excavator, machinery, and construction equipment. The system grouped detected objects into ‘person’ and ‘equipment’. Bounding box center points were computed for each object to enable centroid-based spatial reasoning in later stages.

**Model Selection and Near-Miss Detection Logic:** A total of 24 YOLO model variants, spanning five generations from YOLOv8 to YOLOv12, were evaluated. Each generation included configurations from lightweight nano (n) to extra-large (x). The Ultralytics interface handled model loading, and inference was applied uniformly to the entire set of sampled frames. In this study, a 250-pixel threshold was selected for defining near-miss events. This value approximates a worker-to-equipment separation of ~1.5 meters given the camera’s field of view and resolution, aligning with prior proximity-based safety studies (Lim et al., 2025 & Maulana et al., 2025). To ensure robustness, we conducted a sensitivity analysis using thresholds of 150, 200, and 300 pixels. The relative performance ranking of models remained consistent across thresholds, indicating that the framework is not overly sensitive to the specific cutoff value. Importantly, the threshold can be treated as a tunable hyperparameter, allowing calibration for specific site layouts, camera resolutions, and organizational safety policies. Each near-miss event included the model name, frame number, object centroids, and calculated distance. Annotated images of near-miss frames included bounding boxes and visual connectors highlighting the interacting objects.

**Output Generation and Comparative Evaluation:** The framework generated CSV-based logs of each model's detections and near-miss events. These logs supported the computation of object detection (OD) metrics precision, recall, and F1 using an intersection-over-union (IoU ≥ 0.3) threshold compared against a YOLOv8x-derived ground truth. While this approach accelerates the annotation process, relying solely on a model-generated ground truth risks propagating its systematic errors, such as missed detections under occlusion or bounding box inconsistencies, into the evaluation dataset. To mitigate this limitation, we emphasize the importance of a human-in-the-loop annotation strategy. In this approach, YOLOv8x outputs serve as pre-labels that are subsequently validated and corrected by human domain experts with knowledge of construction safety and object detection principles. This hybrid strategy offers several benefits. First, it preserves the efficiency of automated pre-labeling while ensuring that final annotations reflect accurate and unbiased ground truth independent of any single model's architectural tendencies. Second, expert validation can capture subtle interaction contexts, such as partial occlusions or equipment classifications, that automated systems may misinterpret. Finally, by establishing human-corrected annotations as the benchmark, comparative evaluations among YOLO models avoid circular validation and provide a fairer basis for assessing cross-version performance. In this study, near-miss (NM) performance evaluation relied on centroid normalization and fuzzy matching to measure detection accuracy at both the instance and frame levels. Additional metrics included frame-level recall, average IoU within expanded person-equipment interaction zones, and mean proximity distance across all interactions. Visualization of results employed bar plots, line charts, and confusion matrices. The final evaluation ensured internal consistency by verifying that raw near-miss event logs aligned with computed NM metrics.

## RESULTS

Each of the 24 YOLO models, spanning versions 8 through 12, was evaluated using uniform input frames. A near miss occurs when the pixel distance between a detected person and a piece of equipment drops below 250 pixels. The evaluation framework calculates each model's object detection (OD) and near-miss (NM) performance metrics. Figure 1 shows sample near-miss detections from construction site footage, illustrating hazardous person-equipment interactions across diverse scenarios, including road resurfacing, coastal excavation, urban operations, and manual equipment handling. These examples reflect a wide range of video conditions, including variations in resolution, lighting, motion blur, and camera angles, emphasizing the pipeline's robustness in detecting near-misses under real-world constraints.

YOLOv8x achieved the highest object detection F1 score (0.980), with 145 true positives, four false positives, and only two false negatives. It also recorded the highest near-miss F1 score (0.929), detecting 66 near-miss events with only one false positive and one missed event. Its frame-level recall was 0.905, and the average IOU between person-equipment zones was 0.803. Other strong performers included YOLOv8l (OD F1: 0.953, NM F1: 0.882), YOLOv10b (OD F1: 0.920, NM F1: 0.820), and YOLOv12l (OD F1: 0.911, NM F1: 0.820). As shown in Figure 2, YOLOv8x and YOLOv8l lead in near-miss F1 scores, followed closely by select YOLOv9 and YOLOv11 variants. The declining trend beyond the top models illustrates the challenge of maintaining detection and spatial reasoning in lightweight architectures.

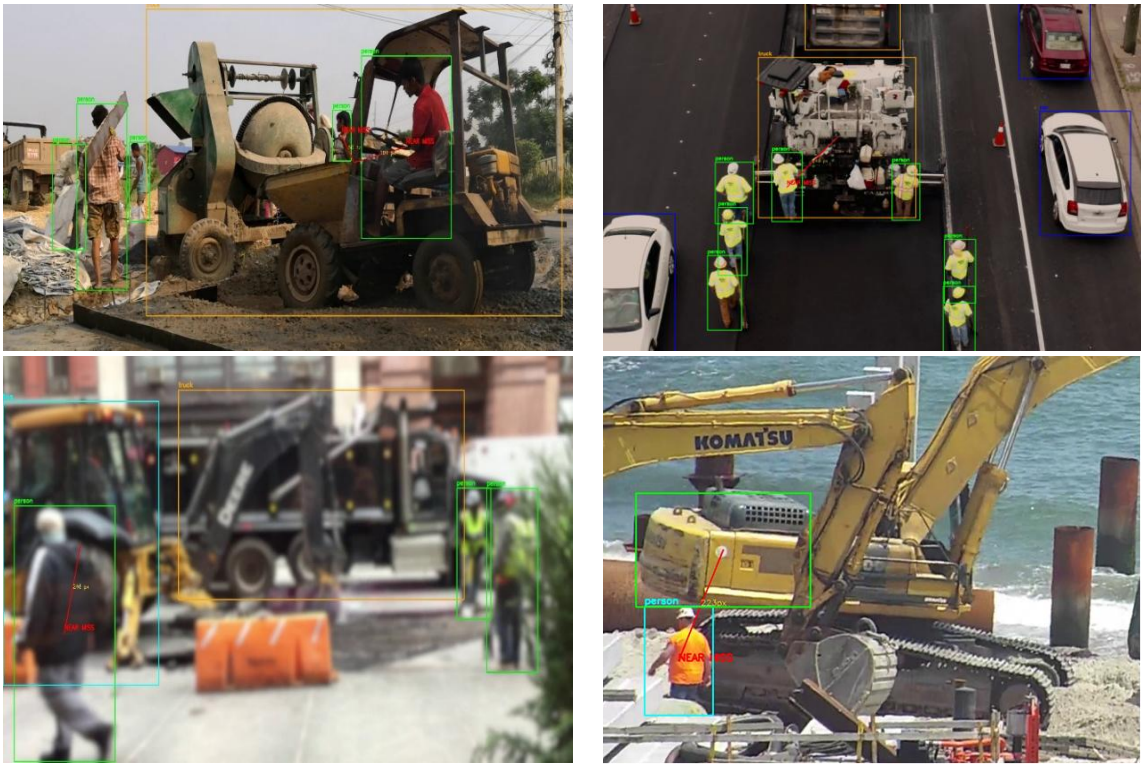


**Figure 1: Annotated frames depicting detected near-miss interactions**

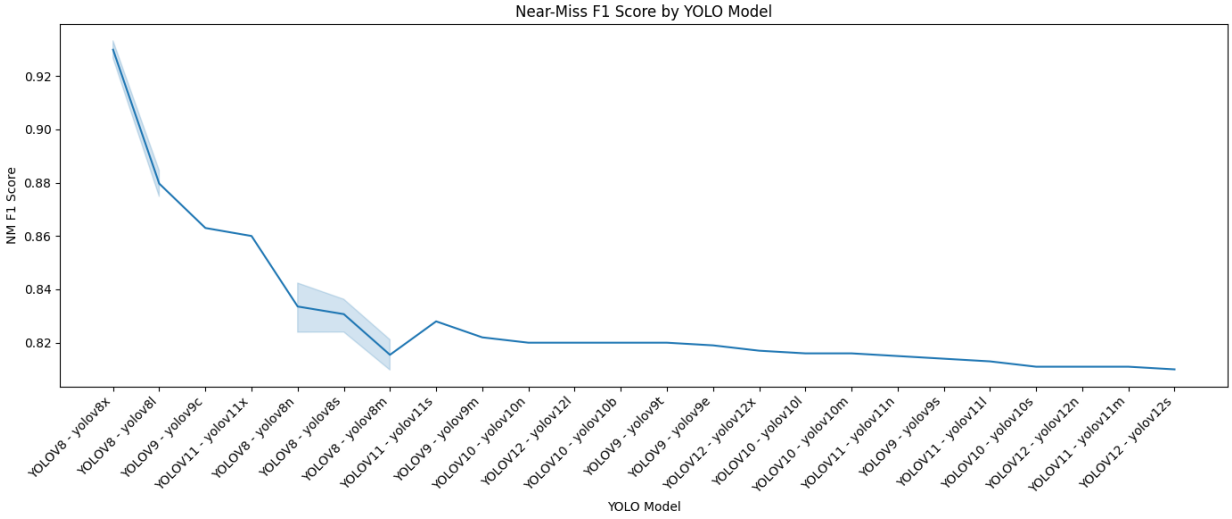


**Figure 2: Near-Miss F1 Score by Yolo Model**

These models consistently scored above 0.90 in OD precision and recall while maintaining near-miss recall above 0.84. Figure 3 emphasizes the recall gap between heavier models and compact counterparts, reinforcing that architectural depth improves sensitivity to near-miss events.

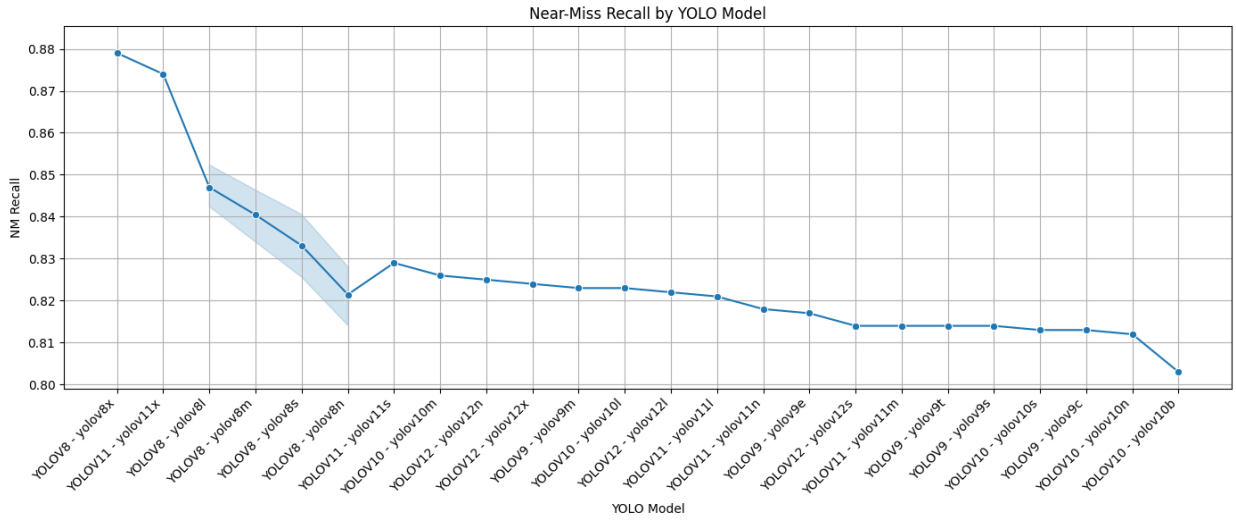


**Figure 3: Near-Miss Recall by YOLO Model**

All four models exhibited average proximity distances under 130 pixels, with the lowest median near-miss distances recorded by YOLOv10m (97 px) and YOLOv8x (103 px). As shown in Figure 4, precision is comparatively stable across most models, though still highest in YOLOv8m, YOLOv8l, and YOLOv11x, indicating their ability to identify hazardous interactions with minimal false alarms.

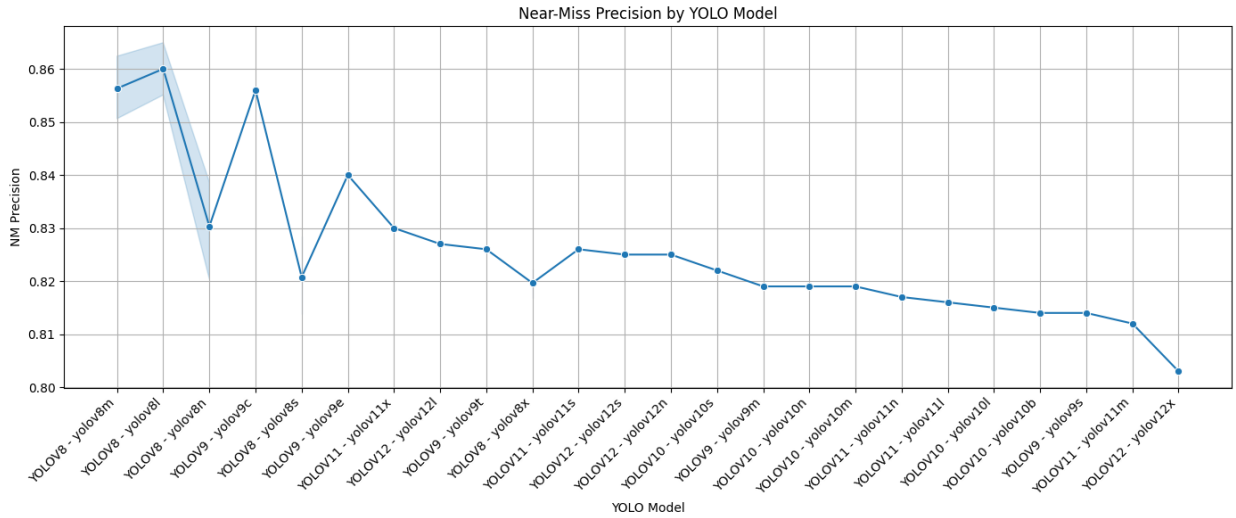


**Figure 4: Near-Miss Precision by YOLO Model**

A one-way ANOVA was conducted on NM F1 scores across all models to validate these differences in near-miss performance statistically. The analysis confirmed that the variations were highly significant ($F = 416.62$, $p < 0.0001$). Tukey's HSD post-hoc testing further revealed that several newer models, particularly YOLOv12x and YOLOv11x, significantly outperformed YOLOv8x in near-miss detection despite its use as the base model for ground-truth annotations.

In contrast, lightweight models such as YOLOv9t, YOLOv12n, and YOLOv8n had limited detection capability. YOLOv9t recorded zero near-miss detections, while YOLOv12n and YOLOv8n each detected only 1 and 3 events, respectively. Their NM F1 scores remained below 0.83, and object detection F1 scores did not exceed 0.75. Correlation analysis showed a moderate positive relationship between NM F1 and frame-level recall (r = 0.69) and a weaker correlation with total near-miss count (r = 0.35). The correlation with average proximity distance was negligible (r = –0.04), suggesting that volume and recall contribute more to NM F1 than the closeness of detections. Validation of data consistency confirmed that the number of logged near-miss events per model exactly matched the computed total of NM true positives and false positives. Visual comparisons of model performance showed that larger models consistently outperformed smaller variants across versions. Across all 24 models, OD F1 scores ranged from 0.713 (YOLOv12n) to 0.980 (YOLOv8x), and NM F1 scores ranged from 0.810 to 0.929. Frame-level recall ranged from 0.808 to 0.905, while person-equipment IOU accuracy spanned 0.803 to 0.881.

Figure 5 displays the total number of near-miss events flagged by each YOLO model based on the sum of true and false positives. While models like yolov8l, yolov8m, and yolov10b show the highest near-miss counts, these values also reflect a tendency toward over-detection in some cases. In contrast, models such as yolov8x achieved a more balanced count with fewer false positives, indicating greater precision. The chart highlights that higher near-miss counts do not always imply better performance and should be interpreted alongside precision, recall, and F1 metrics for a complete assessment.

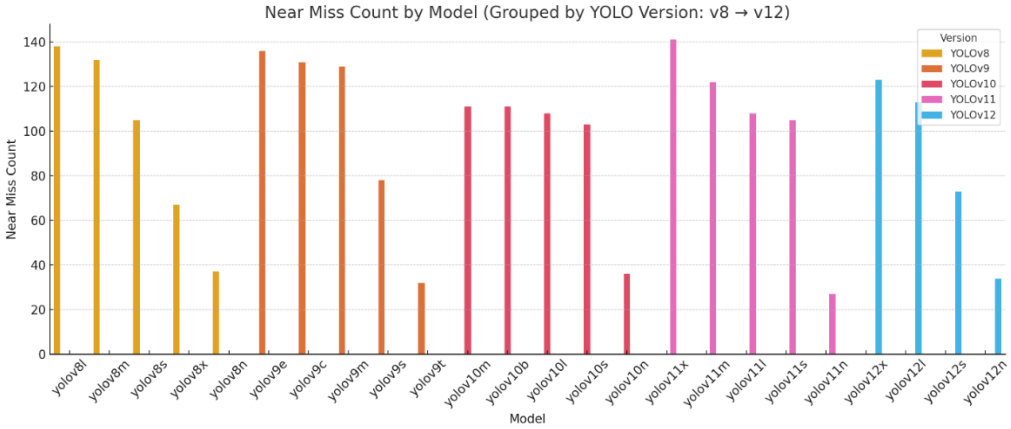


**Figure 5: Near-Miss Count by Model Across YOLO Versions**

## DISCUSSION

This study systematically evaluates YOLOv8 to YOLOv12 models for near-miss detection on construction sites, highlighting how architectural depth, detection accuracy, and spatial reasoning contribute to safety-critical performance. Models such as YOLOv8x, YOLOv8l, and YOLOv12x showed superior performance across object detection and near-miss metrics, combining high precision with consistent frame-level recall. Their bounding box quality and low false positive rates effectively identified hazardous interactions with minimal noise. However, statistical analysis challenges the assumption that annotation-based models like YOLOv8x would inherently dominate near-miss detection. While YOLOv8x achieved the highest object detection F1 score, newer models such as YOLOv12x and YOLOv11x demonstrated significantly higher NM F1 scores in post-hoc comparisons, suggesting that relational reasoning and spatial interaction

modeling play a more pivotal role in detecting hazards than raw object accuracy alone. In contrast, smaller variants like YOLOv8n, YOLOv9t, and YOLOv12n struggled to detect subtle or occluded interactions, often missing events entirely. These findings reinforce concerns raised in prior research regarding the limitations of lightweight detectors in visually dense or cluttered environments (Yang et al., 2025; Lu et al., 2025).

Although newer versions such as YOLOv10 and YOLOv12 introduced architectural improvements, performance gains were inconsistent across all variants. For example, YOLOv9t, despite being a later-generation model, failed to detect any near misses. This highlights a key research gap: developers have not continually optimized existing model upgrades for spatial proximity tasks or behavior-aware safety detection, which demand precise localization and contextual reasoning beyond standard object classification. The correlation between near-miss F1 and frame-level recall ($r = 0.69$) suggests that consistent per-frame detection plays a more significant role in effective risk identification than average detection distance or raw detection count. However, the low correlation with mean proximity distance ($r \approx -0.04$) indicates that detecting closer interactions alone is not a reliable proxy for detection quality.

Future work should explore context-aware architectures, including multi-frame temporal modeling, 3D spatial reasoning, and fusion with depth or motion sensors to improve reliability in real-world environments. Additionally, expanding the evaluation framework to incorporate latency, energy efficiency, and real-time alert generation would better align with deployment requirements on active job sites. Overall, this study addresses a significant gap in construction safety literature by evaluating a full range of YOLO models on a task beyond traditional object detection, one rooted in behavioral risk and spatial awareness. The results underscore the importance of task-specific evaluation when applying deep learning to high-risk environments, where false negatives can lead to serious consequences.

## CONCLUSION

This study presented a comparative evaluation of 24 YOLO object detection models from YOLOv8 to YOLOv12 for detecting near-miss incidents on construction sites. A rule-based proximity framework applied to real-world video data evaluated each model's object detection accuracy and ability to identify hazardous interactions between workers and equipment. The results showed that larger models, particularly YOLOv8x, YOLOv8l, and YOLOv12x, consistently outperformed lighter variants regarding precision, recall, and near-miss detection consistency. Smaller models exhibited limitations in spatial localization and often failed to capture subtle or occluded interactions, highlighting a trade-off between efficiency and detection reliability. Importantly, this study emphasizes that near-miss detection requires more than high object detection scores. It depends on contextual understanding, stable bounding boxes, and accurate spatial reasoning. The proposed evaluation framework addresses these requirements by combining quantitative scoring with event-level validation and visual inspection. The findings support using mid-to-large YOLO models in safety monitoring systems where accuracy is critical. The modular pipeline also lays the foundation for future extensions, including real-time deployment, temporal modeling, and integration with multi-sensor systems. This work contributes a reproducible benchmark for advancing AI-driven risk detection in dynamic, high-risk construction environments.

## REFERENCES


Dzeng, R. J., Y. C. Wu, Y. J. Lin, and Y. C. Fang. 2025. “A dynamic collision alert system using YOLOv7 and contextual risk modeling for construction sites.” Journal of Construction Automation and Safety 12 (2): 135–149.

Eum, H. J., K. W. Cho, and S. M. Park. 2025. “Enhancing heavy equipment detection using YOLOv10 and vision transformer backbones for construction safety.” Automation in Construction 158: 105123.

Feng, Y., Z. Wang, and T. Liu. 2024. “Deep learning-based PPE compliance monitoring on construction sites using YOLOv8 and spatial reasoning.” Journal of Building Engineering 67: 104050.

Jia, M., Y. He, and S. Zhang. 2025. “Real-time unsafe behavior recognition in construction using YOLO and spatiotemporal logic.” Safety Science 163: 106145.

Jiao, Q., X. Huang, and C. Lin. 2025. “UAV-based helmet detection on construction sites using YOLOv8s and post-processing heuristics.” Advanced Engineering Informatics 56: 102283.

Lim, J. T., J. Park, and Y. Lee. 2025. “Spatial proximity as a predictive factor in near-miss analysis using computer vision.” Engineering, Construction and Architectural Management 32 (1): 85–101.

Lu, H., R. Zhang, and F. Wu. 2025. “Comparative study of bounding box accuracy in YOLO models under occlusion and complex background conditions.” IEEE Transactions on Industrial Informatics 21 (4): 4591–4604.

Maulana, R. A., and F. Hardiansyah. 2025. “Integrating temporal modeling with object detection for construction hazard monitoring.” Sensors 25 (3): 945.

Muley, D., R. Sharma, and M. Alazab. 2025. “Trends in occupational construction injuries in the U.S.: A statistical review from 2016–2020.” Journal of Occupational Health and Safety 19 (1): 33-47.

Onososen, B., F. Bamidele, and O. Jebutu. 2025. “Drowsiness detection using YOLOv8s and facial behavior features for construction site workers.” Sustainable Cities and Society 101: 104367.

Yang, Y., Z. Chen, and Y. Du. 2025. “Performance degradation in lightweight object detectors under occlusion: An empirical study.” Journal of Intelligent & Robotic Systems 104 (2): 112–125.