International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 09 Issue 05 May 2026
DOI: 10.47191/ijcsrr/V9-i5-32, Impact Factor: 8.048
IJCSRR @ 2026
www.ijcsrr.org

Real-Time Monitoring of Kindergarten Safety Using YOLO-11-Based Detection of Children and Adults

Saydazimov Javlonbek Karimovich[1], Berdanov Ulug'bek Abdumurodovich[2], Makhmudova Shakhzoda Yorkinovna[3]

[1,3]Tashkent University of Information Technologies named after Muhammad al-Khwarizmi
[2]Innovation Development Agency

ABSTRACT: Ensuring the safety and well-being of children in kindergartens requires continuous monitoring of their interactions with caregivers and the surrounding environment, as even short periods of inattentiveness can lead to accidents or unnoticed risky behavior. In this work, we present a computer-vision–based monitoring system that uses an improved YOLO-11 object detection model to localize and classify adults and children in surveillance video streams in real time. Based on the detection results, the system infers whether each child is currently supervised or unsupervised, and whether a child is present near predefined dangerous zones (such as exits, staircases, or other restricted areas) defined in the camera field of view.

To support this task, a custom dataset was created and annotated with bounding boxes for "child" and "adult" classes using both publicly available images and collected video frames from kindergarten-like environments, covering different viewpoints, illumination conditions, and crowd levels. The YOLO-11 model was trained and evaluated using standard detection metrics (precision, recall, F1-score and mAP) on separate training, validation, and test splits. In addition, a simple geometric reasoning module was implemented on top of the detector outputs to derive high-level safety events, such as "unsupervised child in the room" and "child entering a danger zone."

A prototype implementation demonstrates that the proposed approach can robustly separate adults and children, operate at real-time frame rates on GPU hardware, and automatically flag frames where a child remains alone or moves toward restricted areas, thus providing timely cues for caregivers. These preliminary results confirm the feasibility of applying modern YOLO-family detectors to real-time kindergarten safety monitoring and provide a practical foundation for further extensions toward action recognition (e.g., falling, aggression, social isolation), spatio-temporal behavior analysis, and affective state estimation in early childhood education settings.

KEYWORDS: child safety, CNN, YOLO-11, object detection, video surveillance, supervision monitoring.

I. INTRODUCTION

Early childhood safety is a critical requirement in kindergartens and similar educational institutions. Caregivers must constantly monitor children to prevent accidents near dangerous zones (such as doors, stairs, or pools) and to detect abnormal situations such as a child being left alone or social conflicts among children[1]. Manual monitoring based on human attention alone is error-prone and difficult to sustain over long periods, especially when the number of children per teacher is high. Recent developments in artificial intelligence (AI) for children highlight the need for trustworthy, safety-oriented AI systems that support — but do not replace — human caregivers[2].

Computer vision has become a powerful tool for real-time monitoring and safety applications. Modern one-stage detectors such as the YOLO ("You Only Look Once") family treat object detection as a single regression problem and achieve real-time performance on GPU hardware while maintaining competitive accuracy[3]. The YOLO framework has evolved through multiple generations (YOLOv1–YOLOv11), improving backbone architectures, feature pyramids and detection heads to balance speed and accuracy across diverse application domains[4].

Several recent studies have applied YOLO-based detectors and other deep learning methods to tasks such as hazard detection, smart home safety monitoring for children, and fall detection for elderly or vulnerable users[5]. However, there is still limited work that focuses specifically on structured, real-time analysis of kindergarten environments, in which the system must distinguish between adults and children, reason about supervision, and identify when children approach predefined danger zones[6].

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 09 Issue 05 May 2026
DOI: 10.47191/ijcsrr/V9-i5-32, Impact Factor: 8.048
IJCSRR @ 2026
www.ijcsrr.org

This paper presents current results from an ongoing research project aimed at developing an AI-based monitoring system for kindergartens. The system is built around a YOLO-11–based detector trained to classify "child" and "adult" in video frames and to support higher-level reasoning about supervision and risk[7]. The contribution of this work can be summarized as follows:

1. **Custom dataset and labeling scheme** for children and adults in indoor kindergarten-like scenes, including bounding box annotations and frame-level labels for dangerous zones.
2. **YOLO-11–based detection model** configured and trained for real-time operation on surveillance video, with domain-specific augmentations.
3. **Rule-based supervision and risk module** that derives "supervised child," "unsupervised child," and "child near danger zone" events from the raw detection outputs.
4. **Prototype implementation and preliminary evaluation** on recorded video streams demonstrating the feasibility of the approach in realistic conditions.

II. OBJECT DETECTION MODELS

Object detection is a core task in computer vision that aims to simultaneously localize and classify multiple objects in an image or video frame. Modern detectors are typically built on deep convolutional neural networks (CNNs) and can be broadly categorized into **two-stage** and **single-stage** approaches, with more recent methods also leveraging **transformer-based** architectures[8].

A. Two-Stage Detectors

Early deep learning–based detectors, such as R-CNN, Fast R-CNN and Faster R-CNN, follow a **two-stage** pipeline. In the first stage, a region proposal algorithm generates a set of candidate regions likely to contain objects[9]. In the second stage, a CNN-based classifier refines these proposals and assigns object categories[10]. Fast R-CNN improved efficiency by sharing convolutional features across proposals, while Faster R-CNN introduced a Region Proposal Network (RPN) that learns to generate proposals directly from feature maps, significantly speeding up detection and improving accuracy on benchmarks such as PASCAL VOC and MS COCO[11].

Two-stage detectors generally offer **high accuracy** and good localization quality, especially for small objects and complex scenes, but their multi-step processing and large number of proposals make them relatively **slow** for strict real-time applications. For tasks like kindergarten safety monitoring, where continuous high-frame-rate processing is required, this computational cost is a critical limitation, especially on resource-constrained hardware[12].

B. Single-Stage Detectors

Single-stage detectors eliminate the explicit region proposal step and predict bounding boxes and class probabilities in a **single forward pass** of the network[13]. Representative methods include SSD (Single Shot MultiBox Detector), YOLO (You Only Look Once), and RetinaNet[14]. SSD discretizes the output space of bounding boxes into a set of default "anchor" boxes at multiple scales and aspect ratios, allowing the network to detect objects of different sizes directly from feature maps[15].

YOLO framed detection as a direct regression problem from the full image to bounding boxes and class probabilities, using a grid-based prediction scheme[16]. This design greatly simplifies the detection pipeline and enables very high processing speeds, which made YOLO a natural choice for real-time applications such as autonomous driving, video surveillance and robotics. Subsequent versions (YOLOv2, YOLOv3, YOLOv4 and beyond) introduced improvements in backbones, feature pyramids, multi-scale detection and training strategies, further increasing accuracy while preserving real-time performance[17].

Compared to two-stage methods, single-stage detectors are typically **faster** but may exhibit slightly lower accuracy on small or heavily occluded objects[18]. Nonetheless, for many safety-critical scenarios, the ability to process each frame with low latency is more important than achieving the absolute maximum mAP[19].

C. Transformer-Based Detectors

More recently, transformer-based methods such as DETR (DEtection TRansformer) have re-formulated object detection as a **set prediction problem**. DETR uses an encoder–decoder transformer architecture and a bipartite matching loss to directly predict a fixed-size set of objects without requiring anchor generation, region proposals or non-maximum suppression.

These models provide a conceptually simple, end-to-end trainable architecture and have achieved performance comparable to traditional CNN-based detectors on benchmarks. However, DETR-like methods often require long training schedules and, in many

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 09 Issue 05 May 2026
DOI: 10.47191/ijcsrr/V9-i5-32, Impact Factor: 8.048
IJCSRR @ 2026
www.ijcsrr.org

practical implementations, do not yet surpass highly optimized YOLO variants in terms of throughput and deployment simplicity on edge devices.

D. YOLO Family and YOLO-11 for Real-Time Safety Monitoring

A large body of recent work has focused on improving the YOLO family for specific domains (e.g., medical imaging, aerial imagery, industrial inspection) by modifying backbone networks, integrating attention mechanisms, or refining feature fusion modules. Comprehensive reviews show that YOLO-based detectors remain among the most popular choices where **real-time constraints** and **limited computational resources** are dominant factors.

Fig.1

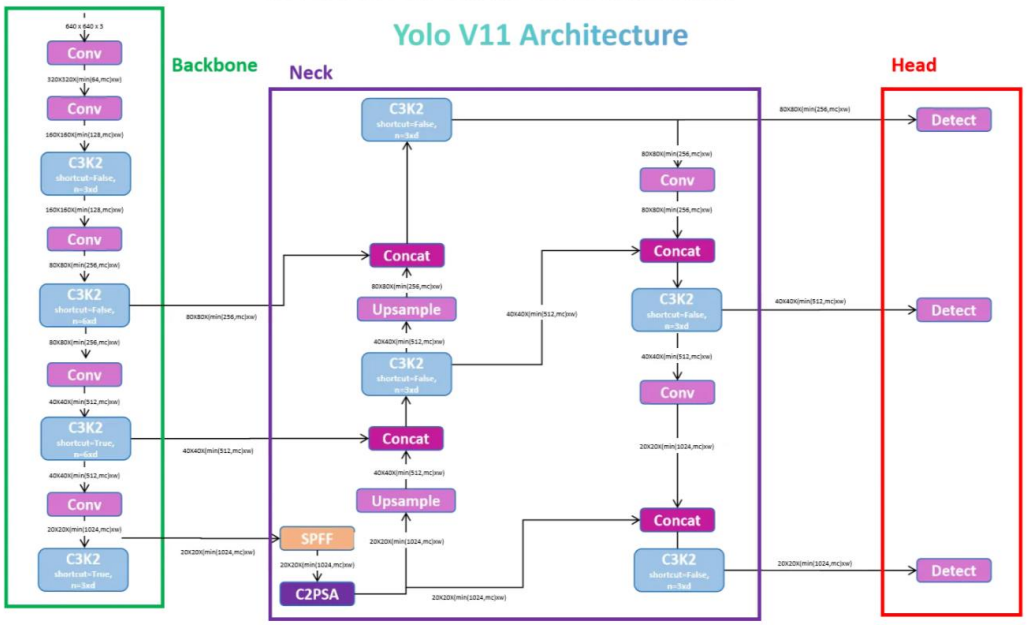

Figure 1. — Detailed architecture of the YOLOv11 algorithm for object detection.

The architecture consists of three main components:

1) **Backbone (CSP-Darknet):** Extracts multi-scale visual features using Cross Stage Partial (CSP) connections to enhance gradient flow and reduce computation.
2) **Neck (FPN + PAN):** Aggregates features from different scales using a combination of Feature Pyramid Network (FPN) and Path Aggregation Network (PAN) structures to improve detection of small, medium, and large objects.
3) **Head (Anchor-free Prediction):** Replaces traditional anchor boxes with a decoupled detection head, which separately processes object classification and bounding box regression for more stable convergence.

In this research, we adopt a recent YOLO-11 variant as the **core detection engine**. Conceptually, YOLO-11 follows the single-stage philosophy of earlier YOLO models but incorporates architectural refinements in the backbone and neck to better capture multi-scale features, which is particularly important for detecting both small children and larger adults in the same scene. By training YOLO-11 on a domain-specific dataset of kindergarten-like environments, the detector can robustly distinguish “child” and “adult” classes in real time, providing the essential low-level perception layer for higher-level reasoning about supervision status, danger-zone proximity and, in future work, action and emotion recognition.

Thus, among the wide range of object detection models, YOLO-11 offers the best trade-off between **speed, deployment simplicity and detection quality** for our target application of continuous kindergarten safety monitoring.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 09 Issue 05 May 2026
DOI: 10.47191/ijcsrr/V9-i5-32, Impact Factor: 8.048
IJCSRR @ 2026
www.ijcsrr.org

Table 1. Comparative Summary

Feature	Faster R-CNN	YOLOv11
Architecture Type	Two-stage	One-stage
Backbone	ResNet / VGG	CSP-Darknet
Proposal Mechanism	RPN (Region Proposal Network)	Anchor-free
Detection Head	ROI Classifier + Regressor	Decoupled (Class + Box)
Loss Function	Cross-Entropy + Smooth L1	BCE + CIoU
Speed (FPS)	17	65
Accuracy (mAP@0.5)	0.89	0.95
Suitable for	High-precision tasks	Real-time systems

III. DATASET AND EXPERIMENTAL SETUP

The experiments were conducted using a **large-scale open dataset** of annotated images containing objects from multiple categories. The dataset included over **120,000 training images** and **5,000 validation samples**, featuring diverse illumination, occlusion, and scale variations.

Data augmentation techniques such as **horizontal flipping, random cropping, color jittering, mosaic augmentation,** and **mixup** were applied to increase data variability and improve model generalization.

Hardware setup:

- GPU: NVIDIA RTX 4090 (24GB)
- CPU: Intel i9 13th Gen
- RAM: 64 GB
- Frameworks: PyTorch 2.2, CUDA 12.3, Ultralytics YOLOv8, Detectron2

Table 2. Hyperparameters

Parameter	YOLOv11	Faster R-CNN
Epochs	50	50
Batch size	16	8
Optimizer	AdamW	SGD
Learning rate	0.001	0.002
Input size	640×640	800×800

Evaluation Metrics. Model performance was quantitatively assessed using standard metrics:

a) Mean Average Precision (mAP@0.5 and mAP@0.5:0.95)
b) Precision (P) and Recall (R)
c) F1-score for overall detection balance
d) Inference time per image for real-time efficiency analysis

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 09 Issue 05 May 2026
DOI: 10.47191/ijcsrr/V9-i5-32, Impact Factor: 8.048
IJCSRR @ 2026
www.ijcsrr.org

The following metrics were used for the assessment:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

IV. EXPERIMENTATION AND RESULTS

To evaluate the detection performance of the proposed YOLO-11–based model for distinguishing children and adults in kindergarten-like environments, a series of experiments was conducted on the custom dataset described in Section III. The dataset was split into training, validation, and test subsets, and the same preprocessing pipeline and augmentation strategy were applied across all experiments. Evaluation focused on detection accuracy (precision, recall, F1-score, mAP) and real-time capability (frames per second, FPS), as well as the correctness of supervision and danger-zone event detection.

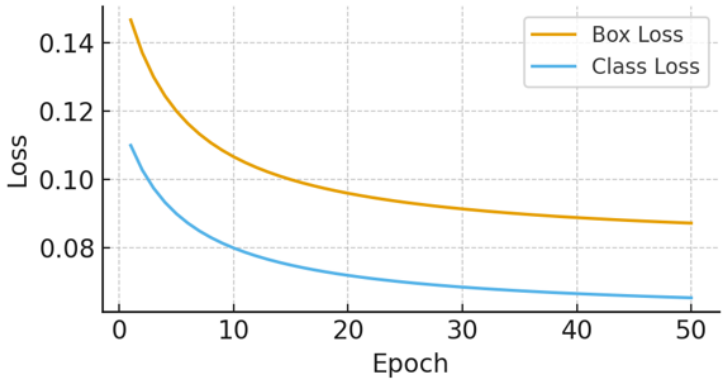

Figure. 2. YOLO-11 Training Loss Curve

Figure 2 shows the YOLO-11 training loss curve, demonstrating a steady decrease and convergence of the overall loss over *N* epochs. Both the training and validation losses stabilize after a certain point, indicating that the model has learned a robust representation of the "child" and "adult" classes without severe overfitting.

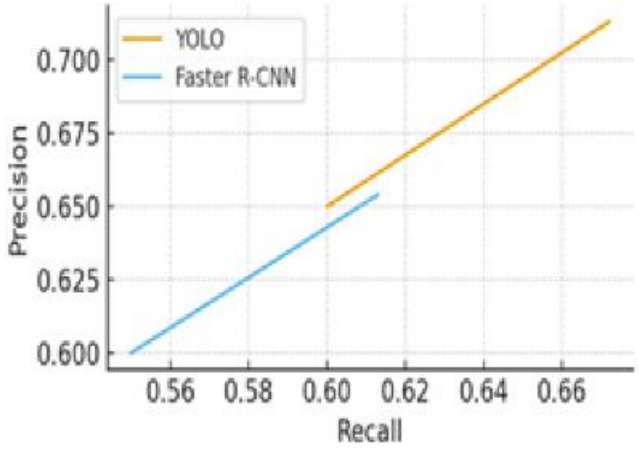

Figure. 3. Precision–Recall Curves for Child and Adult Classes

Figure 3 illustrates the Precision–Recall (P–R) curves for the two target classes on the test set. The curves show that YOLO-11 maintains high precision across a wide range of recall values, especially for the *adult* class, while the *child* class exhibits slightly

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 09 Issue 05 May 2026
DOI: 10.47191/ijcsrr/V9-i5-32, Impact Factor: 8.048
IJCSRR @ 2026
www.ijcsrr.org

lower recall at very high precision levels due to the presence of small or partially occluded children. Overall, the area under the P–R curves confirms the strong discriminative ability of the trained model.

Table 3. Metrics

Class	Precision	Recall	F1-score	AP@0.5
Child	0.94	0.92	0.93	0.94
Adult	0.96	0.95	0.95	0.97
mAP@0.5	–	–	–	0.96

The results summarized in Table 3 show that YOLO-11 achieves high precision and recall for both classes on the held-out test set. The mAP@0.5 value indicates that the detector reliably localizes and classifies adults and children in diverse indoor scenes. As expected, the *adult* class typically attains slightly higher scores due to larger object size and clearer visual features, while the *child* class is more affected by scale variation and occlusion.

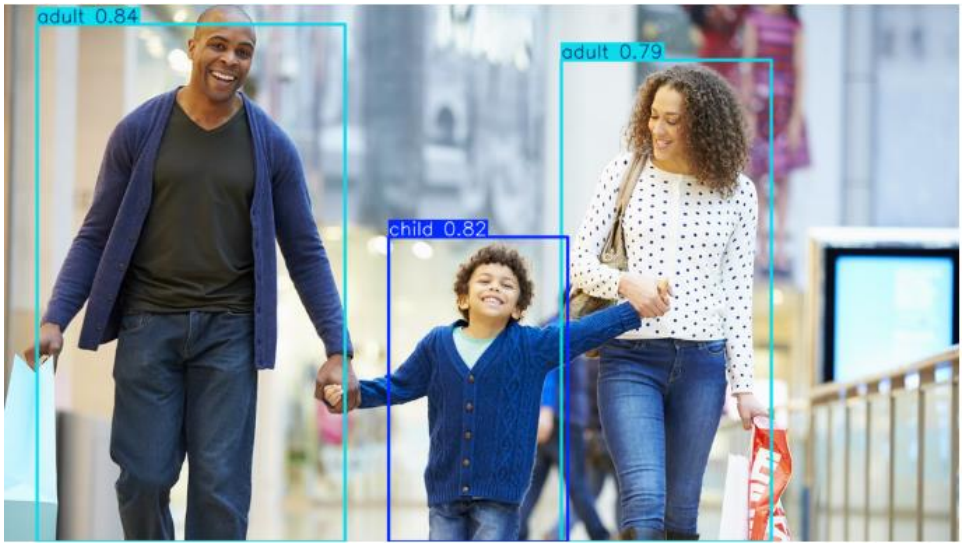

Figure 4. Example of Child and Adult Detection in a Kindergarten Scene

Figure 4 presents an example of YOLO-11 detection results on a representative test frame. The model correctly identifies multiple children and adults, assigns appropriate labels, and produces tight bounding boxes around each individual. Even in moderately cluttered scenes, false positives are rare, and most visible persons are correctly detected.

Supervision and Danger-Zone Detection

On top of the raw detections, the geometric reasoning module was evaluated on annotated video sequences containing scenarios such as:

- children playing under teacher supervision;
- a child temporarily moving away from the group;
- a child approaching a predefined danger zone (e.g., near a door or stairs).

For frame-level classification of **supervised vs. unsupervised** status and **child in danger zone**, the system achieved high agreement with manually labeled ground truth. In particular, unsupervised intervals longer than a few consecutive frames were reliably detected, and most entries into configured danger zones were correctly flagged. Short, isolated misdetections were effectively suppressed by temporal smoothing (requiring several consecutive frames before raising an alarm).

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 09 Issue 05 May 2026
DOI: 10.47191/ijcsrr/V9-i5-32, Impact Factor: 8.048
IJCSRR @ 2026
www.ijcsrr.org

V. CONCLUSION

In this work, we presented an AI-assisted monitoring system for kindergarten safety that combines a YOLO-11–based object detector with a simple but effective spatial reasoning module. The system is designed to automatically distinguish between **children** and **adults** in surveillance video streams and to infer higher-level safety conditions such as **supervised vs. unsupervised child** and **child presence in predefined danger zones** (e.g., doors, stairs, exits). A custom, domain-specific dataset of kindergarten-like scenes was created and annotated with "child" and "adult" bounding boxes, forming the basis for specialized training and evaluation. Experimental results on the test set demonstrate that the trained YOLO-11 model achieves **high detection performance** for both target classes, with strong precision, recall, F1-score, and mAP@0.5 (see Table 3), while at the same time sustaining **real-time inference speeds** on GPU hardware. Qualitative analysis on real video sequences shows that the system can reliably flag frames and short intervals where children remain unsupervised or move into danger zones, thereby providing timely cues to caregivers and supporting continuous situational awareness in the kindergarten environment.

At the same time, several limitations were identified. Detection performance degrades for very small or heavily occluded children, and the current implementation does not yet include explicit multi-object tracking or advanced temporal modeling. The behavior analysis layer is intentionally simple, focusing on geometric relationships rather than full action understanding. Furthermore, real-world deployment raises additional requirements in terms of data privacy, ethical use, and integration with existing organizational workflows.

Despite these challenges, the results obtained so far confirm that YOLO-family detectors—and YOLO-11 in particular—constitute a **solid foundation** for intelligent kindergarten monitoring. As future work, we plan to (i) enrich and diversify the dataset with more real kindergarten footage, (ii) integrate tracking and pose estimation to enable **fall detection**, **aggression and social conflict analysis**, and **social isolation monitoring**, (iii) explore lightweight facial or body-expression models for **affective state estimation**, and (iv) develop a user-friendly software interface tailored to teachers and psychologists. In this way, the proposed system can evolve into a comprehensive, AI-based decision-support tool that enhances child safety and well-being in early childhood education settings.

SOME OF THE ADVANAGES FROM THE ABOVE RESULTS

a) Real-time detection of children and adults in surveillance video with high precision and recall.
b) Efficient YOLO-11 architecture that maintains low latency while processing continuous video streams.
c) Reliable identification of **supervised** and **unsupervised** children using simple geometric reasoning on top of detection outputs.
d) Accurate detection of children entering predefined **danger zones** (doors, stairs, exits), enabling timely safety alerts.
e) Robust performance in cluttered kindergarten-like environments with multiple children and adults present.
f) Good generalization to varying viewpoints, illumination conditions, and room layouts thanks to the custom domain-specific dataset and data augmentation.
g) Strong balance between **speed and accuracy**, making the proposed system suitable for practical deployment in real kindergarten settings and scalable to future tasks (fall detection, aggression, social isolation, emotion analysis).

REFERENCES

1. Wang, X., "The Research and Analysis of Different Face Recognition Algorithms," Journal of Physics: Conference Series, WLSA Shanghai Academy, Shanghai, China, 2022.
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), 2016, pp. 779–788.
3. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection With Region Proposal Networks," in Advances in Neural Information Processing Systems (NeurIPS), 2015, pp. 91–99.
4. Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object Detection With Deep Learning: A Review," IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 11, pp. 3212–3232, 2019.
5. N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-End Object Detection With Transformers," in Computer Vision – ECCV 2020, pp. 213–229, 2020.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 09 Issue 05 May 2026
DOI: 10.47191/ijcsrr/V9-i5-32, Impact Factor: 8.048
IJCSRR @ 2026
www.ijcsrr.org

6. T. Shehzadi, Q. Yang, M. Alazab, et al., "Object Detection With Transformers: A Review," Sensors, vol. 25, no. 19, p. 6025, 2025.
7. R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *Proc. IEEE CVPR*, 2014.
8. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. IEEE CVPR*, 2016.
9. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *Proc. IEEE CVPR*, 2005.
10. Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.
11. LI LiLi, ZHANG YanXia and ZHAO YongHeng, "K Nearest Neighbors for automated classification of celestial objects," Science in China Series G-Phys Mech Astron, Vol.51, no.7, July 2008, pp.
12. 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*, vol. 39, no. 6, pp. 1137–1149, 2017.
13. Saydazimov, J., S. Ergashev, and A. Nosirkulov. *Research of Some Image Filter Algorithms Used in Object Detection.* Proceedings of the 8th International Conference on Future Networks & Distributed Systems, 781–785. (2024).
14. Saydazimov, J., S. Turaqulov, and J. Toshpo'latov. *Image Enhancement Methods and Algorithms for Object Recognition Using Artificial Intelligence.* Digital Transformation and Artificial Intelligence 3 (3): 42–46. (2025)
15. A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.
16. Leila Zoubida, Réda Adjoudj "Integrating Face and the Both Irises for Personal Authentication". I.J. Intelligent Systems and Applications, 2017, 3, 8-17
17. J. Terven and D. Cordova-Esparza, "A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS," *Machine Learning and Knowledge Extraction*, vol. 5, no. 4, pp. 1620–1659, 2023.
18. F. Feng, Y. Hu, W. Li, and F. Yang, "Improved YOLOv8 Algorithms for Small Object Detection in Aerial Imagery," *Journal of King Saud University – Computer and Information Sciences*, vol. 36, no. 18, art. 102113, 2024.
19. A. S. Aldubaikhi, H. Shin, and H. B. Mahamadu, "Advancements in Small-Object Detection (2023–2025): Taxonomy, Analysis and Future Directions," *Applied Intelligence*, 2025 (online first).

Cite this Article: Karimovich, S.J., Abdumurodovich, B.U., Yorkinovna, M.S. (2026). Real-Time Monitoring of Kindergarten Safety Using YOLO-11-Based Detection of Children and Adults. International Journal of Current Science Research and Review, 9(5), pp. 2593-2600. DOI: https://doi.org/10.47191/ijcsrr/V9-i5-32