A Computer Vision Based Child Safety Solution Using YOLOv8 Architecture

Muhammad Haroon Ahmad[1], Umair Ud Din[1], Fawad Hussain[1], Muhammad Farooq[1], Abid Khan[1], Ihsan Ul Haq[2]
[1]Department of Electrical Engineering, University of Engineering and Technology, Peshawar, Pakistan
[2]Energy Engineering, University of Hull, United Kingdom
*Correspondence: 21pwele5850@uetpeshawar.edu.pk, muhammad.farooq@uetpeshawar.edu.pk, ihsankhan97@gmail.com

Citation | Ahmad. M. H, Din. U. U, Hussain. F, Farooq. M, Khan. A, Haq. I. U, "A Computer Vision Based Child Safety Solution Using YOLOv8 Architecture", IJIST, Vol. 07 Special Issue. pp 297-306, May 2025.
Received| April 14, 2025 Revised| May 15, 2025 Accepted| May 18 2025 Published| May 20, 2025.

Child safety continues to be a major concern in homes, public spaces, and schools. Physical barriers and supervision by parents or guardians are often not enough to prevent accidents in restricted or high-risk areas such as swimming pools, staircases near sharp objects, electrical sockets or places where drugs are stored. This project proposes a real-time computer vision-based solution to enhance child safety by detecting the presence of children in restricted zones and alerting guardians, caregivers or authorities immediately. The system is built using YOLOv8 (You Only LOOK Once version 8) for object detection, combined with distance estimation and an alarm-triggering mechanism. A custom dataset containing over 30,000 labeled images across eight categories was used for model training and validation. The euclidean distance formula was applied to measure the spatial relationship between the detected children and nearby hazards, enabling accurate risk assessment in real-time. The proposed model achieved a mean Average Precision (mAP) of 90% and showed high accuracy in detecting critical proximity scenarios instantly. The solution is scalable and deployed in various environments, offering a proactive approach to preventing accidents. This project aims to deliver and effective system using readily available hardware, making it easy to install in both private and public spaces. Early testing demonstrated high levels of accuracy, speed, and real-time performance, positioning this system as a potential breakthrough in child safety technology.

Keywords: Computer Vision, YOLOv8, Euclidean Distance, Ultralytics, Annotation, Confusion Matrix.

Introduction:

Children's safety has become a major concern for parents, educators, and communities in today's fast-paced and often unpredictable world. Traditional methods of ensuring child safety are increasingly challenged by the complexities of modern environments, where many domestic accidents involve children coming into contact with dangerous objects such as electrical sockets, stairs, or sharp items. Conventional child safety practices mainly rely on passive protections, like installing barriers, which are not always effective. Since the launch of YOLOv1 in 2015, the YOLO series of algorithms has gained significant attention due to its innovative approach of treating object detection as an end-to-end regression problem. This method differs sharply from two-stage detection approaches, simplifying the process and noticeably increasing detection speed, making YOLO ideal for real-time detection applications such as video monitoring and autonomous driving [1].

Thus, we selected the YOLO architecture for our work. In addition to object detection, we integrated a distance estimation algorithm to calculate the spatial relationship between a child and nearby hazards. The system uses Euclidean distance calculations between the centers of detected bounding boxes, a widely used method in computer vision for estimating object proximity [2]. This enables real-time identification of potential threats based on predefined safety thresholds. Distance estimation techniques have been extensively studied in vision-based monitoring systems, proving their effectiveness in fields such as autonomous navigation and safety monitoring [3].

The objective of this research is to develop a computer vision based real-time solution that continuously monitors the distance between the child and surrounding hazards triggering an alarm when a critical proximity threshold is crossed to help prevent accident [4]. Our system is designed to activate an alert mechanism whenever the calculated distance falls below the set limit providing timely warnings to caregivers and enhancing child safety.

Literature review:

In today's rapidly-evolving world, child monitoring has become a significant concern for parents, making it an increasingly area of research. Traditional methods such as baby-proofing, hiring caregivers for constant supervision, and installing physical barriers have often proven to be insufficient. However, with the rapid advancements in technology—particularly in artificial intelligence and computer vision-there is growing potential to develop intelligent systems that significantly enhance child safety. Several studies have explored object detection models. Notably, YOLO (You Only Look Once), introduced by Redmon and Farhadi [5], is a real-time object detection algorithm known for achieving high accuracy. Hybrid object detection approaches have also been investigated to improve both speed and accuracy. For example, in [6], a combined YOLOv5 and Faster R-CNN model was proposed for vehicle detection and traffic density estimation, achieving 90% accuracy on a dataset of 40,000 images. The study revealed that YOLOv5 performed better in low-object-density scenarios, while Faster R-CNN was more effective in high-density environments. In terms of processing speed, YOLOv5 required only 7.5 seconds per prediction compared to 18.5 seconds for Faster R-CNN. These findings support the selection of YOLO-based architectures for real-time applications due to their superior efficiency.

YOLOv8 [7], the latest version, offers further improved performance and is particularly well-suited for child safety applications due to its ability to detect multiple objects in dynamic environments. Distance estimation is another critical component of hazard detection. Many researchers have employed Euclidean distance calculations to measure the proximity between detected objects. During the COVID-19 pandemic, several studies demonstrated the effectiveness of combining YOLO models with distance estimation to monitor social distancing [8]. Similar techniques have been applied in vehicle safety systems to estimate distances between vehicles on highways, helping to prevent collisions [9]. Inspired by these approaches, our system aims to minimize risk by integrating YOLOv8 with real-time distance measurement.

Furthermore, real-time alert mechanisms have been widely implemented across various domains, including surveillance, vehicle detection, and healthcare. The integration of computer vision with alarm systems as suggested in [10], ensures immediate user alerts in the presence of potential dangers.

This review highlights the importance of combining object detection, distance estimation, and alert generation into a unified framework for enhancing child safety. Our research builds upon these advancements by leveraging YOLOv8, proximity measurements, and real-time alert systems to develop a scalable and robust child safety solution. In addition, existing studies on child monitoring have proposed using robots [11], wearable sensors [12], and smart cameras [13]. However, most of these solutions fall short in delivering real-time hazard detection and accurate distance estimation. Our approach addresses these limitations by integrating state-of-the-art YOLOv8 object detection with real-time spatial analysis offering a more effective solution for child protection.

Methodology:

Dataset preparation*:*

To develop an accurate child safety detection model, this dataset combines custom images from different real-world environment such as rooms, staircases and outdoor areas along with sources like Kaggle and Roboflow. It includes various lightning conditions, different object angles and various background types. This enhances the model's ability to generalize across real-world situations. The dataset consisting of over 32,000 images for training and more than 5,000 images for validation across 8 classes. The annotation process was carried out using the CVAT (Computer Vision Annotation Tool) to ensure the precise labeling of babies and hazardous objects including as pools, broken glass, knives, stairs, sockets, drugs, and reptiles.

Table 1. Contains class name, class id and total dataset

No	Class name	Class ID	Dataset size
1	Baby	0	7000 images + annotations
2	Pool	1	5000 images + annotations
3	Stairs	2	5000 images + annotations
4	knife	3	4000 images + annotations
5	Broken glass	4	5000 images + annotations
6	Electrical socket	5	5000 images + annotations
7	drugs	6	4000 images + annotations
8	reptiles	7	4000 images + annotations

Table 1 provides an overview of the dataset. It consists of annotated images covering eight classes relevant to child safety scenarios. The "Baby" class includes 7000 annotated images, while hazardous objects, such as pools, stairs, broken glass, and electrical sockets, each have 5,000 images. Additionally, the classes for knives, drugs, and reptiles are represented with 4,000 images each.

Directory structure:

Figure 1, shows the directory structure in YOLO format There are two main folders i.e images and labels. Inside the images folder, there are two sub folders: train and val. The train folder contains 80% of the total dataset, for training while the val folder holds 20% of the dataset for validation. Similarly the labels folder mirrors this structure and contains the annotations for all the images.

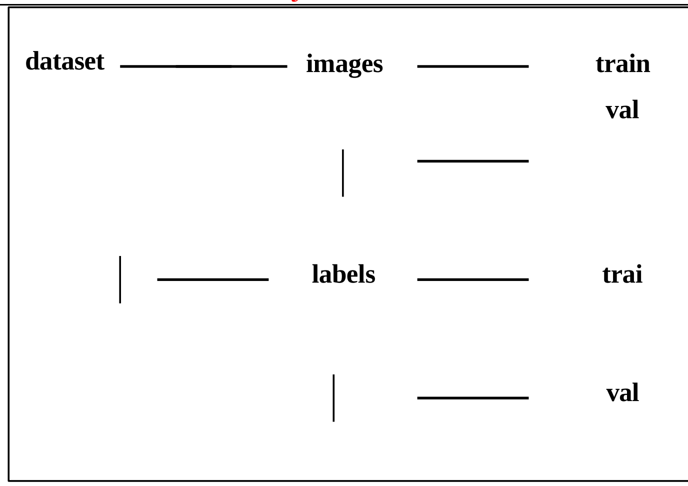

Figure 1. Directory structure to train the model

Model training:

After preparing the dataset we trained a YOLOv8 large model on Google Colab using an Nvidia A100 GPU with 70 compute units. We used YOLOv8 large model for its optimal trade-off between accuracy and real-time capabilities. While smaller version offers faster interference, our primary objective was reliable detection of small objects like knife, broken glass, drugs and socket, which requires better feature representation. The Command showen in Figure 2, was used to train the mode.

```
task=detect mode=train  model=yolov8l.pt
data=dataset.yaml  epochs=100  imgsz=640
batch=16
```

Figure 2. Code use to train the model

After training the model we used the best.pt file (which corresponds to the model that achieved the best evaluation metric) to test the model's accuracy on random images and videos.

Distance estimation:

We loaded YOLOv8 model using Ultralytics [14] and run inference on an image to detect objects. The model identifies bounding boxes for detected objects, extracting coordinates (x_1, y_1, x_2, y_2). The center of each detected object is computed using equation (1).

$$X_{center} = \frac{x_1+x_2}{2},\ Y_{center} = \frac{y_1+y_2}{2} \qquad (1)$$

This determines the exact location of baby and the hazard.)

The Euclidean distance between the baby and a hazard is computed as:

$$\sqrt{(x_h - x_b)^2 + (y_h - y_b)^2} \qquad (2)$$

In Equation (2), $(x_b\ , y_b)$ represents the center of the baby's bounding box, and $(x_h\ , y_h)$ represents the center of the hazard's bounding box. Therefore Equation (2) provides an accurate measure of the proximity between the child and the hazardous object.

Threshold based safety:

A critical distance threshold is defined, if the distance is less than the critical threshold, the system displays a warning message (WARNING: Baby too close to hazard!) "Otherwise", a "Safe" status is shown. The system annotates the image with bounding boxes, using a blue box for the baby and green boxes for hazards. The processed image is then displayed using Matplotlib [15].

Alert system:

When the system detects a hazardous situation, it immediately triggers an alert using a GSM module (e.g., SIM800L). A predefined warning message is sent via SMS to the caregiver's phone, notifying them of the potential danger. At the same time, a buzzer sounds to provide an immediate on-site warning. This ensures that even if the caregiver is not physically present, they are instantly informed and can take the necessary action to protect the child.

Phase flow chart:

Figure 3, shows the five phases of our framework, which include data collection, preprocessing, classification and detection, distance estimation and evaluation.

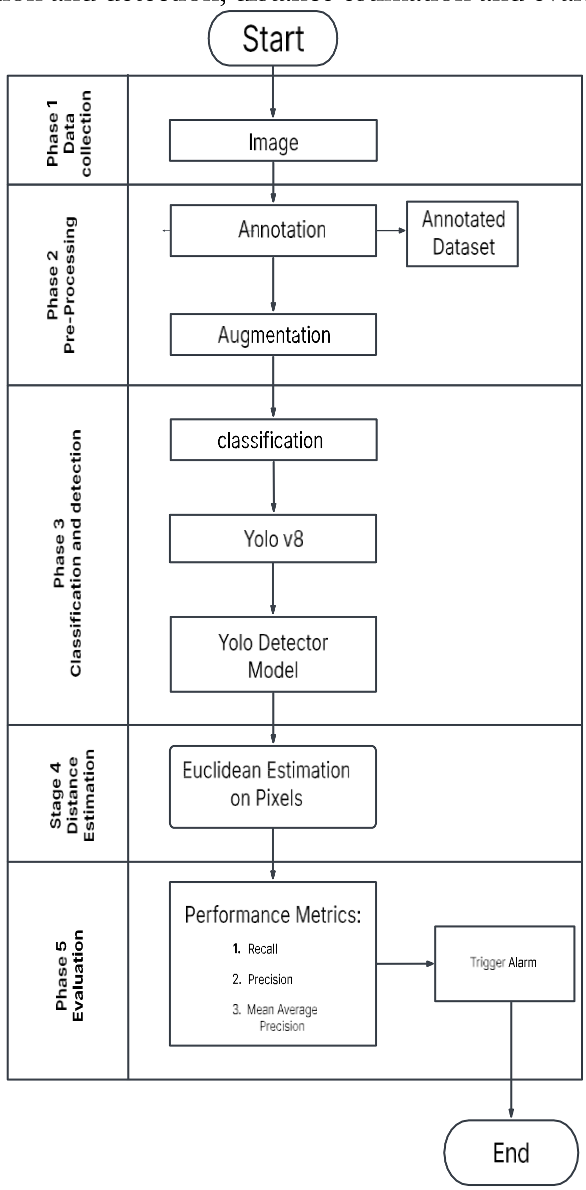

Figure 3. Project framework

Results:

Figure 4, shows the confusion matrix of the model. The model demonstrates strong performance in detecting most classes with highest accuracy observed for the "Pool" class, where predictions were 100% correct. Other classes also exhibit high correct classification rates including "Broken glass" (94%), "Drug" (93%), "Socket (87%), and "Stairs" (85%). The "baby" was correctly identified in 86% of the cases but also shows some misclassification on the background and with minor confusion into “Reptile” (1%). Slight confusion was noted between certain classes. For example, “knife” was correctly classified 84% of the time, but was misclassified as “Reptile” and “background” in 16% of cases. Class “Reptile” has also slightly lower true positive rate of 81%. Despite these challenges the matrix overall illustrates strong multi-class detection performance and confirms the model reliability in distinguishing diverse object in complex scenes

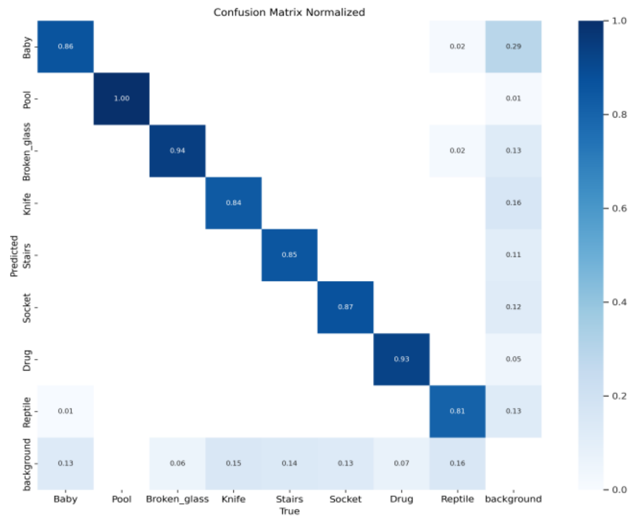

Figure 4. Confusion matrix of model

Figure 5, shows the box loss, classification loss and DFL loss all of which are effectively decreasing indicating successful learning, it also presents precision and recall metrics both of which show an upward trend meaning the model is improving in correctly identifying objects. Additionally, the mAP50 and mAP50-95 metrics are also improving further indicating enhanced detection performance. From Figure 5, the training box loss decreased from initial value of 3.0 to 0.6 approximately, while the classification loss dropped for 4.0 to 1.0. The dfl loss also declined from 3.6 to around 1.1 over 50 epochs. Similar trends were reflected in the validation losses, indicating that the model generalized well to unseen data and showed no sign of overfitting. In parallel with the loss reduction, performance metrics steadily improved. The precision improved from 0.2 to approximately 0.89, and the recall rose from 0.35 to around 0.86. These quantitative confirms the model’s stable convergence and its growing capability to accurately detect and classify across training iterations.

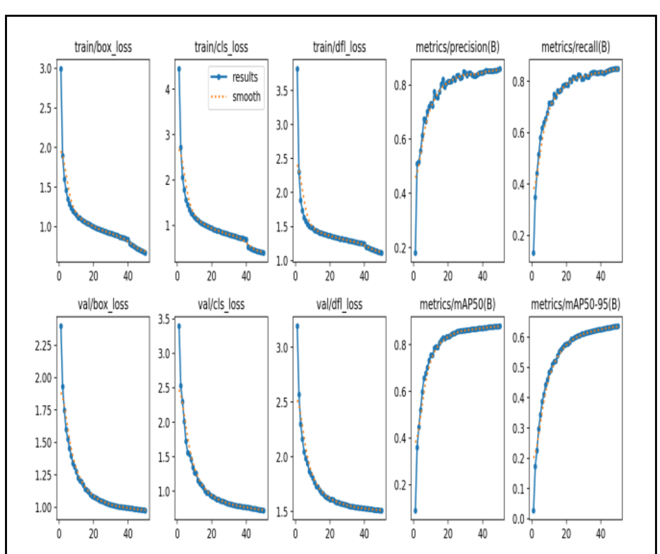

Figure 5. Different curves of the model

Figure 6, shows the model's detection, where it identifies the baby and various hazards. The bounding boxes are colour-coded as follows blue for the baby, light green for the knife, purple for the stairs, white for broken glass, pink for the socket and red for the drugs. The model detects the baby and hazards with great accuracy.

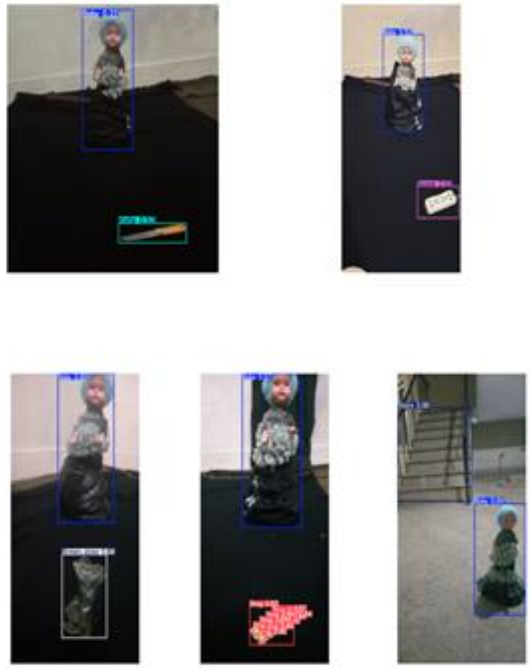

Figure 6. Detection of baby and single hazard

Figure 7, shows that the model performs well even with multiple objects. In the figure, the baby is near several hazards, such as socket and knife, socket and broken glass, and knife and drugs. The model accurately detects both the baby and the hazards in each scenario.

Figure 7. Detection of baby and multiple hazards

Figure 8, presents two scenarios: one where the baby is far from the hazard (knife and socket) and one where the baby is near the hazard. The model calculates the Euclidean distance between the baby and the hazard. When the baby is far from the hazard and, the critical distance exceeds the threshold of 700 pixels, the system generates a "Safe" message in green indicating that the baby is safe. If the critical distance is less than the threshold, the system generates a "WARNING: Baby too close to knife"" or WARNING: Baby too close to socket" message in red indicating potential danger.

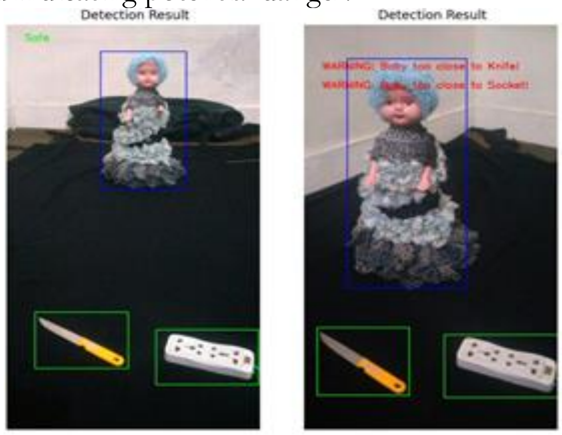

Figure 8. Baby near and far from hazard

Figure 9, shows baby is near different hazard such as broken glass, socket, stairs and the warning message is generated. The results show the model is trained very fine and it detects when the baby is close to hazard with accuracy and at same time generates the alert.

Discussion:

To evaluate our model effectiveness, we compared our results with existing object detection system used in related child safety and hazard detection studies. Traditional approaches often relied on wearable sensors [12] or child monitoring robot [11], which lacked dynamic detection and real time alerting. In contract our system leverages YOLOv8 architecture to perform accurate, multiclass object detection in real time, with recorded mean average precision (mAP) of 90%. This significantly exceeds the performance of earlier YOLOv5-based systems used in COVID distancing studies [7], where typical mAP ranged between 70%-80%. Furthermore, our Euclidean distance estimation for dynamic alert triggering introduces a novel layer of safety which absent in previous works. The use of diverse dataset and real-world hazard classes ensures greater generalization capabilities, making our approach more adaptable and reliable in uncontrolled environment.

Figure 9. Baby near different hazards

Conclusion:

The project aims to enhance child safety in situations where parental supervision may not be available. A system was developed to detect potential hazards and accurately identify the presence of children, allowing for timely alerts to mitigate risks. The results show that the model achieves high accuracy in recognizing both children and hazardous objects, establishing it a reliable tool for safety monitoring in various environments. This work is particularly relevant to home security and public spaces, where child safety remains a critical concern. By providing a proactive approach to monitoring, the system has the potential to significantly reduce the risk of accidents and support caregivers in maintaining a safer environment. Despite these promising results, certain limitations remain such as occasional false positives and sensitivity to changes in lighting conditions. Future research should focus on improving detection robustness and enhancing real-time processing efficiency. In conclusion, this study highlights the valuable role that artificial intelligence can play in advancing child safety solutions in both private and public settings.

Acknowledgement:

Our work is supported by Department of Electrical Engineering, University of Engineering and Technology Peshawar. We also want to thank Professors of Electrical Department for the Mentorship.

References:

[1] Liang Junbiao, "A review of the development of YOLO object detection algorithm," *Appl. Comput. Eng.*, vol. 71, no. 1, pp. 39–46, 2024, doi: 10.54254/2755-2721/71/20241642.

[2] A. F. Joseph Redmon, "YOLOv3: An Incremental Improvement," *arXiv:1804.02767*, 2018, doi: https://doi.org/10.48550/arXiv.1804.02767.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-December, pp. 770–778, Dec. 2016, doi: 10.1109/CVPR.2016.90.

[4] S. Tripathi, A. Sharma, and G. Krishnan, "Child Monitoring and Data Analysis via Computer Vision," *Proc. 6th Int. Conf. Smart Syst. Inven. Technol. (ICSSIT), Kalpa Publ. Comput.*, vol. 19, pp. 455–442, Aug. 2024, doi: 10.29007/MJXM.

[5] A. F. Joseph Redmon, Santosh Divvala, Ross Girshick, "You Only Look Once: Unified, Real-Time Object Detection," *arXiv1506.02640 [cs*, 2016, doi: https://doi.org/10.48550/arXiv.1506.02640.

[6] U. Mittal, P. Chawla, and R. Tiwari, "EnsembleNet: a hybrid approach for vehicle detection and estimation of traffic density based on faster R-CNN and YOLO models," *Neural Comput. Appl.*, vol. 35, no. 6, pp. 4755–4774, Feb. 2023, doi: 10.1007/S00521-022-07940-9/METRICS.

[7] S. M. Om M. Khare, Shubham Gandhi, Aditya M. Rahalkar, "YOLOv8-Based Visual Detection of Road Hazards: Potholes, Sewer Covers, and Manholes," *arXiv:2311.00073*, 2023, doi: https://doi.org/10.48550/arXiv.2311.00073.

[8] M. Ş. Gündüz and G. Işık, "A new YOLO-based method for social distancing from real-time videos," *Neural Comput. Appl.*, vol. 35, no. 21, pp. 15261–15271, Jul. 2023, doi: 10.1007/S00521-023-08556-3/METRICS.

[9] F. V. and H. M. Vítor Freitas, Richardson Menezes, "Estimation of Safety Distance Between Vehicles on Highways Using YOLOv4 from Aerial Images," *Proc. Brazilian Symp. Intell. Comput.*, 2021, [Online]. Available: https://sbic.org.br/eventos/cbic_2021/cbic2021-148/

[10] S. Kumar, S. Swetha, V. T. Kiran, and P. Johri, "IoT based smart home surveillance and automation," *2018 Int. Conf. Comput. Power Commun. Technol. GUCON 2018*, pp. 786–790, Mar. 2019, doi: 10.1109/GUCON.2018.8674999.

[11] J. A. C. Jose *et al.*, "A Vision-Based Detection and Tracking Algorithm for a Child Monitoring Robot," *2019 4th Asia-Pacific Conf. Intell. Robot Syst. ACIRS 2019*, pp. 169–173, Jul. 2019, doi: 10.1109/ACIRS.2019.8936015.
[12] S. C. Ofir Arbili, Lior Rokach, "Wearable Sensors for Ensuring Sports Safety in Children with Autism Spectrum Disorder: A Comprehensive Review," *Sensors*, vol. 25, no. 5, p. 1409, 2025, doi: https://doi.org/10.3390/s25051409.
[13] H. A. Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *arXiv:1704.04861*, 2017, doi: https://doi.org/10.48550/arXiv.1704.04861.
[14] Ultralytics, "Ultralytics YOLOv8 Documentation," *Ultralytics*, 2025, [Online]. Available: https://docs.ultralytics.com/
[15] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 90–95, 2007, doi: 10.1109/MCSE.2007.55.