Fall Detection in Elderly People at Home Using the YOLOv11-Pose Model

Qaseem Riyadh Khawam and Muhaned Al-Hashimi
Department of Cybersecurity, College of Computer Science and Mathematics, Tikrit University, 34001 Tikrit, Salah al-Deen, Iraq
qassem.r.khawam00429@st.tu.edu.iq , muhaned@tu.edu.iq

Keywords: Fall Detection, YOLO, Deep Learning, Pose Estimations, Real-Time Detection.

Abstract: Falls are a leading cause of serious harm and unintentional death among the elderly, especially in homes where direct monitoring or rapid assistance is unavailable. Most of the models based on computer vision developed to determine falls lack decent performance in the real world, e.g., they cannot work properly in the case of views lighting conditions. Such deficiencies result in the realization of detections or false alarms particularly in residences. The study will focus on promoting the accuracy of fall detection among older adults, by proposing a strong real-time system capability, where object detection is combined with human pose estimation. In this study, we propose an automated fall detection system using computer vision techniques, specifically the YOLOv11-pose model. A phone camera was used to capture high-resolution image data, which was then properly processed, labelled, and optimized before training multiple versions of YOLO11-pose using standardized hyperparameter settings to ensure fair comparison. Experimental results showed that all models performed well in fall detection and pose estimation, with smaller models ideal for rapid implementation, achieving high frames per second (FPS) of around 95 fps and larger models, such as YOLOv11m, achieved an accurate mAP@50 of 99.43%. This study determined that YOLOv11-Pose is an effective model for detecting falls in the elderly, with promising accuracy and speed. Future directions include improving performance in challenging imaging conditions, expanding the model to include more activities.

1 INTRODUCTION

As people reach the age of 60, their health and well-being become a top priority, as older adults face numerous challenges, including the risk of falls, which can have serious consequences. The World Health Organization predicts that by 2050, the number of people over 60 will double, making the situation even more critical [1]. There is an urgent need to address the issue of spotting human falls using cutting-edge technology to ensure that the older population is not left alone and receives prompt attention following a fall [2], [3]. Therefore, preventative techniques for residential settings are of fundamental importance [4]. A fall can cause significant injuries or health consequences, necessitating immediate attention. To address this issue, several strategies and procedures have been developed to detect falls early and give rapid aid [5], [6]. Fall detection systems are now grouped into three categories [7]: fall detection based on sensors deployed in ambient situations, wearable sensor devices, and computer vision.

To detect falls, wearable sensors equipped with accelerators, gyroscopes, and magnetic needles are worn on the elderly's waist, limbs, chest, or back. This technique is simple to implement and has a high detection rate. However, the gadget must be worn constantly, which may disrupt everyday activities. If the elderly do not wear it, the gadget cannot detect their condition in time and has to be charged, making it less convenient [8], [9]. The sensor-based strategy involves installing monitoring equipment in the geriatric activity area and collecting data on pressure, vibration, and sound to detect falls. This method's detection area is limited, and sensors are susceptible to interference from external influences, resulting in poor detection accuracy [10], [11].

Computer vision-based fall detection involves processing captured footage to determine whether there is a fall behaviour. This method has gained popularity in fall detection research due to its features, including a fixed camera with continuous power supply for real-time monitoring, the absence of wearable devices to prevent interference, and high detection accuracy [12], [13]. Using the YOLO framework for object identification and position estimation, this study proposes a deep learning-based strategy for detecting falls in the elderly at home. To determine the causes of falls, real-life photographs of falls and natural activities were captured. The dataset was utilized to train the YOLO model, whose performance was assessed using cutting-edge computer vision techniques. This research seeks to increase the accuracy of real-time fall detection and create scalable solutions. The contributions of this study are as follows:

1) Create a realistic dataset with numerous falls and daily activities.
2) Use the YOLOv11-pose model to analyze body postures and identify falls.
3) Evaluate the model's ability to identify falls from other comparable behaviours.
4) Present a concept for a smart alert system based on this model that would provide instant notifications when a fall happens.

This paper explores efforts to ensure the safety of the elderly in their homes through posture analysis and real-time fall detection. Given the paucity of studies on the use of this method, which excels in combining object detection and pose estimation, and the promising results of recent object detection studies using the YOLOv11 pose detection model, the remainder of this paper is divided into: Section 2 discusses related work; Section 3 presents the YOLOv11 pose perspective; Section 4 describes the methodology; Section 5 presents experimental results and discussions; and Section 6 concludes with future research directions.

2 REVIEW OF RELATED LITERATURE

Human fall detection systems has recently gained attention in the healthcare field due to its crucial role in preventing injuries and improving the quality of life for the elderly population. It is used to identify an old person who has tripped, stumbled, collapsed, crashed, kneeled, or is discovered positioned on the lower surface [14]. Developing sophisticated fall detection systems for the elderly can benefit them in a variety of ways. These include maintaining a regular lifestyle, following routines autonomously, lowering healthcare costs, and avoiding isolation. Overall, it can improve both physical and mental health, leading to a higher quality of life [1].Traditional fall detection technologies, such as wearable sensors or motion detectors, have limits in accuracy, reliability, and scalability. These methods frequently suffer from high false alarm rates, limited coverage, and discomfort for the user [15].

However, recent technology advancements, notably in artificial intelligence and deep learning, have resulted in automated and intelligent monitoring strategies. Among these strategies, object detection algorithms and pose estimation, notably the YOLO (You Only Look Once) algorithm, have shown to be extremely successful in real-time fall detection [16]. Research indicates that YOLO models provide remarkable speed and accuracy for fall detection and posture predictions, paving the way for early warning systems and real-time monitoring [17].

The (You Only Look Once) family of models has proven its amazing effectiveness in health monitoring tasks, such as analyzing and detecting body postures to detect falls in homes [18]. Due to YOLO's popularity for high-accuracy real-time object identification and its capacity to analyze pictures in a single network pass, the YOLO family of models has recently been increasingly used for human fall detection. For example, Lu and Chu [19] used YOLOv3 detection with Camshift tracking to create an image-based fall detection system for the elderly who struggle to sit or stand. Similarly, Keet al. [20] examined the YOLOv5 variations on the CAUCA Fall dataset and found that YOLOv5x was the most accurate. However, it was slower than YOLOv5s, which had somewhat lower accuracy but had high speed for real-time fall detection. Several research has examined lowering the number of layers and employing ghost convolution techniques to reduce computational burden and complexity in deep networks, such as YOLO. For instance, Kan et al. [21] created CGNS YOLO for fall detection by redesigning YOLO5's neck network with the incorporation of the GSConv and GDCN modules. They also employed a normalization-based attention module (NAM) to improve the method's accuracy. Zheng et al. [22] introduced the FDT-YOLO technique, which enhanced fall detection by modifying the YOLOv8 framework and replacing the C2f module with the Faster R-CNN module to improve feature reuse and minimize computational cost. It also has a deformable convolution module and a triplet attention mechanism, which increases detection accuracy and resilience. Chen et al. [23] upgraded the YOLOv5 architecture by including an improved Efficient Channel Attention (ECA) network. The researchers used average pooling layers in a Spatial Pyramid Pooling (SPP) network to extract features at many scales. The work accurately detects fall occurrences of varying magnitude and subtlety. Khan et al. [24] modified the YOLOV8S model by including a focus module in the backbone to maximize feature extraction and Convolutional Block Attention Modules (CBAMs) at appropriate locations. Similarly, Qin et al. [25] updated the YOLOv8S model by replacing C2f with C2Dv3 modules in the backbone and adding a DyHead block in the neck layer to integrate different attention processes. Despite advances in deep learning models such as YOLOv11-pose, their implementation in real-world settings is hampered by computing costs and data quality constraints. The performance of these models is strongly dependent on availability to high-quality, well-annotated information and appropriate operational circumstances. In the context of fall detection among the elderly, YOLOv11-pose has shown great accuracy and real-time responsiveness, exceeding existing approaches in terms of detection speed and reliability.

However, practical restrictions remain, such as differences in illumination, body shapes, and privacy problems in household settings. The goal of this research is to create an intelligent, cost-effective, and efficient system based on YOLOv11-pose for detecting falls accurately, improving safety, and enabling quick response for older people in residential settings.

3 OVERVIEW OF THE YOLOV11-POSE MODEL

The YOLOv11 pose model is a deep convolutional neural network based on YOLOv11 that employs a unified end-to-end architecture rather than the usual two-branch architecture commonly employed in posture estimation. This single-stage network provides simultaneous detection, pose estimation, and prediction, contributing to its speed and accuracy, allowing pose estimation through combined object and keypoint detection training. This model is built and shows promising results in fall detection and pose estimation. [26]. Three main components are included in the YOLOv11-pose design shown in Figure 1 below: The head, neck, and backbone [27]. After processing the input picture at 640 x 640 pixels, the backbone applies successive convolutional layers (Conv), which increase the amount of feature mappings while decreasing the spatial dimension. Additionally, this backbone has C3K2 blocks, which are specialized modules that use Cross-Stage Partial Connections (CSP) and convolutional procedures to efficiently sample the important parts of an image in order to increase computational efficiency. Multi-scale feature maps produced by the backbone network are further aggregated by the Neck component. Additionally, it includes certain sophisticated feature-fusion layers like Concat (which combines different feature scales) and Upsample (which increases the feature map resolution). YOLOv11 brings notable improvements to the neck section, which includes a C2PSA (Cross-Stage Partial Spatial Attention) module that employs spatial attention to highlight key areas of the feature maps and the SPPF (Spatial Pyramid Pooling Fast) layer to pool the features at various scales [28]. In the last stage, the head of the YOLOv11-position architectural is responsible for creating the final detect outputs, which include object bounding boxes, classification scores, and human pose keypoints. Unlike standard object detection heads, the pose-specific head in YOLOv11 is intended to handle multi-task output in a unified and computationally efficient manner.

To increase pose estimate accuracy, YOLOv11 incorporates keypoint heatmaps into the head module, allowing it to regress exact joint locations with lightweight layers. These outputs are optimized with a composite loss function that incorporates IoU loss (for bounding boxes), binary cross-entropy (for class and confidence), and MSE loss (for keypoints), resulting in balanced learning across all tasks. [29]. In comparison to previous versions like YOLOv7-pose or YOLOv8-pose, YOLOv11-pose adds architectural changes such as the usage of C2f modules, coordinate attention techniques, and better spatial decoupling, which improves localization precision while keeping real-time inference capability [30].

4 METHODOLOGY

This section explains the processes of the proposed fall detection and posture analysis for the elderly utilizing the YOLOv11-pose model. The methodology includes data collection, pre-processing, model training, validation, and evaluation of model performance. The arrangement of these steps in Figure 2 demonstrates the study's conceptual framework, which guarantees the choice of the best model for precise fall detection.

4.1 Dataset Collection

In this study, movies were recorded using the phone's camera at 30 frames per second and 1080p quality in order to gather data for training and assessing a fall detection model utilizing computer vision techniques. In order to guarantee visual quality and discern minute details during a variety of tasks, this resolution was selected.

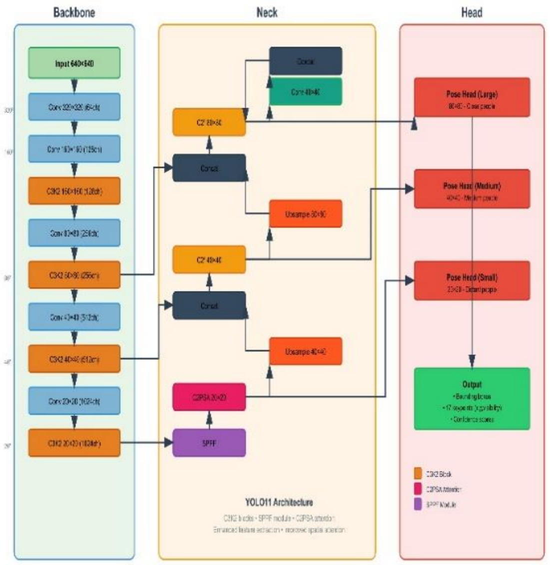

Figure 1: The architecture of YOLOv11- pose model.

Numerous fall scenarios, including forward, side, and backward falls, falls from a chair, and falls while carrying things, are included in the data. These scenarios were all recorded in domestic settings. Walking, standing, sitting on a chair, praying, bending over to tie shoes, ascending and descending stairs, and bending over to pick up articles from the floor were among the other natural activities that were included to symbolize every day, natural gestures.

The videos were shot with people between the ages of 21 and 29. This age group was purposefully chosen owing to the lower frequency of falls within this group compared to older persons, which contributes to their safety during data collecting. Furthermore, choosing this age group allows a diverse range of body compositions and motor abilities while keeping research participants safe. When recording, a range of lighting conditions and viewpoints were taken into account. To make sure the data was complete and addressed all the difficulties the model may encounter in real-life settings, videos were shot in a variety of interior locations and lighting circumstances, including bright and low light. To make sure the model could handle falls from any angle, care was also taken to employ a range of camera angles, including overhead, side, and straight on. The dataset was divided into three major groups based on the type of action conducted in the video, as follows: falling, bending, and normal.

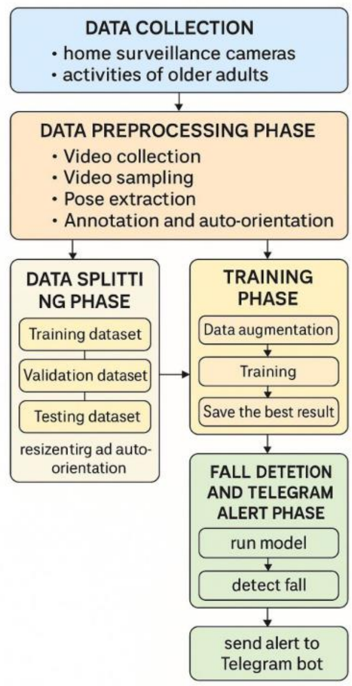

Figure 2: Conceptual framework of the study.

4.2 Dataset Preparation

The fall dataset, which includes categories such as normal motion, bending, and falling, was processed in a methodical manner to prepare it for training the YOLO-Pose model. Videos were taken in a standardized indoor setting with different viewpoints and consistent lighting conditions. Images were manually extracted at 24 frames per second and converted to distinct JPEG images. The photos were reoriented and standardized at 640 by 640 pixels.

The dataset was partitioned into training, validation, and test sets in a 70:20:10 ratio. After augmentation, there were 3784 photos in the training set, 1022 in the validation set, and 745 in the test set, totaling 5551 images. Data augmentation approaches were used to improve data quality and variety, which helped the model perform better, increase its generalization capacity, and reduce the dangers associated with class imbalance. An automated script was used to tag photos and classify them based on their originating folder. The data were translated to the COCO format, which enables preserving the keypoints and tags required to train a human posture identification model.

4.3 Model Training

In this study, the YOLOv11-pose model was chosen for its proven effectiveness and performance in pose estimation and real-time object detection tasks, and all iterations were implemented and trained using an annotated fall detection dataset. To reduce training time and improve detection accuracy, the model was initialized with pre-trained weights from large-scale datasets like COCO through transfer learning, and hyper parameter tuning was carried out methodically, adjusting parameters like the learning rate, batch size, time intervals, and anchor box dimensions to achieve optimal performance. These hyper parameters, listed in Table 1, were carefully selected to create a standardized training environment, allowing for fair comparison between various models.

Table 1: Model hyperparameters.

Training Parameters	Values
Learning Rate	0.001
Batch Size	8
Epochs	100
Image Size	640*640
Momentum	0.937
Weight Decay	0.0005

4.4 Evaluation Metrics

Evaluating the performance of object recognition and pose estimation models, such as YOLOv11-Pose and YOLOv8-Pose, is a crucial aspect of deep learning and computer vision research. This assessment typically relies on several widely accepted metrics, including precision, recall, F1-score, mean average precision (mAP), confusion matrix, and frames per second (FPS) [31], [32]. Precision measures the model's ability to correctly identify relevant targets among all predictions, with high precision indicating a low rate of false positives, which is particularly important in applications where incorrect detections are costly. Recall quantifies how effectively the model detects all relevant instances in the dataset, with high recall signifying that few true positives are missed. The F1-score provides a balanced measure by combining precision and recall into a single metric, offering a comprehensive view of the model's accuracy, especially when both false positives and false negatives have significant consequences. Mean Average Precision (mAP) summarizes the precision-recall relationship across all classes, reflecting overall detection performance and accounting for both accuracy and completeness in multi-class scenarios. Frames Per Second (FPS) represents the number of frames the model can process per second, indicating computational efficiency and suitability for real-time deployment. All these metrics are standard in object detection and pose estimation studies, with the definitions of true positives, true negatives, false positives, and false negatives following widely accepted conventions in the literature [31], [32].

5 RESULTS AND DISCUSSION

All trials in this study were carried out on a personal computer running the Windows 10 operating system. The system specs are shown below: CPU: AMD Ryzen 5 5600.

GPU: Nvidia GeForce GTX 1070 Ti. These hardware resources were used to train the object and fall detection models with the YOLOv11-pose method. The device's computing capabilities enabled efficient model training and assessment.

5.1 Result of the Evaluation Metrics Performance

This section compares the evaluation performance of the YOLOv11-pose detection model to previous versions of the YOLOv8-pose model. Figure 3 compares mAP at 0.5 and 0.95 thresholds for different variations of the YOLOv11-pose and YOLOv8-pose.

YOLOv11m-pose outperforms other models, obtaining 99.43% accuracy at mAP@50 and 92.78% for mAP@50-95. YOLOv11 variations, particularly YOLOv11n-pose, YOLOv11s-pose, and YOLOv11m-pose, excel in detecting junk with high accuracy.

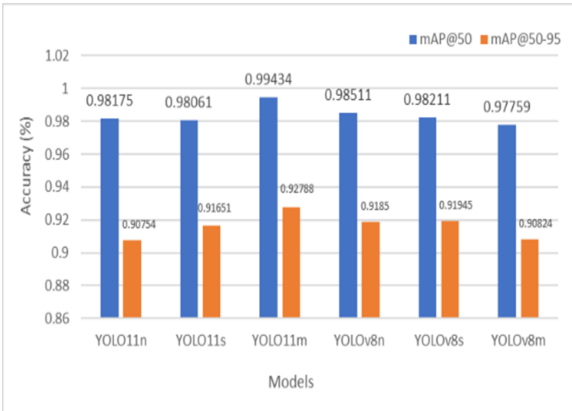

Figure 3: Comparison of mAP@50 and mAP@95 across the YOLO models.

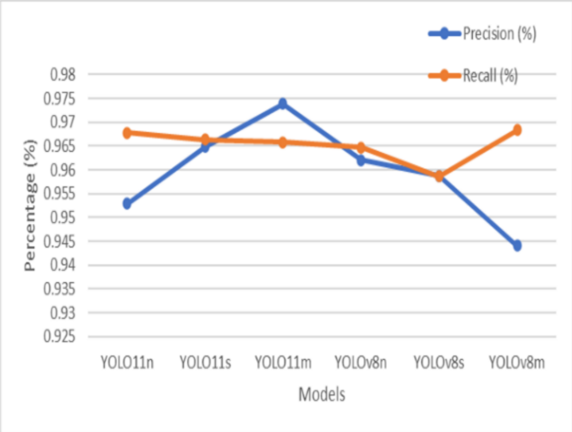

Figure 4: Comparison of precision and Recall of the YOLO models.

Figure 4 compares precision and recall percentages across various YOLO model variants, specifically YOLOv8-pose (n, s, m,) and YOLOv11 (n, s, m,). All models exhibit consistently high accuracy in recognizing true positives, with precision values ranging from around 95% to above 97%. With a precision of over 97%, the YOLOv11m-pose model is the most accurate, indicating that it is better at correctly recognizing junk with fewer false positives. The models' recall percentages differ more pronouncedly, with results ranging from around 96% to 97%. It is noteworthy that the YOLOv11m-pose model has an impressive balance, demonstrating the highest accuracy and very competitive recall performance, indicating its efficacy in acquiring the majority of true positive instances. On the other hand, the YOLOv8s-pose model has the lowest recall, suggesting that it has trouble identifying some objects, maybe as a result of model complexity constraints. It is worth noting that the YOLOv11-pose and YOLOv8-pose models with larger architectures (m version) perform better overall in terms of recall and accuracy, suggesting that detection ability and model complexity are connected. However, the observed variability suggests that, depending on the requirements of a certain application, the ideal YOLO model version should strike a compromise between detection performance, processing resources, and model size.

5.2 Inference Time Performance

Figure 5 compares the frames per second (FPS) achieved by several object detection models, including the YOLOv11-pose and YOLOv8-pose series. Smaller models, such as YOLOv8n-pose and YOLOv11n-pose, produce much faster inference speeds, with YOLOv8n-pose attaining a peak performance of 97 frames per second. Larger models, notably YOLOv8m-pose and YOLOv11m-pose, perform much slower, with YOLOv11m-pose recording the slowest inference rate of only 44 frames per second. The findings show a key trade-off between inference efficiency and model complexity, implying that selecting the optimum YOLO variation necessitates a careful balance of accuracy needs, processing restrictions, and real-time application requirements. Medium-sized models (such as YOLOv11s-pose or YOLOv8s-pose) strike a reasonable balance between speed and accuracy, meeting the balanced needs of developing a detection model.

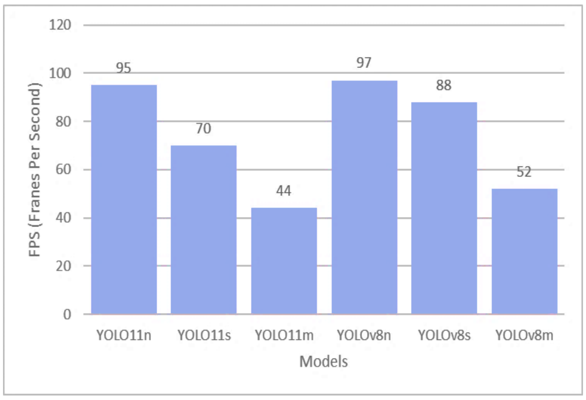

Figure 5: Comparison of Inference speed among the YOLO models.

5.3 Test the Model Fall Detection

5.3.1 Performance Assessment of Fall Detection

The YOLOv11-pose model was evaluated for detecting falls among elderly people in their homes. The model uses computer vision techniques to recognize human body postures in real time through pose estimation algorithms. It exhibits high sensitivity for identifying fall-related patterns by recording motion sequences and recognizing aberrant movement behaviors. A multi-level confidence grading system allows the model to distinguish between true falls and similar actions, such as sitting or bending. The results show that the model performs well across numerous fall situations and remains robust under changing environmental conditions, such as illumination and furniture density (Fig. 6).

5.3.2 The Alert System

The alarm system is a feature intended to instantly inform family members or caretakers when a fall is suspected. For the person who has fallen, this method guarantees quick attention, which might be crucial in averting more harm or difficulties. A Telegram Bot API-based alert system is part of our system. The system notifies a pre-specified chat ID when a fall is detected Figure 7.

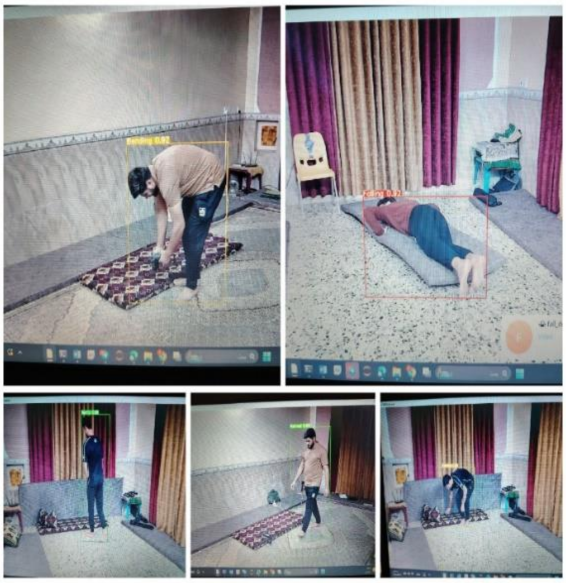

Figure 6: Detection results of Fall Detection for Elderly in Homes.

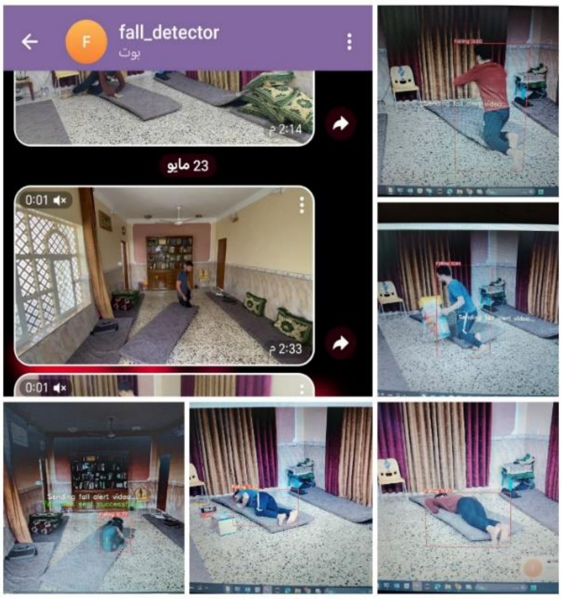

Figure 7: Real-time fall detection.

When a fallout is detected, the system may use the Telegram Bot API to deliver automatic messages to a predetermined conversation ID that include a video of the fall state. This integration includes the following steps:

- Bot Setup. A Telegram bot is built and assigned a unique API key. This token is used to verify and send messages on behalf of the bot;
- Chat ID Configuration. The system assigns a predetermined chat ID to the receiver (a caregiver or family member). This ensures that notifications are sent to the appropriate individual or group.

5.4 Comparative Analysis with the Related Detection Models

This section compares the YOLOv11- Pose based detection model to previous research that employed various versions of YOLO for pose estimation and real-time fall detection. The comparison is based on important factors such as model type, dataset source, detection accuracy (mAP), research restrictions, and how previous work handles or overcomes them. The results are summarized in Table 2. The YOLOv11-Pose model outperforms previous models, with a mAP@50 of 99.43%, indicating improved accuracy and real-time response speed. Zhao et al. [17] utilized YOLOv7-fall improved and attained an accuracy of 95.2% on UR Fall Detection Dataset, Multiple Cameras Fall Dataset. (4016 Image), indicating strong but inferior detection performance compared to YOLOv11-pose. Tîrziu et al. [33] utilized the YOLOv7-W6-Pose model with Le2i data and live videos to identify falls in the elderly with a real-time accuracy of 96.15%, The main problem was differentiating between falls and other actions like extension and bending. Zhongzi Luo et al. [34] built the YOLOv5s-GCC fall detection model using structural improvements that included GhostNetV2, CARAFE, and the CBAM attention module. The model has a mAP@0.5 rate of 93.5% while lowering the number of parameters by 27.5%. Key obstacles included a lack of fall data and the complexity of multi-angle situations, which necessitated increasing the model's generalization and context extraction skills. Improved YOLOv8, as reported by Qin et al. [25], showed only a marginal increase of 5% mAP, and it faced challenges in effectively extracting features objects detection. Using improved YOLOv8, Zheng et al. [22] demonstrated increased performance with a mAP@50 of 96.2% across 5806 images, although they fell short of the excellent findings of YOLOv11.

Table 2: A comparison of our proposed method with other studies.

Study	Model Used	Dataset	Metrics	Result
Zhao et al. [17]	YOLOv7-fall	UR Fall Detection Dataset, Multiple Cameras Fall Dataset. (4016 Image)	mAP	95.0%
Tîrziu et al. [33]	YOLOv7-W6-Pose	Le2i Fall Detection Dataset	F1 Score	95.5%
Zhongze Luo et al. [34]	YOLOv5s-GCC	UR Fall Detection Dataset, Multiple Cameras Fall Dataset. (4303 Image)	mAP	93.5%
Qin et al. [25]	Improved YOLOv8	URFall Detection Dataset, and internet images.	mAP	88.7%
Sarıçayır et al. [5]	YOLOv8 and YOLOv7	Fallen New Version Dataset, Fallsdata2 Dataset. (3290 Image)	mAP	v8: 78%, v7: 81%
Zheng et al. [22]	Improved YOLOv8	Fall Detection Dataset, Le2i Fall Detection Dataset, Multiple Cameras Fall Dataset. (5806 Image)	mAP	96.2%
Proposed model	YOLOv11-Pose	Custom Fall Detection Dataset	mAP	99.4%

6 CONCLUSIONS

The YOLOv11 position model is used to identify falls among elderly individuals in their homes. Initially, a fixed smartphone camera was used to capture videos in interior areas. These clips were then transformed into frames and pre-processed. The dataset was partitioned into training, validation, and test sets. Bounding boxes and keypoint annotations were used to ensure that all data adhered to the YOLOv11 posture. The model was trained on falls, bending, and everyday activities in various settings. After training, the model attained an average accuracy (mAP) of 99.43 using the YOLOv11 posture, and an automated alarm was delivered over Telegram when a fall was detected.

This study determined that YOLOv11-Pose is an effective model for detecting falls in the elderly, with promising accuracy and speed. Future directions include improving performance in challenging imaging conditions, expanding the model to include more activities. The results confirm that YOLOv11-Pose is an effective and efficient model for elderly fall detection. Future work will focus on enhancing its performance under challenging imaging conditions, expanding its activity set for broader behavior monitoring, and integrating it with IoT-based smart home systems for continuous, unobtrusive health surveillance.

REFERENCES

[1] A. R. Khekan, H. S. Aghdasi, and P. Salehpoor, “Fast and high-precision human fall detection using improved YOLOv8 model,” 2024.

[2] A. Bansal, R. Sharma, and M. Kathuria, “A vision-based approach to enhance fall detection with fine-tuned Faster R-CNN,” in Proc. Int. Conf. Advanced Computing & Communication Technologies (ICACCTech), 2023, pp. 678–684.

[3] R. G. Cumming, G. Salkeld, M. Thomas, and G. Szonyi, “Prospective study of the impact of fear of falling on activities of daily living, SF-36 scores, and nursing home admission,” Journals of Gerontology: Series A, vol. 55, no. 5, pp. M299–M305, 2000.

[4] M. Montero-Odasso et al., “New horizons in falls prevention and management for older adults: A global initiative,” Age and Ageing, vol. 50, no. 5, pp. 1499–1507, 2021.

[5] B. Sarıçayır, E. Alican, and C. Özcan, “Deep learning-based human detection for fall injuries,” Computational Technologies in Healthcare, vol. 1, no. 2, pp. 83–92, 2023.

[6] A. R. Khekan, H. S. Aghdasi, and P. Salehpoor, “The impact of YOLO algorithms within fall detection applications: A review,” 2024.

[7] T. Chen, Z. Ding, and B. Li, “Elderly fall detection based on improved YOLOv5s network,” IEEE Access, vol. 10, pp. 91273–91282, 2022.

[8] X. Lv, Z. Gao, C. Yuan, M. Li, and C. Chen, “Hybrid real-time fall detection system based on deep learning and multi-sensor fusion,” in Proc. 6th Int. Conf. Big Data and Information Analytics (BigDIA), 2020, pp. 386–391.

[9] P. V. Er and K. K. Tan, “Wearable solution for robust fall detection,” in Assistive Technology for the Elderly. Elsevier, 2020, pp. 81–105.

[10] L. Ma, M. Liu, N. Wang, L. Wang, Y. Yang, and H. Wang, “Room-level fall detection based on ultra-wideband monostatic radar and convolutional long short-term memory,” Sensors, vol. 20, no. 4, p. 1105, 2020.

[11] K. Wang, G. Zhan, and W. Chen, “A new approach for IoT-based fall detection system using commodity mmWave sensors,” in Proc. 7th Int. Conf. Information Technology: IoT and Smart City, 2019, pp. 197–201.

[12] A. Benkaci, L. Sliman, and H. N. Dellys, “Vision-based human fall detection systems: A review,” Procedia Computer Science, vol. 241, pp. 203–211, 2024.

[13] L. Zhu, Z. Chen, and C. Tian, “Review of fall detection methods based on wearable devices,” Computer Engineering and Applications, vol. 55, no. 18, pp. 8–14, 2019.

[14] L. Ren and Y. Peng, “Research of fall detection and fall prevention technologies: A systematic review,” IEEE Access, vol. 7, pp. 77702–77722, 2019.

[15] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, “YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,” in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), 2023, pp. 7464–7475.

[16] X. Huang et al., “SDES-YOLO: A high-precision and lightweight model for fall detection in complex environments,” Applied Sciences, vol. 15, no. 1, p. 2026, 2025.

[17] D. Zhao, T. Song, J. Gao, D. Li, and Y. Niu, “YOLO-fall: A novel convolutional neural network model for fall detection in open spaces,” IEEE Access, vol. 12, pp. 26137–26149, 2024.

[18] H. Hwang, D. Kim, and H. Kim, “FD-YOLO: A YOLO network optimized for fall detection,” Applied Sciences, vol. 15, no. 1, p. 453, 2025.

[19] K.-L. Lu and E. T.-H. Chu, “An image-based fall detection system for the elderly,” Sensors, vol. 8, no. 10, p. 1995, 2018.

[20] A. Priadana et al., “HFD-YOLO: Improved YOLO network using efficient attention modules for real-time one-stage human fall detection,” 2025.

[21] X. Kan, S. Zhu, Y. Zhang, and C. Qian, “A lightweight human fall detection network,” Sensors, vol. 23, no. 22, p. 9069, 2023.

[22] X. Zheng, J. Cao, C. Wang, and P. Ma, “A high-precision human fall detection model based on FasterNet and deformable convolution,” Electronics, vol. 13, no. 14, p. 2798, 2024.

[23] H. Chen, W. Gu, Q. Zhang, X. Li, and X. Jiang, “Integrating attention mechanism and multi-scale feature extraction for fall detection,” Healthcare, vol. 10, no. 10, 2024.

[24] H. Khan et al., “Visionary vigilance: Optimized YOLOv8 for fallen person detection with large-scale benchmark dataset,” Measurement, vol. 149, p. 105195, 2024.

[25] Y. Qin, W. Miao, and C. Qian, “A high-precision fall detection model based on dynamic convolution in complex scenes,” Electronics, vol. 13, no. 6, p. 1141, 2024.

[26] L. Liu, A. Kurban, Y. Liu, and J. Ma, “Improved YOLOv11pose for posture estimation of Xinjiang Bactrian camels,” International Journal of Advanced Computer Science and Applications, vol. 15, no. 12, 2024.

[27] R. Khanam and M. Hussain, “YOLOv11: An overview of the key architectural enhancements,” 2024.

[28] J. Bento, T. Paixão, and A. B. Alvarez, “Performance evaluation of YOLOv8, YOLOv9, YOLOv10, and YOLOv11 for stamp detection in scanned documents,” Applied Sciences, vol. 15, no. 6, p. 3154, 2025.

[29] Ultralytics, “YOLOv11 pose estimation guide,” May 2024. [Online]. Available: https://docs.ultralytics.com.

[30] L. He, Y. Zhou, L. Liu, and J. Ma, “Research and application of YOLOv11-based object segmentation in intelligent recognition at construction sites,” Buildings, vol. 14, no. 12, p. 3777, 2024.

[31] A. T. Khalaf and S. K. Abdulateef, “Ophthalmic diseases classification based on YOLOv8,” Journal of Robotics and Control, vol. 5, no. 2, pp. 408–415, 2024.

[32] M. T. Abdulhadi, Recognition and Tracking System for Human Action Using Deep Learning, University of Technology, 2023.

[33] E. Tîrziu, A.-M. Vasilevschi, A. Alexandru, and E. Tudora, “Enhanced fall detection using YOLOv7-W6-Pose for real-time elderly monitoring,” Sensors, vol. 16, no. 12, p. 472, 2024.

[34] Z. Luo et al., “Elderly fall detection algorithm based on improved YOLOv5s,” Multimedia Tools and Applications, vol. 53, no. 2, pp. 601–618, 2024.