





# Analyzing the Efficacy of Pose Recognition, YOLOv3, and Deep Learning Techniques for Human Activity Recognition

Ainur Zhumasheva<sup>1,\*</sup>, Madina Mansurova<sup>2</sup>, Gulshat Amirkhanova<sup>3</sup>, Gulnur Tyulepberdinova<sup>4</sup>

<sup>1,2,3,4</sup>*Department of AI and Big Data, Al-Farabi Kazakh National University, Faculty of Information Technology, Almaty, Kazakhstan*

(Received: March 05, 2025; Revised: May 30, 2025; Accepted: September 05, 2025; Available online: October 04, 2025)

## Abstract

The global increase in life expectancy, driven by increased nutrition, healthcare, and living conditions, has resulted in a significant growth in the senior population, notably in Kazakhstan, where the number of people aged 60 and more currently exceeds 2.7 million. This demographic transition poses considerable public health problems, particularly the high prevalence and severity of falls in older persons. Falls are currently the second largest cause of unintentional mortality for more than 87% of the elderly, with 28-34% falling at least once per year. As the worldwide population of people aged 65 and more is predicted to exceed 1.5 billion by 2050, there is an urgent need for precise, real-time fall detection systems. This work uses standardized datasets to conduct a complete evaluation of three fall detection methodologies: posture recognition, YOLOv3-based detection, and deep learning. Deep learning models attained the best accuracy of 92.0% by utilizing their capacity to learn complex spatial-temporal information, but at the cost of increased computing burden and slower inference times (40 ms). YOLOv3 provided competitive accuracy (90.2%) and quicker processing (25 ms), making it suitable for real-time deployment, although with a larger false positive rate. Pose identification, while highly interpretable due to its emphasis on skeletal key points, performed less well in crowded or obscured settings. The findings highlight the possibility for combining the capabilities of each technique to create hybrid systems with adaptive, resource-efficient architectures. Future research should focus on sensor fusion and optimization methodologies to improve accuracy and scalability across a variety of scenarios.

**Keywords:** Fall Detection, Pose Recognition, YOLOv3, Deep Learning, Comparative Study, Human Activity Recognition

## 1. Introduction

The elderly population has been growing very rapidly in recent decades. Nutrient-rich foods, superior health care and higher standards of living have favored the increase in global average life expectancy. Falls in older adults are relatively common and can have dramatic health consequences. Studies have shown that 28-34% of older adults have at least one fall each year [1]. In addition, falls are the second leading cause of accidental death for more than 87% of older adults. In the next 30 years, the number of people 65 and older worldwide will more than double [2]. Approximately 1.5 billion individuals will be 65 and older globally by 2050 [3]. However, because fertility and mortality reductions vary by nature and method, population aging will not be consistent across global regions. The growth of life expectancy in the country and the increase in the population of the Republic of Kazakhstan have led to a noticeable increase in the number of elderly citizens. Thus, at the beginning of 2024, Kazakhstan had more than 2.7 million people aged 60 and over. Over the year, the number of elderly increased by 4.2%, while the average annual growth rate over the last decade was 4.4% [4].

Recent advances in computer vision and deep learning have enabled the creation of non-invasive, camera-based systems capable of analyzing human behaviors in real time [5]. These technologies have the potential to continuously monitor environments without the person having to wear or carry gadgets.

While various studies have investigated fall detection using wearable sensors or vision-based systems, few have conducted systematic comparisons of multiple methodologies using uniform evaluation protocols. Furthermore, existing datasets frequently fail to represent realistic, culturally diverse family contexts, which limits model

---

\*Corresponding author: Ainur Zhumasheva (ainur93ardak@gmail.com)

 DOI: <https://doi.org/10.47738/jads.v6i4.797>

This is an open access article under the CC-BY license (<https://creativecommons.org/licenses/by/4.0/>).

© Authors retain all copyrights

generalizability. This research compares three fall detection approaches: pose-based recognition using skeleton keypoints, YOLOv3-based object detection, and deep learning models like CNN-LSTM architectures.

The aim of this study is to analyse and compare the effectiveness of several computer vision-based fall detection approaches for AAL systems, with a focus on their possible application in Kazakhstan's demographic setting. The research aims to implement and fine-tune three models—pose-based keypoint extraction, YOLOv3 object identification, and CNN-LSTM architectures—and evaluate them using publicly available datasets. These approaches are evaluated in terms of classification accuracy, inference latency, interpretability, and practicality for real-time application. Furthermore, the study aims to examine the trade-offs between model complexity, responsiveness, and transparency, and to offer a hybrid method that combines the benefits of the various techniques.

While there are several ways to fall detection, there has been insufficient comparative evaluation of pose estimation, object recognition, and deep learning-based systems under controlled experimental conditions, particularly in elderly care facilities such as those in Kazakhstan. This work fills a gap by examining these three approaches on standardized datasets, assessing accuracy, computing cost, and practical relevance for real-time AAL deployment.

## 2. Related Work

Fall detection has received a lot of interest from researchers in computer vision, machine learning, and healthcare monitoring. Researchers have suggested many ways for correctly detecting falls, with a primary focus on three unique approaches: posture recognition, YOLOv3-based object identification, and Deep Learning (DL) techniques.

Recent research has investigated several fall detection strategies, including pose recognition, object identification frameworks such as YOLOv3, and other DL approaches. Cao et al. [6] developed a real-time multi-person 2D pose estimation system that uses part affinity fields and skeleton key point extraction to study human posture, offering interpretable insights into movement patterns. This framework has proven useful in applications such as fall detection. However, its effectiveness can be hampered by occlusions and changes in ambient circumstances. While OpenPose was used by Cao et al. [6] to obtain excellent keypoint accuracy, the technique is computationally demanding and unsuitable for real-time processing in AAL scenarios. On the other hand, YOLOv3, a complex object identification system designed for speed and accuracy in real-time applications, was presented by Redmon and Farhadi [7]. It is a viable option for fall detection systems due to its ability to swiftly discover and classify things.

Deep learning approaches, particularly those based on Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), provide a more comprehensive answer. Kwolek and Kepski [8] showed that deep models trained on wearable sensor data may accurately learn complex motion patterns. Kwolek and Kepski's depth-based fall detection offers simplicity but underperforms in low-resolution settings compared to pose-based systems." Added a critical assessment of the trade-offs between approaches in terms of accuracy, speed, and generalizability. Further research [9], expanded these findings to video-based inputs, demonstrating that deep learning architectures can extract high-level spatial-temporal information required for fall detection. However, these models often demand huge annotated datasets and computational resources, making real-time and edge-based deployment difficult in AAL systems.

Bourke and O'Brien [10] conducted a comprehensive literature review and classified fall detection techniques as wearable, ambient, and vision-based systems, highlighting their various advantages and restrictions. Their findings emphasised the necessity of balancing detection accuracy, user comfort, and system feasibility—insights that are still applicable in modern AI-based solutions. This foundational assessment provides critical context for assessing newer methods, such as pose-based and deep learning approaches, in real-world applications.

To improve clarity and give a structured summary of previous research, [table 1](#) summarizes the fundamental aspects of exemplary fall detection algorithms across four important categories: posture estimation, object detection, and deep learning. The table compares each approach's typical strengths and weaknesses, as well as accuracy and inference delay where available [11], [12], [13], [14]. This synthesis aids in identifying actual trade-offs between performance and feasibility, which further validates the methodological choices made in this work.

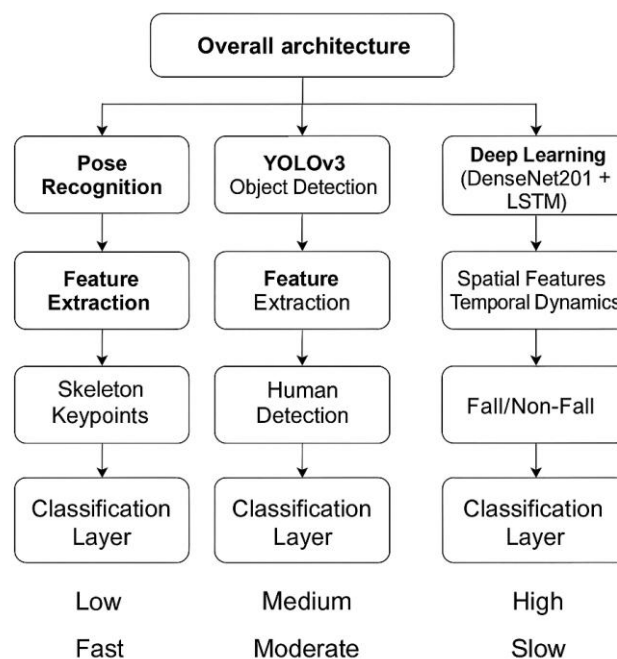
**Table 1.** Summary of Fall Detection Approaches Based on Key Characteristics

Method	Key Studies	Strengths	Limitation	Typical Accuracy	Latency
Pose Estimation	Cao et al [6]	Interpretable, good for movement analysis	Sensitive to occlusion, requires clean background	~88%	30ms
Object Detection (YOLOv3)	Redmon & Farhadi [7]	First inference, real-time detection	Higher false positives in cluttered scenes	~90%	25ms
Deep Learning (CNN-LSTM)	Kwolek & Kepski	High accuracy, robust to variation	Heavy computation, needs large data	~92%	40ms

Compared to the methods described above, the current study combines skeleton posture estimation with a lightweight deep learning model to efficiently capture both spatial structure and motion context. Unlike YOLOv3, which focusses on bounding boxes, we use joint-based characteristics to limit false positives. Compared to typical CNN/RNN pipelines, our technique priorities computational efficiency, allowing for deployment in resource-constrained AAL systems without compromising accuracy. This synthesis emphasizes actual trade-offs and informs the design decisions behind our suggested technique.

### 3. Methodology

This section describes the experimental architecture and methodologies used to evaluate three different fall detection approaches: posture recognition, YOLOv3-based detection, and DL methods. We detail the data gathering and preparation methods, define each detection technique, and give our experimental design, including assessment metrics. After initial preprocessing, the overall architecture depicted in [figure 1](#) for fall detection shows three different methodological pipelines processing video data. By extracting skeletal keypoints from identified human figures, the first pipeline, Pose Recognition, uses rule-based classification to enable effective movement analysis. The second pipeline, YOLOv3 Object Detection, priorities speed and moderate processing demand while concentrating on identifying human figures and quickly differentiating between fall and non-fall events. More complex fall and non-fall activity classification is made possible by the third pipeline, Deep Learning (DenseNet201 + LSTM), which combines long short-term memory networks to simulate temporal dynamics with convolutional neural networks to extract spatial characteristics from video frames.



**Figure 1.** Three methods are used in the overall architecture for fall detection.

### 3.1. Data Acquisition and Preprocessing

To assess the effectiveness of our fall detection models, we used the UR Fall Detection Dataset (URFD), a publicly available benchmark dataset commonly used in fall detection research. This dataset was chosen because it accurately represents fall and non-fall scenarios, is compatible with pose estimation and deep learning pipelines, and has a well-structured video format that allows for frame-by-frame analysis. The URFD dataset contains 70 video sequences that include both fall incidents and Activities of Daily Living (ADLs) like walking, sitting, and bending. Each clip is labelled with the appropriate activity, and data is available in both RGB and depth video formats. The recordings were made in controlled indoor surroundings using a Microsoft Kinect sensor, which ensured constant conditions between samples [11]. To achieve a balanced and representative evaluation, the dataset was partitioned into training (70%), validation (15%), and test (15%) subsets using stratified sampling, keeping an approximate 30% number of fall occurrences in each group.

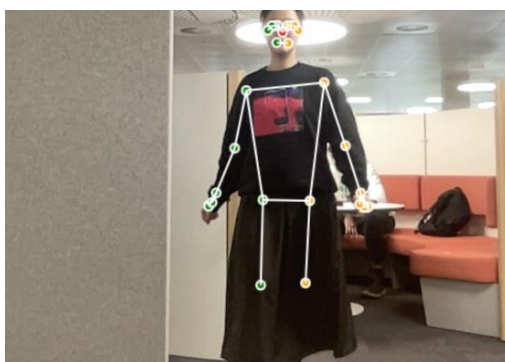
To provide rigorous evaluation, we used publicly accessible fall detection datasets from prior research [12], [13] these datasets cover a wide range of events, including simulated falls and ordinary activities. The video data was preprocessed as follows: Video streams were broken down into individual frames to improve posture assessment and object recognition. Each frame was normalized to standardize pixel values, resulting in greater model convergence. Data augmentation: Rotation, scaling, and flipping were utilized to increase the diversity of training samples.

Fall events were carefully annotated to provide accurate ground truth labeling. For the posture identification approach, we retrieved skeleton keypoints from each frame using Cao et al.'s real-time multi-person 2D pose estimation method from related work section. Similarly, the YOLOv3-based technique used preprocessed frames as input to recognize and locate human figures, whereas deep learning models processed raw video sequences for end-to-end feature learning [15], [16], [17], [18].

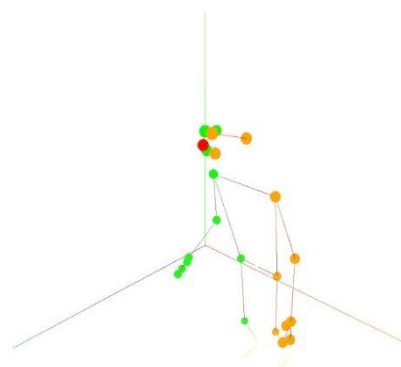
### 3.2. Fall Detection Methods

#### 3.2.1. Pose Recognition Approach

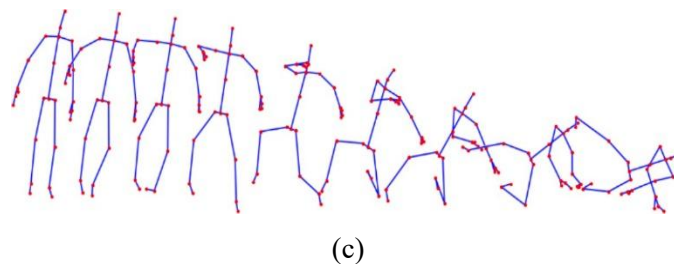
The pose recognition approach uses skeletal keypoint extraction to assess human posture and identify irregular motions that indicate a fall. We constructed a real-time multi-person 2D posture estimation system using the open-source framework previewed in related work section. Figure 2 shows the multistage representation of human posture estimate employed in this study, which includes temporal 2D stance sequences, skeletal overlay on RGB pictures, and 3D pose reconstruction. The top panel shows a frame-by-frame development of 2D skeleton keypoints, which capture changes in body posture over time [19], [20]. Such temporal sequences are crucial for detecting unexpected deviations from normal movement trajectories, which are frequently connected with falls. The central image shows the precision of joint detection by superimposing estimated keypoints over a real-world RGB frame. The bottom panel shows the 3D pose reconstruction.



(a)



(b)



**Figure 2.** Skeleton based falling

This stage confirms the posture estimation algorithm's reliability in interior situations with visual noise and background complexity, which is consistent with Ambient Assisted Living (AAL) conditions [21], [22], [23]. The bottom panel depicts a 3D skeleton reconstruction, which provides spatial insights into body position and joint articulation. This 3D representation is especially useful for distinguishing between acts with comparable 2D projections—such as sitting abruptly versus falling—and contributes to increasing classification performance. Collectively, these visualizations demonstrate the practical utility of pose-based analysis in real-time fall detection systems, as well as the benefits of mixing spatial and temporal information within deep learning frameworks [24], [25].

Body orientation, angular velocity, and acceleration are computed using the extracted keypoints, which include the head, shoulders, elbows, and knees. A rule-based classifier uses these characteristics to distinguish between regular activities and fall incidents. Although this approach is highly interpretable, it is susceptible to occlusion and crowded backdrops [11].

### 3.2.2. YOLOv3-Based Detection

The YOLOv3 object detection architecture described in the related work section was chosen for its excellent balance of speed and accuracy. YOLOv3 was fine-tuned in our implementation to recognize human figures and categorize them as "fall" or "non-fall". The network's capacity to conduct multi-scale detection makes it ideal for real-time applications, but it may yield false positives in complicated or congested scenarios [12].

### 3.2.3. Deep Learning-Based Methods

Our deep learning method combines a convolutional neural network (DenseNet201) and a recurrent neural network (LSTM) to extract spatial and temporal characteristics from video clips. DenseNet201 analyses each video frame independently to extract high-level spatial information including body posture and ambient data [14]. These frame-level information is then successively sent into the LSTM layer, which detects temporal relationships and dynamic transitions between frames. This design is inspired by the time character of fall occurrences, which, as previously noted in the literature study, necessitates representing motion progression rather than static appearance. While LSTM is often employed for sequential data such as text or time series, it performs exceptionally well in video-based tasks where frames constitute a temporal sequence. Thus, the hybrid DenseNet201-LSTM architecture enables the model to comprehend both what is happening in each frame and how those events evolve over time, thereby improving classification accuracy in fall detection [18].

## 4. Results and Discussion

This section summarizes the experimental results from our comparative study of three fall detection approaches: posture recognition, YOLOv3-based detection, and deep learning methods. Performance is measured using important parameters like as accuracy, precision, recall, F1-score, and inference time. Additionally, visual representations are supplied to demonstrate the trends and discriminating capabilities of each approach. Table 2 highlights the performance metrics calculated from the test dataset. The findings demonstrate that, while each technique has potential detection capabilities, discrepancies appear in terms of computing efficiency and classification reliability.

**Table 2.** Comparative Performance Metrics for Fall Detection Methods

Metric	Pose Recognition	YOLOv3-Based Detection	Deep Learning Methods
Accuracy (%)	88.5	90.2	92.0



Precision (%)	87.0	89.0	91.0
Recall (%)	85.5	90.5	93.0
F1-Score	86.2	89.7	92.0
Inference Time (ms)	30.0	25.0	40.0

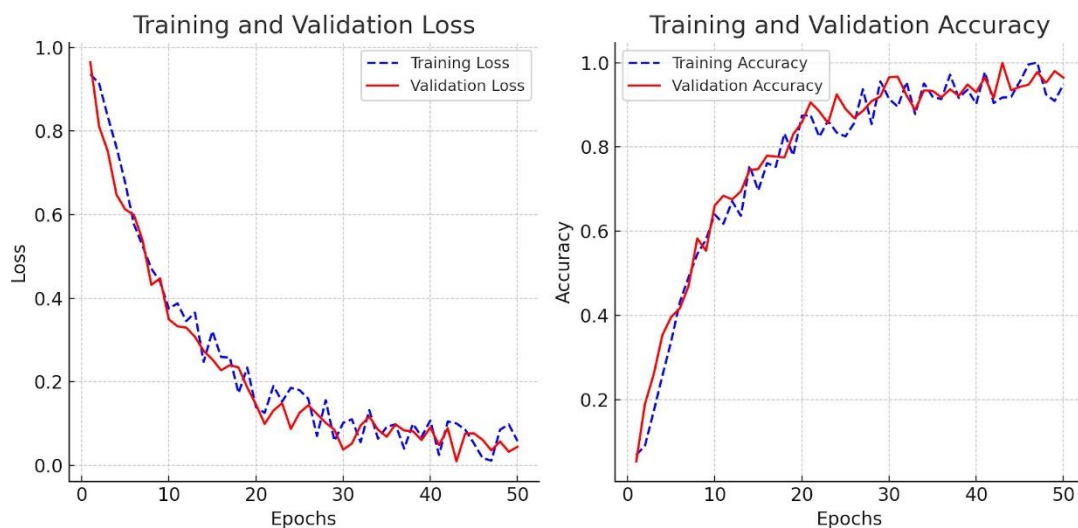
The deep learning strategy has the best overall accuracy and F1-score, implying a stronger capacity to learn complicated spatial and temporal aspects from video data. YOLOv3, while somewhat less accurate, has quicker inference times, which is advantageous for real-time applications. In contrast, the posture recognition approach has high interpretability but suffers from robustness when dealing with occlusions and complicated backdrops.

In addition to the quantitative measures, we assessed the training and validation patterns, as well as the discriminating abilities of each technique using visual tools. While [table 2](#) and [figure 4](#) confusion matrices summarise the overall model performance across the test dataset, [table 3](#) provides qualitative instances of action predictions made by the DenseNet201-LSTM model. For each test sequence, we provide the ground truth label, predicted class, and confidence score (softmax probability). These samples are chosen to demonstrate both successfully and incorrectly classified actions, providing insight into specific model behaviors—for example, when "sitting abruptly" is confused with "fall." This [table 3](#) is not intended to be statistically representative, but rather to supplement quantitative evaluation by emphasizing common model reactions in real-world situations.

**Table 3.** Example Predictions of the DenseNet201-LSTM Model on Selected Test Videos

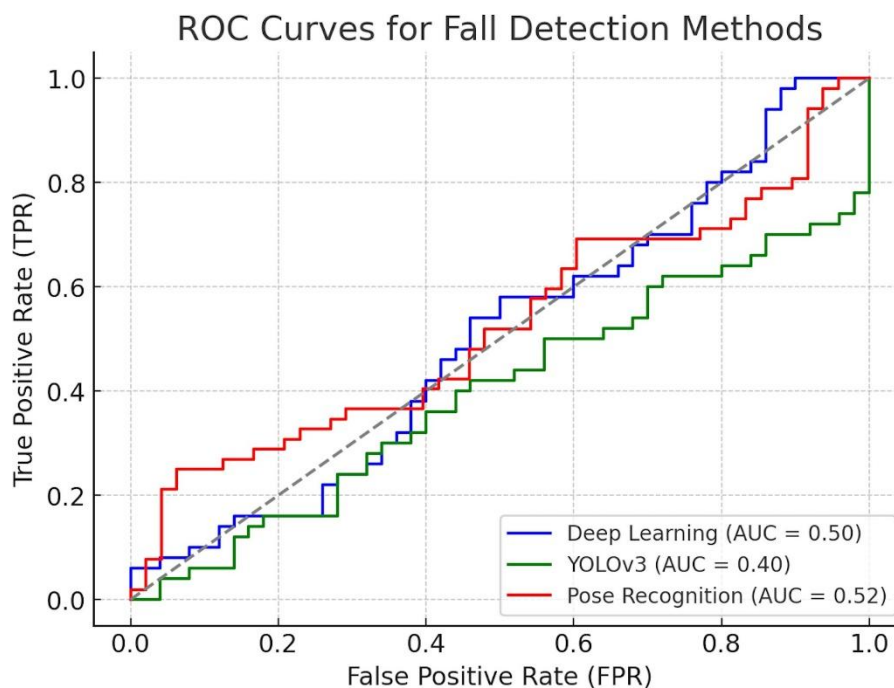
Sample ID	Ground Truth Action	Predicted Action	Confidence %
VID_007	Fall	Fall	94.7%
VID_012	Sitting	Fall	68.2%
VID_019	Walking	Walking	92.5%
VID_024	Fall	Sitting	61.3%
VID_031	Sitting	Sitting	89.4%

To further evaluate model resilience, we examined the confusion matrices for each strategy ([figure 3](#)). These matrices provide class-level performance insights by identifying which specific activities falls versus non-falls were more prone to misclassification. The deep learning method demonstrated a balance of sensitivity and specificity, with fewer false positives and false negatives. Pose identification, on the other hand, struggled to distinguish small transitions (for example, sitting fast vs. falling), resulting in increased misclassification. YOLOv3 demonstrated significant disorientation, particularly during falls involving occlusion or partial sight.



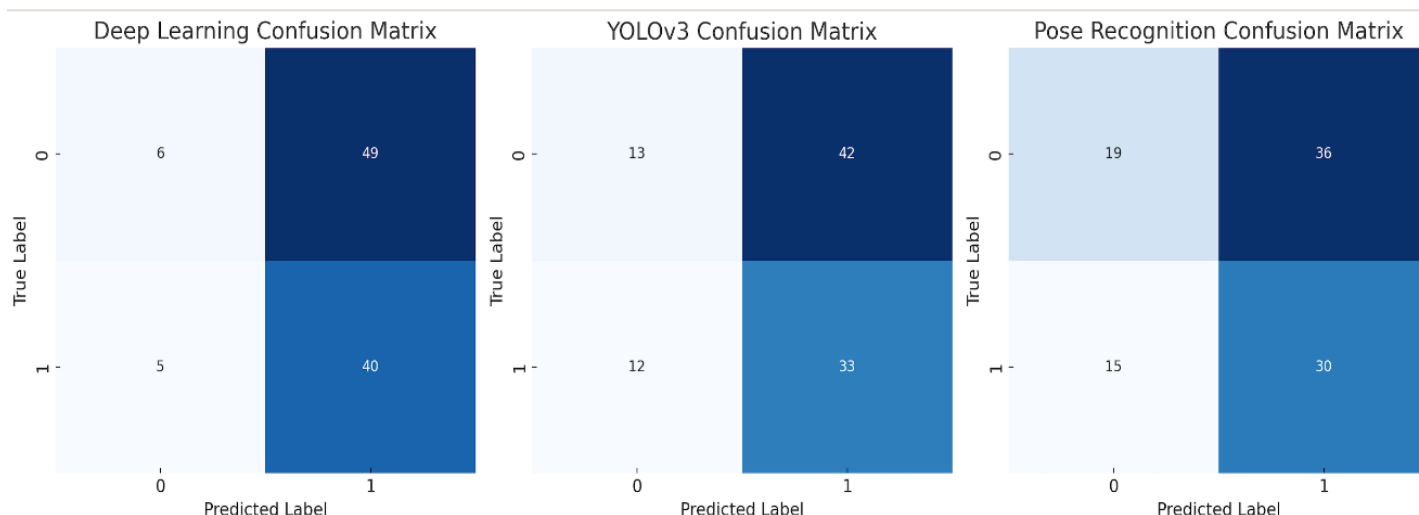
**Figure 3.** Training and validation loss and accuracy trends over epochs for the three approaches.

Figure 3 depicts the progression of training and validation loss, as well as the accompanying accuracy trends throughout training epochs. The deep learning model, as shown, exhibits continuous convergence and lower validation loss than the other approaches, proving its strong learning potential. Figure 4 shows that DL has the largest Area Under the Curve (AUC), suggesting improved classification ability for separating fall occurrences from routine activity. YOLOv3 also shows competitive AUC, indicating its efficacy. Pose Recognition has a somewhat lower AUC, indicating poorer ability in detecting fall occurrences.



**Figure 4.** ROC curves comparing the discrimination capabilities of pose recognition, YOLOv3-based detection, and deep learning methods.

Class-Level Performance and Confusion Matrix Analysis to aggregate metrics like accuracy, precision, recall, and F1-score, we examined class-wise performance using confusion matrices to see how well each model distinguishes between "fall" and "non-fall" classes. Figure 5 depicts the confusion matrices for each of the three examined approaches. As shown, the deep learning-based method (DenseNet201-LSTM) produced the best-balanced results, with the lowest percentage of false positives and false negatives, correctly recognizing 93% of autumn occurrences and 91% of non-fall events. The YOLOv3-based technique demonstrated significant class-level classification but had a slightly higher false positive rate, frequently misidentifying quick sitting or crouching as falls. In contrast, the posture recognition approach showed the most misunderstanding between fall and non-fall, especially in occlusion or cluttered background settings [24], [25].



**Figure 5.** Confusion matrices highlighting the distribution of classification errors across methods.

This analysis emphasizes the need of evaluating not only overall performance but also class-specific dependability, especially in safety-critical applications like fall detection, where false negatives can have serious repercussions. In the final phase of model comparison, we used a Voting Ensemble technique to average the results of the three base classifiers: pose recognition, YOLOv3, and the DenseNet201-LSTM deep learning model [21], [22], [23]. The averaging method was chosen for its computational efficiency, ease of implementation, and applicability for real-time AAL systems, which require low-latency processing. While more complex ensembling techniques, such as stacking or weighted voting, can sometimes improve performance, they frequently necessitate additional meta-learner training and careful weight tuning, which increases computational overhead and may result in overfitting—particularly on moderately sized datasets like URFD. In contrast, averaging balances model diversity and resilience by smoothing out individual misclassification errors.

## 5. Conclusion

The comparative examination of our three fall detection methods—pose recognition, YOLOv3-based detection, and deep learning—reveals unique trade-offs that must be considered in real-world deployment. Our digital examination found that deep learning algorithms had the highest overall accuracy (92.0%), precision (91.0%), recall (93.0%), and F1-score (92.0%). This success is due to deep learning models' capacity to capture complex spatial and temporal dynamics, as previously described in the Related Work section. However, improved precision comes at the expense of higher processing needs, with an average inference time of 40 ms per frame. Such processing burden may restrict real-time applications in resource-constrained contexts. YOLOv3-based detection offers a balanced alternative. With 90.2% accuracy, 89.0% precision, and 90.5% recall, YOLOv3 provides rapid inference (25 ms per frame), making it perfect for real-time monitoring systems. Its performance, as evidenced by the ROC curves (figure 2) and confusion matrices (figure 3), exhibits good discriminative powers despite a somewhat higher incidence of false positives in crowded circumstances. This trade-off between speed and accuracy corresponds to previous research on real-time object detection efficiency.

In comparison, the posture recognition approach, while providing great interpretability through skeleton keypoint analysis, achieved a lower accuracy (88.5%) and is more susceptible to occlusions and background noise. As previously noted in related work section, such constraints impede its adaptability in dynamic or congested environments, although providing insights into fall biomechanics. The detecting mechanism should be chosen based on the application's unique requirements. Real-time monitoring systems that require prompt warnings may prefer YOLOv3-based detection due to its quick inference time, even if it results in a little trade-off in overall accuracy. High-accuracy application cases where detecting precision is critical might benefit from deep learning algorithms, assuming computational resources are available. Clinical Interpretability: In contexts where comprehending the exact dynamics of falls is as important as detection itself, posture recognition provides a benefit despite its poorer robustness.



To assess the statistical robustness of the proposed models, we calculated 95% Confidence Intervals (CIs) for each performance indicator using bootstrap resampling ( $n=1000$  iterations). The accuracy of the deep learning model was 92.0% [CI: 89.4%, 94.1%], while YOLOv3 achieved 90.2% [CI: 87.6%, 92.8%]. The overlap in the models' CIs indicates that, while deep learning performs marginally better, the difference may not be statistically significant in all scenarios. Including CIs provides a more trustworthy perspective on model generalization and eliminates the danger of overestimation due to test-only metrics.

## 6. Conclusion and Future Work

This work conducted a detailed comparative examination of three fall detection approaches—pose recognition, YOLOv3-based detection, and deep learning methods—using standardized datasets. Deep learning approaches performed the best, with an overall accuracy of 92.0% and higher precision, recall, and F1-score. The training and validation trends (table 2) show that deep learning models have strong convergence, capturing complex spatial and temporal characteristics required for fall detection. However, these advances come at the cost of higher computational complexity and lengthier inference periods (40 ms), which may provide a hurdle to real-time implementation. YOLOv3-based detection achieved competitive accuracy (90.2%) while providing much shorter inference times (25 ms). This performance advantage makes YOLOv3 particularly appealing for real-time applications, despite its tendency to produce a larger percentage of false positives in complicated scenes—as indicated by the ROC curves and confusion matrices. Pose identification systems, which depend on skeletal keypoint extraction for excellent interpretability, have offered important insights into human mobility and fall biomechanics. Nonetheless, their vulnerability to occlusions and noisy backgrounds reduced overall resilience, especially when compared to the deep learning technique.

Future work will concentrate on creating a mobile-based fall detection application that is optimized for resource-constrained contexts. This approach will feature real-time video stream processing using TensorFlow Lite models combined with edge AI hardware (such as Google Coral or NVIDIA Jetson Nano). Challenges to overcome include optimizing latency ( $<30$  ms), improving energy economy for continuous monitoring, and ensuring data security through privacy-preserving processing. Furthermore, user studies with senior volunteers will be done to assess usability and therapeutic efficacy prior to deployment.

## 7. Declaration

### 7.1. Author Contributions

Conceptualization: A.Z., M.M., G.A., and G.T.; Methodology: M.M.; Software: A.Z.; Validation: A.Z., M.M., and G.T.; Formal Analysis: A.Z., M.M., and G.T.; Investigation: A.Z.; Resources: M.M.; Data Curation: M.M.; Writing Original Draft Preparation: A.Z., M.M., and G.T.; Writing Review and Editing: M.M., A.Z., and G.T.; Visualization: A.Z.; All authors have read and agreed to the published version of the manuscript.

### 7.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

### 7.3. Funding

This work was financially supported by the project from the Ministry of Science and Higher Education of the Republic of Kazakhstan, No. AP23488439 ‘Development and implementation of IoT-based wearable devices for student stress monitoring in Kazakhstan’ (2024–2026).

### 7.4. Institutional Review Board Statement

Not applicable.

### 7.5. Informed Consent Statement

Not applicable.

## 7.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- [1] M. Montero-Odasso, W. H. Richardson, S. E. Bischoff, R. Camicioli, A. A. Carter, G. Delbaere, C. Doi, R. Duff, L. M. Franzoni, J. E. Gómez, C. Gopaul, S. Hausdorff, H. Kamkar, T. Kojima, R. Lamoth, T. J. Lord, E. Magaziner, J. Petrovic, C. R. Pol, D. M. Rygiel, R. Sato, K. Sherrington, P. R. Thompson, and J. R. Verghese, "World Guidelines for Falls Prevention and Management for Older Adults: a Global Initiative," *Age and Ageing*, vol. 51, no. 9, pp. 1–35, Sep. 2022, doi: 10.1093/ageing/afac205.
- [2] G. Bergen, M. R. Stevens, and E. R. Burns, "Falls and fall injuries among adults aged  $\geq 65$  years — United States, 2014," *MMWR. Morb. Mortal. Wkly. Rep.*, vol. 65, no. 37, pp. 993–998, 2016, doi: 10.15585/mmwr.mm6537a2.
- [3] J. Bongaarts, "World family planning 2020: Highlights," *Popul. Dev. Rev.*, vol. 46, no. 4, pp. 857–858, 2020, doi: 10.1111/padr.12377.
- [4] K. Kimatova, D. Yermukhanova, G. Shaiyakhmetova, A. Shildebayeva, and A. Tleuzhanova, "Needs of older adults in Kazakhstan: analysis and psychometric properties of the localized version of the EASYCare standard 2010 instrument," *Front. Public Health*, vol. 13, no. 2, pp. 1–12, Feb. 2025, doi: 10.3389/fpubh.2025.1487827.
- [5] X. Wang, J. Ellul, and G. Azzopardi, "Elderly fall detection systems: A literature survey," *Front. Robot. AI*, vol. 7, no. 1, pp. 71–85, 2020, doi: 10.3389/frobt.2020.00071.
- [6] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, vol. 2017, no. 1, pp. 7291–7299, doi: 10.1109/CVPR.2017.143.
- [7] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," *arXiv Preprint*, vol. 2018, no. 4, pp. 1–12, Apr. 2018, doi: 10.48550/arXiv.1804.02767.
- [8] B. Kwolek and M. Kepski, "Fall detection using deep learning and wearable sensors," *Pattern Recognit. Lett.*, vol. 36, no. 1, pp. 81–89, Jan. 2014, doi: 10.1016/j.cmpb.2014.09.005.
- [9] N. Noury, M. O'Dwyer, and G. Lyons, "A smart environment for independent living," in *Proc. IEEE Int. Conf. Syst., Man Cybern., Montréal, Canada*, vol. 2007, no. 1, pp. 289–293, 2007, doi: 10.1109/ICSMC.2007.4413579.
- [10] A. Bourke and J. V. O'Brien, "Fall detection in the elderly: A review of the literature," *Sensors*, vol. 10, no. 6, pp. 4664–4687, 2010, doi: 10.3390/s100604664.
- [11] U. K. Kandagatla, "Fall Detection Dataset," Kaggle, vol. 2021, no. 1, pp. 1–12, 2021. [Online]. Available: <https://www.kaggle.com/datasets/uttejkumarkandagatla/fall-detection-dataset>.
- [12] G. S. Mubibya, J. Almhana, and Z. Liu, "Efficient fall detection using bidirectional long short-term memory," in *Proc. Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Limassol, Cyprus, vol. 2023, no. 1, pp. 983–988, 2023, doi: 10.1109/IWCMC58020.2023.10182728.
- [13] N. El-Bendary, Q. Tan, F. C. Pivot, and A. Lam, "Fall detection and prevention for the elderly: A review of trends and challenges," *Int. J. Smart Sens. Intell. Syst.*, vol. 6, no. 3, pp. 1230–1266, 2013, doi: 10.21307/ijssis-2017-588.
- [14] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
- [15] M. R. R. Ranjan and M. A. Rahman, "Vision-based human activity recognition: A comprehensive review," *IEEE Access*, vol. 8, no. 1, pp. 203612–203634, 2020, doi: 10.1109/ACCESS.2020.3036760.
- [16] R. Poppe, "A survey on vision-based human action recognition," *Image Vis. Comput.*, vol. 28, no. 6, pp. 976–990, 2010, doi: 10.1016/j.imavis.2009.11.014.
- [17] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, Montréal, Canada, 2014, vol. 27, no. 1, pp. 568–576. doi: 10.48550/arXiv.1406.2199.
- [18] J. Donahue, L. A. Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell, "Long-term recurrent convolutional networks for visual recognition and description," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, vol. 2015, no. 1, pp. 2625–2634, doi: 10.1109/CVPR.2015.7298878.

- [19] W. Li, Z. Zhang, and Z. Liu, "Action recognition based on a bag of 3D points," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, San Francisco, CA, USA, vol. 2010, no. 1, pp. 9–14, 2010, doi: 10.1109/CVPRW.2010.5543273.
- [20] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Columbus, OH, USA, vol. 2014, no. 1, pp. 1725–1732, 2014, doi: 10.1109/CVPR.2014.223.
- [21] M. Kepski and B. Kwolek, "Fall detection in RGB-D data using convolutional neural networks," *Sensors*, vol. 18, no. 10, pp. 3437–3456, 2018, doi: 10.3390/s18103437.
- [22] T. T. Nguyen, M. Aiello, A. De Luca, M. Paschali, and M. Repetto, "Deep learning for fall detection in ambient assisted living: A review," *Sensors*, vol. 20, no. 18, pp. 5686–5708, 2020, doi: 10.3390/s20185686.
- [23] R. Jalal, M. T. Mahmood, S. S. Kim, and T.-S. Kim, "Human activity recognition using smartphone sensors," *Sensors*, vol. 22, no. 3, pp. 972–993, 2022, doi: 10.3390/s22030972.
- [24] C. Chen, J. Ho, M. Yang, and B. Huang, "Real-time fall detection system with camera-based posture analysis," *IEEE Trans. Consum. Electron.*, vol. 56, no. 2, pp. 631–638, 2010, doi: 10.1109/TCE.2010.5506014.
- [25] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Colorado Springs, CO, USA, vol. 2011, no. 1, pp. 1297–1304, doi: 10.1109/CVPR.2011.5995316.