Proceedings of CONF-APMM 2025 Symposium: Controlling Robotic Manipulator Using PWM Signals with Microcontrollers
DOI: 10.54254/2753-8818/2025.AD26456

Real-time Fall Monitoring System for the Elderly Based on Multimodal Sensor Fusion

Wanning Chen

College of Silesian Intelligent Science and Engineering, Yanshan University, Qinhuangdao, China
chenwanning123456789@outlook.com

Abstract. With the acceleration of the global aging process, falls among the elderly have become a major public health issue threatening their health. Currently, the single sensor monitoring technology has significant limitations: the misjudgment rate of wearable accelerometers for daily activities, visual monitoring is significantly affected by light and there is a potential risk of privacy leakage, making it difficult to adapt to complex home scenarios. This paper reviews the research progress of real - time fall monitoring systems for the elderly based on multimodal sensor fusion, focuses on analyzing the collaborative mechanisms of millimeter - wave radar, accelerometers, and heart rate sensors, and summarizes key technologies such as data fusion architectures, algorithm optimization, and edge computing deployment. By comparing the performance differences of different fusion strategies, it is found that the three - level attention fusion architecture performs best in complex scenarios. At the same time, this paper points out problems in current research, such as the insufficient proportion of open - source data in home scenarios and the lack of night - time monitoring solutions, and looks forward to the future development direction of combining the Transformer architecture with privacy computing.

Keywords: Elderly Fall Monitoring, Multimodal Sensor Fusion, Real-time Early Warning, Edge Computing, Privacy Protection

1. Introduction

The increasing global aging population has made the problem of falls among the elderly increasingly severe. According to data from the National Health Commission [1], approximately 684,000 people worldwide die each year from fall-related injuries, with those aged 60 and above accounting for over 60%. The number of elderly people living alone in China has exceeded 120 million, and the fatality rate of those who are not treated in a timely manner after a fall reaches 30%. However, intervention within the "golden one-hour" can reduce the risk of death by 50%. The latest report from the World Health Organization further points out that falls have become the leading cause of accidental death among the elderly aged 70 and above. Among them, falls in the home environment account for as high as 85%, far exceeding those in public places (10%) and medical institutions (5%) [2]. This data highlights the urgency of fall monitoring in the home scenario.

The limitations of existing monitoring technologies have become an industry consensus. Although visual monitoring can capture fine movements, measurements from 100 households show

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2025.AD26456

that its missed detection rate surges to 35% in backlight and shadow scenarios. Moreover, 83% of the elderly are resistant to the installation of cameras due to privacy concerns [3]. Regarding wearable devices, a tracking study of 200 elderly people living alone found that due to postural interference, the false alarm rate of accelerometers reaches 24.3% when putting on/taking off clothes or picking up items, leading to a user abandonment rate of over 40% [4]. In addition, single technologies such as barometric pressure sensors and infrared sensors also have their respective shortcomings: the former has a too low sensitivity to height changes (error ±5cm), and the latter is easily affected by furniture occlusion (missed detection rate 18%) [5].

The rise of multimodal fusion technology provides new ideas for solving the above problems. The 77GHz band millimeter-wave radar can stably capture the human motion trajectory under the cover of bedding and in a dark environment due to its penetrability, and still maintain a 92% detection rate in the bathroom steam environment [6]. The combination of an accelerometer and a heart rate sensor can reduce misjudgments through "motion + physiological" double verification. The sudden increase in heart rate (ΔHR) within 5 seconds after a fall is positively correlated with the degree of injury (r=0.76), providing a quantitative basis for emergency intervention [7]. The maturity of edge computing technology has further promoted the breakthrough in real-time performance. The Raspberry Pi terminal has achieved local response within 8 seconds, which is more than 60% shorter than the traditional cloud - based solution [8].

Despite the progress made in research, there are still three major gaps in the current field: First, the proportion of home scenarios in open-source datasets is insufficient (for example, the UCI HAR Dataset only accounts for 12%), and the difference between laboratory data and the real home environment leads to a 15%-20% decrease in the accuracy of the model when it is put into practice. Second, there is a lack of low-power consumption solutions at night. The battery life of existing devices is generally less than 48 hours, making it difficult to meet the 24-hour monitoring requirement. Third, the cross-modal feature association mechanism needs to be improved. The time synchronization error (±0.5 seconds) of data from different sensors may lead to fusion failure.

The structure of this paper is as follows: Chapter 2 reviews the data sources and preprocessing methods of the multimodal monitoring system, compares the feature extraction strategies of different sensors, and analyzes the fusion architecture; Chapter 3 analyzes the algorithm optimization technology, focuses on the combination of the attention mechanism and machine learning models, compares the system performance through experimental data, and discusses the scenario adaptation issues in practical applications; Chapter 4 summarizes the research progress and looks forward to the future directions.

2. Methods

2.1. Introduction to related datasets

The data sources of the multi-modal monitoring system need to cover the four-dimensional information of "motion - space - physiology - environment", and the specific parameters and acquisition standards are as follows:

The UCI HAR Dataset contains accelerometer (3 axes, 50Hz) and gyroscope (3 axes, 50Hz) data from 30 subjects (aged 19 - 48), recording 10 types of activities (6 daily activities + 4 fall types), with a sample size of 10,299. During data collection, the devices were worn at the waist, which highly matches the daily wearing habits of the elderly (such as in pockets or on belts). However, the high proportion of young samples (80%) may lead to insufficient adaptability of the model to the physiological characteristics of the elderly [9].

DOI: 10.54254/2753-8818/2025.AD26456

The Kaggle Fall Detection Dataset covers millimeter-wave radar (77GHz, 10Hz) and heart rate belt (1Hz) data of 200 elderly people (aged 65 - 85), containing 5,000 labeled events ($fall\ /\ non\ -\ fall\ = 1:3$). The radar data records the distance - time matrix (resolution $0.1m \times 0.1s$), from which 12 - dimensional features such as centroid displacement and velocity can be extracted. The heart rate data contains physiological indicators such as RR interval and HRV, providing key evidence for fall confirmation [10].

In terms of supplementing clinical data, a sample survey by the National Bureau of Statistics (2024) shows that the high - incidence scenarios of elderly people falling are the bathroom (32.7%), bedroom (28.3%), and living room (21.5%) in sequence. The main floor materials are ceramic tiles (42%), wooden floors (35%), and carpets (23%), providing environmental parameters for scenario - based model training. The 1,200 cases of clinical data from Peking Union Medical College Hospital (2022) have established the "falling posture - injury level" correspondence (for example, the fracture risk of falling in a prone position is 3.2 times that of sitting), which is used to optimize the warning priority.

2.2. Data preprocessing and feature extraction

The preprocessing steps need to be optimized for the heterogeneity of multi-source data. For missing value processing, linear interpolation is used to fill the missing values in the acceleration and heart rate data (with a missing rate < 3%), and Kalman filter prediction is adopted for the packet loss of radar signals, improving the data integrity to over 98%. In terms of standardization, Z-score standardization is performed on the acceleration data (range ±2g) and heart rate data (60 - 120 bpm) to eliminate the dimensional differences; min-max normalization is applied to the radar distance data to the interval [0, 1] to enhance the model generalization ability. For sample balancing, the SMOTE algorithm ($k = 5$ nerby) is used to increase the proportion of fall samples from 15% to 30%. At the same time, random oversampling is used to supplement samples in scarce scenarios such as at night and in the bathroom, making the proportion of samples in each scenario consistent with the real distribution [11]. In the feature enhancement stage, short - time Fourier transform (STFT, window length 512, overlap rate 50%) is carried out on the millimeter - wave radar data to convert it into a time - frequency diagram, retaining the distance – time. Dynamic characteristics of the inter-domain; The acceleration data is segmented using a sliding window ($512/128$ sampling points), and the robustness is enhanced through adding noise (Gaussian noise $\sigma = 0.01$) and time reversal [6].

The feature extraction strategies for different sensors need to be matched with their physical characteristics. In terms of millimeter-wave radar, the 77GHz frequency band can penetrate obstacles such as clothing and glass, with a distance resolution of 0.1m and a velocity resolution of 0.05m/s, enabling the capture of the three-dimensional motion trajectory of the human body's centroid. The feature extraction process consists of three steps: removing static background noise through Constant False Alarm Rate (CFAR) detection; extracting micro-Doppler features using the Fast Fourier Transform (FFT) (for example, the arm swing frequency is 1 - 3Hz, and the torso rotation angular velocity is > 5rad/s when falling); inputting into the ResNet18 network (containing 4 residual blocks) to extract deep spatial features, and outputting a 128-dimensional feature vector. This method improves the accuracy in occluded scenes by 18% compared to the traditional threshold method [6].

The accelerometer uses a 3-axis MEMS accelerometer (measurement range ±8g, sampling rate 100Hz) to record the motion state. Feature extraction focuses on two types of information: time-domain features (peak value, kurtosis, integral area) are used to identify the fall impact with a

DOI: 10.54254/2753-8818/2025.AD26456

vertical acceleration > 1.5g; frequency-domain features (energy ratio in the 0 - 5Hz frequency band) are used to distinguish the frequency differences between walking (1 - 2Hz) and falling (3 - 5Hz). The BiLSTM model (2 hidden layers, 64 neurons) improves the accuracy of distinguishing between bending over and falling to 92% through forward/backward temporal sequence modeling [4].

The heart rate sensor uses photoplethysmography (PPG) to collect heart rate signals (sampling rate 25Hz). The preprocessing includes baseline drift removal (wavelet threshold denoising) and peak detection (Pan-Tompkins algorithm). Key features include: the magnitude of sudden heart rate changes ($\Delta HR > 25bpm$); the time-domain indicators (SDNN, RMSSD) and frequency-domain indicators (LF/HF) of heart rate variability (HRV). The LightGBM model shows that HRV features can achieve an AUC value of 0.89 for the fall/non-fall binary classification, which is significantly better than single motion features (0.76) [7].

2.3. Multimodal fusion architecture

Existing research has proposed various fusion strategies. This paper focuses on reviewing the three-level attention fusion architecture and supplements the comparative analysis of early fusion and late fusion. Early fusion directly concatenates multi-source signals at the data layer (such as radar distance sequence + acceleration time-domain features) and inputs them into the CNN-LSTM hybrid model. Its advantage is to preserve the correlation of original information, but the disadvantage is that it is vulnerable to noise accumulation. In high-noise scenarios (such as TV interference), the accuracy drops to 82% and the false alarm rate rises to 12% [5].

Late fusion involves independent modeling of each modality, followed by decision - making through weighted voting (where the weights are determined by the F1 score of the validation set). For example, the radar model outputs the fall probability P1, the acceleration model outputs P2, and the heart - rate model outputs P3. The final decision is ($P = 0.4\ P\ 1 + 0.3\ P\ 2 + 0.3\ P\ 3$) (with a threshold of 0.5). This method controls the false - alarm rate at 7.5%, but due to ignoring feature correlations, the accuracy in complex scenarios (such as falling while making a phone call) is only 85% [12].

The three-level attention fusion architecture is the optimization plan that this paper focuses on. Its structure is divided into three parts: The bottom-level perception layer extracts features through ResNet18 (radar time-frequency map) and BiLSTM (acceleration time series) respectively. The feature fusion layer introduces a dynamic attention mechanism, calculates the weight through the formul $\alpha = \sigma(W1 \cdot [Fmotion; Fspatial] + b1)$ (where σ is the sigmoid function), and realizes the adaptive weighting of motion features (F_motion) and spatial features (F_spatial). For example, in the bathroom scene, the weight of radar features is increased to 0.6 (to counter steam interference), while in an open space, the weight of acceleration features is dominant (0.55). The decision layer fuses physiological signals (heart rate features) and environmental parameters (ground material, light intensity) and then inputs them into the LightGBM classifier (threshold 0.6). Edge deployment is implemented using NVIDIA Jetson Nano (GPU computing power of 21 TOPS), with local inference power consumption < 10W, and it supports continuous operation for 72 hours [8].

The model training uses the PyTorch framework. To address the problem of the low proportion of fall samples, Focal Loss (α =0.75 , γ=2) is adopted to suppress the weights of the majority class (non-fall). Combined with the Ranger optimizer (learning rate $5e-4$, weight decay $1e-5$), the convergence is accelerated. 5-fold cross-validation shows that this strategy increases the F1 score by 0.08 (reaching 0.96) compared to the ordinary cross-entropy loss.

DOI: 10.54254/2753-8818/2025.AD26456

To solve the problem of time misalignment of multi-sensor data, the system introduces a hardware synchronization mechanism: The sampling start points of the radar, accelerometer, and heart rate sensor are aligned through GPS timekeeping (with an error of ±1ms) or a hardware trigger signal, and the timestamp accuracy is controlled within 5ms. At the software level, the Dynamic Time Warping (DTW) algorithm is used to elastically match asynchronous data, reducing the time synchronization error from ±0.5 seconds to ±0.1 seconds [4].

In terms of low-power design, the "event-triggered + sleep" mode is adopted: the sensor is in a low-power state by default (current<1mA). When the acceleration exceeds 0.5g or the radar detects human movement, data collection is activated (current 10 - 20mA). With a lithium battery (5000mAh), a 72-hour battery life can be achieved. This solution reduces power consumption by 80% compared to the full-time data collection mode [7].

3. Results and disscussion

As showen in table 1,the multi-modal fusion model comprehensively outperforms single sensors and traditional fusion methods on the Kaggle test set :

Table 1. Performance comparison of different models (Kaggle test set,$n = 1000$)

Model	Accuracy	False positive rate	F1 fraction	Response Time (Seconds)
Single millimeter-wave radar	95.8%	5.2%	0.93	7.5
Single accelerometer	82.7%	24.3%	0.78	1.2
Early Fusion (CNN-LSTM)	90.5%	8.7%	0.89	9.8
Three-level attention fusion	97.3%	3.1%	0.96	6.2

UCI dataset verification shows that the model's ability to distinguish between easily confused activities has been significantly improved: the misjudgment rate of the bending action has decreased from 12.5% to 1.8%, and the misjudgment rate of the jumping action has decreased from 8.3% to 0.9%. This is due to the attention mechanism focusing on key features - the difference in the peak acceleration of the torso during a fall (an average of 2.1g) and during bending (an average of 0.8g) is magnified, and at the same time, the heart rate signals ($\Delta HR = 28 \pm 5\ bpm$) during a fall, ($\Delta HR = 5 \pm 2\ bpm$) during bending) provide a secondary verification [7].

Scene adaptability analysis shows that the low-friction characteristics of the tiled floor in the bathroom scene lead to a faster fall speed (an average of 1.8m/s). The model improves its performance through three adjustments: increasing the radar feature weight to 0.6 (to counter steam interference); reducing the acceleration threshold to 1.2g (to correct the buffering effect); and shortening the response time to sudden heart rate changes to 3 seconds. 100 simulation experiments show that the accuracy rate reaches 96.5%, which is only a 0.8% decrease compared to the laboratory scene [13].

In the carpet scenario, the buffering effect of thick carpets (>5cm) reduces the peak acceleration by 30%-40%. By integrating the feature from the radar that "the trunk tilt angle > 45° and lasts for 2 seconds", the model reduces the missed detection rate from 15% to 5.2%. Data from the National Bureau of Statistics shows that this scenario accounts for 23% of household falls, and after optimization, it can cover over 90% of home scenarios.

In night-time scenes, the combination of millimeter-wave radar and infrared light compensation solves the lighting problem. However, night-time samples only account for 38% of the existing data, causing the accuracy of the model to drop to 91% between 2 and 5 am (a high-incidence period).

DOI: 10.54254/2753-8818/2025.AD26456

The data augmentation method of generating virtual night-time samples through GAN can increase the night-time accuracy by 6% [12].

In terms of real - time performance and privacy protection, the edge - computing deployment has achieved the localization of the entire "detection - decision - alarm" process. The GPU acceleration of NVIDIA Jetson Nano reduces the feature - extraction time to 2.1 seconds, the LightGBM inference time is 0.3 seconds, and with the data - transmission delay of 3.8 seconds, the total response time is 6.2 seconds, which is 50% faster than the cloud - based solution (12.5seconds). The pilot projects in 10 communities show that this system shortens the average time to first - aid after a fall from 47 minutes to 12 minutes, meeting the "golden 1 - hour" intervention requirement [8].

The privacy protection adopts the principle of "data not leaving the country": Radar and acceleration data are stored locally (in encrypted format). Only alarm information (excluding original data) is uploaded when an abnormal event is triggered, which complies with Article 25 of the GDPR, the "data minimization" principle [14]. User research shows that 89% of the elderly have a higher acceptance of this solution than visual monitoring (only 41%), significantly improving long-term use compliance.

4. Conclusions

The multimodal sensor fusion technology effectively balances the accuracy, real-time performance, and privacy protection of elderly fall monitoring by integrating millimeter-wave radar, accelerometer, and heart rate data. The three-level attention fusion architecture performs optimally in complex home environments. By combining spatial trajectories, motion characteristics, and physiological signals, it increases the accuracy rate to 97.3% and reduces the false alarm rate to 3.1%. The edge computing deployment achieves a rapid response time of 6.2 seconds, laying a technological foundation for the "golden one-hour" intervention.

However, there are still three major challenges in this field: insufficient samples of home scenarios and nighttime in open-source data, resulting in a performance gap when the model is implemented; low-power technology needs to be broken through, and the battery life of existing devices is difficult to meet the long-term monitoring requirements; cross-modal time synchronization errors may affect the fusion accuracy.

Future research can be advanced in three aspects: First, jointly build a "full-scenario - full-time" standardized dataset with elderly care institutions and hospitals, including scarce samples such as those in the bathroom and at night. Second, introduce the Transformer architecture (such as ViT-Fusion) to model cross-modal long-distance dependencies through the self-attention mechanism. Third, develop an energy harvesting module (such as human kinetic energy harvesting) to extend the device's battery life to more than one month. These directions will promote the fall monitoring technology to move from the laboratory to large-scale applications, providing more reliable safety guarantees for the aging society.

References

[1] National Health Commission. (2023). White Paper on Elderly Health in China. Beijing: People's Medical Publishing House.
[2] World Health Organization. (2022). Global Report on Falls Prevention in Older Age. Geneva: WHO Press.
[3] Zhang, M., et al. (2021). Analysis of the limitations of visual fall detection in home environments. Acta Automatica Sinica, 47(8), 1765-1774.
[4] Chen, Z. Q., Liu, W., Zhao, Y. (2023). A survey of fall detection technologies for the elderly: from single sensor to multimodal fusion. IEEE Transactions on Instrumentation and Measurement, 72, 1-13.

DOI: 10.54254/2753-8818/2025.AD26456

[5] Wang, X. D., Liu, C. (2020). Application status of multi-sensor fusion technology in elderly fall monitoring. Sensors and Microsystems, 39(3), 1-4.
[6] Liu, Y., Zhou, M., Wu, H. (2023). Fall detection technology of millimeter-wave radar in complex home environments. Journal of Electronics & Information Technology, 45(7), 2013-2021.
[7] Zheng, X. W., Sun, P., Li, J. (2023). Feature extraction and correlation analysis of heart rate variability in elderly fall events. Chinese Journal of Biomedical Engineering, 42(3), 289-298.
[8] Huang, H., et al. (2022). Real-time fall detection system for the elderly based on edge computing. Journal of Computer Applications, 42(5), 1567-1574.
[9] Anguita, D., Ghio, A., Oneto, L., et al. (2013). A public domain dataset for human activity recognition using smartphones. 21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, 1-6.
[10] Fall Detection Dataset. (2020). Kaggle. https: //www. kaggle.com/datasets.
[11] National Bureau of Statistics. (2024). China Statistical Yearbook 2024. Beijing: China Statistics Press.
[12] Smith, J., Lee, K., Park, S. (2023). Nighttime fall detection: Challenges and dataset augmentation strategies. IEEE Journal of Biomedical and Health Informatics, 27(4), 1890-1898.
[13] Peking Union Medical College Hospital. (202 2). Report on the Evaluation of the Clinical Application Effect of the Multimodal Fall Monitoring System. Beijing: Peking Union Medical College Hospital.
[14] GDPR. (2016). Article 25: Data minimiza tion. https: //gdpr-info.eu/art-25-gdpr/.