Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 11
TPCEE 2025

Analysis of Non-Contact Fall Detection Technologies

Yanqi Huang

School of Flexible Electronics (Future Technologies), Nanjing Tech University, Nanjing, 21000, China

hyanqi11@gmail.com

Abstract. Against the backdrop of an aging society, the health of middle-aged and elderly individuals has become a growing concern for families. Falls are the leading cause of injury-related deaths, with a significant number of older adults suffering from disabilities or fatalities due to falls each year. Therefore, detecting falls is critically important. Traditional wearable sensors face challenges such as algorithmic limitations and wearability issues, making it difficult to meet the increasing demands of the elderly. Thus, research on non-contact fall detection technologies is essential. To explore various non-contact fall detection methods, this paper examines four types of technologies: those based on WiFi, computer vision, radar, and the multimodal fusion of WiFi and RGB. The paper is divided into three parts: the first introduces the theoretical foundations of each technology, briefly explaining their mechanisms and highlights; the second compares the advantages, disadvantages, and applicable experimental scenarios of each technology; and the third summarizes current technological limitations and offers prospects for future development. By analyzing the principles and characteristics of various technologies, this study presents suitable application scenarios and comparative strengths and weaknesses, providing a theoretical reference and practical guidance for future research on non-contact fall detection, thereby facilitating the optimization and implementation of these technologies.

Keywords: Fall Detection; Radar; YOLOv8; CSI.

1. Introduction

In recent years, with continuous social and economic development, population aging has become a major challenge for many countries. Falls in daily life pose a significant threat to the safety of the elderly. They often result in substantial medical expenses and increase the healthcare burden on society. Research indicates that the time taken to provide assistance largely determines the severity of physical harm from a fall, making early detection essential. Wearable sensor-based approaches were among the earliest proposed detection methods, but they face challenges such as weak adaptability to individual differences, insufficient feature extraction and model generalization, contradictions between real-time performance and computational load, and user experience barriers. Therefore, non-contact fall detection technology is essential for the elderly population. In recent years, non-contact fall detection technology has seen significant advancement, meeting the critical needs of an aging society while offering substantial economic and data value. This paper investigates various fall detection technologies, compares their strengths and weaknesses, and provides a theoretical reference for future research in non-contact fall detection.

2. Technical Analysis

2.1. WiFi (CSI)-Based Non-Contact Fall Detection Technology

2.1.1. CSI Signal Model.

With the widespread use of wireless networks and continuous advancements in WiFi, MIMO technology has improved WiFi reliability and throughput, while OFDM has enhanced spectral efficiency and reduced multipath effects. Based on MIMO and OFDM, WiFi provides Channel State Information (CSI) for each transmitter-receiver antenna pair. Specifically, OFDM converts high-speed signals into multiple orthogonal subcarriers for transmission to increase data rates. CSI characterizes the current channel state. Each pair of amplitude and phase information describes the subcarrier's status, reflecting environmental effects on the transmitted signal—including delay, amplitude attenuation, and phase shift. As Ma et al. point out in their review, CSI, as a fine-grained wireless sensing signal, is highly suitable for non-contact fall detection [1].

2.1.2. Deep Learning Models.

Deep learning, a subfield of machine learning, aims to enable computers to learn, understand, and process complex nonlinear data by mimicking the structure and function of the human brain's neural networks. The advancement of this technology has been driven by increased computational power and the availability of large datasets, factors that collectively make the training of deep neural networks possible. Unlike traditional machine learning algorithms, which often require manual feature extraction and domain expertise to convert raw data into recognizable formats, deep learning automatically discovers representations for detection or classification from raw data without costly human intervention. Among these, Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) models are widely applied. RNNs are neural network architectures designed for processing sequential data, primarily suited for data with temporal dependencies. While capable of handling sequences, RNNs face challenges with long sequences, such as the vanishing gradient problem—where gradients diminish during backpropagation, hindering the learning of long-range dependencies. Consequently, LSTM networks were developed to address these limitations. Long Short-Term Memory networks are a specialized autoregressive model of recurrent neural networks, achieved by incorporating a long-term memory unit and modifying the gate structure. LSTMs are employed to process the state information (CSI) extracted from Wi-Fi routers to identify fall behaviors. The core of LSTM lies in its gating mechanism, comprising three primary gates: the forget gate, the input gate, and the output gate. These gates control the inflow and outflow of information, enabling effective management of both long-term and short-term information. LSTM can efficiently learn and retain long-term dependencies in long sequences, effectively resolving the limitations inherent in RNNs.

2.1.3. Traditional Machine Learning (SVM).

Support Vector Machine (SVM) is a classic supervised learning model widely applied in classification and regression tasks, demonstrating exceptional performance, particularly in pattern recognition with small samples and high-dimensional spaces. Its core principle involves finding the optimal separating hyperplane to efficiently partition data into distinct categories.

2.1.4. Specific Research.

Jörg Schäfer and his team introduced a non-contact fall detection method utilizing Channel State Information (CSI) from WiFi signals. Their experiment consisted of three stages: In the first stage, the research team collected WiFi CSI in a corridor environment, recording 200 samples of five activities, including walking, and used wavelet analysis combined with SVM/LSTM for activity recognition. In the second stage, the researchers recorded 140 fall samples and employed a dual-SVM approach to distinguish real falls from other activities. In the final stage, the team recorded 1500 samples of 1-3 person actions in a laboratory setting and utilized LSTM for people counting, thereby validating the complete device-free CSI sensing pipeline. The study innovatively combined wavelet analysis and SVM for refined feature extraction, achieving, for the first time, multi-category activity recognition using CSI in a single MIMO system. By using only ordinary network cards and machine learning techniques, the system accomplished daily activity recognition, fall alarm triggering, and indoor people counting, offering a low-cost, non-intrusive solution for home health monitoring [2].

Yue Zihao proposed a non-contact, real-time, high-precision indoor fall detection method using commercial WiFi (CSI), addressing issues such as wearability inconvenience, privacy concerns, and high deployment costs in traditional solutions. For the first time, multi-sample rate fusion and attention mechanisms were introduced to CSI fall detection, resolving feature inconsistency caused by activity cycle variations and increasing multi-user AUC to 0.990. A wireless signal enhancement strategy via time-axis scaling (0.8–1.2×) was also proposed, expanding the dataset ninefold and improving accuracy by an average of 16%. This significantly mitigated overfitting and completed CSI data augmentation. A semi-supervised dataset was constructed, greatly improving data collection efficiency. The specific experimental process involved setting up one transmitter and one receiver (Intel 5300 network card, 1×3 MIMO) in a conference room, continuously collecting CSI data for nine activities, including falls at 1000 Hz rate. Subsequently, the raw CSI undergoes Butterworth 80Hz low-pass filtering, linear interpolation to compensate for packet loss, phase correction, and amplitude-phase stitching processing. Data was then segmented into samples using a sliding window (2000 points/2s, stride 500 points/0.5s) and divided into training (60%), validation (20%), and test (20%) sets. Offline training of the MS-LSTM network (incorporating multi-scale sampling, attention mechanisms, and temporal data augmentation) followed. Finally, a two-stage cascaded network is deployed online. The first stage performs high-confidence coarse screening of "non-fall" samples, while the second stage performs fine-grained discrimination of "fall" events. Accuracy, F1, and AUC metrics are tested under single-user, multi-user, new environment, and real-time scenarios [3].

2.2. Millimeter-Wave Radar-Based Non-Contact Fall Detection Technology

2.2.1. Frequency-Modulated Continuous Wave (FMCW).

FMCW radar continuously transmits signals with linearly varying frequency, using the frequency difference between echoed and transmitted waves for ranging and Doppler shift for velocity measurement, achieving high precision with low power. Distance resolution is determined by bandwidth, velocity resolution by observation time, and linearity and bandwidth are key design factors.

2.2.2. 4D Imaging Radar.

4D imaging radar is an advanced millimeter-wave sensor based on a 77GHz FMCW system combined with MIMO virtual aperture technology. Building upon the traditional three-dimensional information of "range, radial velocity, and azimuth angle" in conventional millimeter-wave radar, it further introduces the "elevation angle (height)" dimension. This creates four-dimensional space-velocity information, outputting high-density point clouds with sub-degree angular resolution. This enables imaging-level perception of target contours and motion states.

2.2.3. RDSNet Design.

RDSNet primarily consists of Convolutional Neural Network (CNN) and LSTM networks. It uses CNN's efficient feature extraction capability to process single Doppler heatmaps and LSTM's temporal correlation modeling to describe feature map sequences. Input data are range-Doppler heatmaps, which offer several advantages: high sensitivity for quickly and accurately capturing motion information, low data complexity for computational efficiency, and no imaging involved, better protecting the privacy of detected objects.

2.2.4. CNN.

CNN processes the temporal data collected by radar as "sequences," automatically extracting key features such as human body contours, joint displacements, and velocity changes to distinguish between four actions: standing, sitting, lying down, and falling. Compared to traditional threshold methods or wearable sensors, CNN maintains over 95% accuracy even under complex backgrounds, varying lighting conditions, and partial occlusion. It also eliminates the need for users to wear devices, balancing privacy and comfort.

2.2.5. Histogram of Oriented Gradients (HOG).

HOG is a classic manual image feature extraction method first proposed by Dalal and Triggs in 2005. It characterizes object edges and contours by statistically capturing the gradient direction distribution in local image regions, demonstrating strong robustness to lighting, pose, and partial occlusion.

2.2.6. Specific Research.

Sejong Ahn et al. jointly proposed a non-contact fall detection system for the elderly based on 4D imaging radar and a CNN model. The study performed lightweight preprocessing on point clouds through zero-padding, k-means clustering, and sorting, and pioneered the integration of real-time Unity 3D virtual avatar visualization into the radar monitoring system. It achieved 98.66% accuracy in classifying three postures (standing, sitting, lying) and 95% accuracy in fall detection, with real-time alerts synchronized on web and Unity platforms. The system works through clothing, furniture, and low-light conditions. The specific process involved collecting point clouds in a 7m × 7m space with a radar at 50ms intervals, normalizing each frame of 500 points, applying zero-padding, k-means dimensionality reduction, and lexicographical sorting; then using CNN for posture output, combined with zmax height and velocity thresholds via rule-based fall detection. Results are pushed via UDP to Unity virtual avatars and React management dashboards for real-time display and emergency notification triggering [4].

Yuan Zhi'an, Zhou Xiaoyu et al. proposed a millimeter-wave radar-based RDSNet method for human fall detection. They first collected 8-frame range-Doppler heatmap sequences using a 77 GHz FMCW radar. Features are extracted frame-by-frame through two layers of CNN (4-channel and 8-channel 3×3 convolutions + 3×3 pooling). A 208-dimensional LSTM layer then models temporal relationships. Finally, a ReLU fully-connected classifier outputs real-time results for six motion categories: fall/wave/stand/ stationary/walking/turning. Single-frame processing latency is <50ms. The study innovatively integrated CNN and LSTM into millimeter-wave radar range-Doppler heatmap sequences for the first time, building an end-to-end lightweight RDSNet.Systematic experiments confirm the optimal structure of 2-layer CNN + 1-layer LSTM. Under privacy-preserving and light-independent conditions, it achieves 93.33% detection accuracy for falls. The accuracy for falls alone reached 96.67%, with a false alarm rate of only 3.33%. The system can run in real-time on edge devices like Raspberry Pi, balancing high accuracy, low latency, and strong generalization capabilities [5].

Yun Seop Yu et al. proposed the "HOG-LSTM-Attention" fall detection system for elderly individuals living alone. First, a 60GHz FMCW millimeter-wave radar collects 10 seconds of x-y scatter plots and Doppler-range plots (100 frames/event). Then, HOG extracts gradient feature vectors, which are fed into an LSTM network with self-attention mechanisms for temporal modeling. Finally, a fully connected layer outputs five results: "Fall/Walking/Lying Down/ sitting/stationary." The optimal ratio for the attention head and channel numbers is determined to be 2:64. This study innovatively conducts the first systematic comparison of four LSTM architectures (Raw-LSTM, Raw-LSTM-Attention, HOG-LSTM, HOG-LSTM-Attention) in millimeter-wave radar fall detection tasks, demonstrating that the combination of HOG feature engineering and self-attention significantly enhances performance. It also provides quantitative guidance for attention hyperparameter tuning, filling a gap in prior comparative studies that relied solely on FMCW millimeter-wave x-y scattering +Doppler imaging. Operating non-contact, light-independent, and privacy-preserving while maintaining high accuracy, the system runs in real-time on standard PC+GPU configurations, demonstrating strong deployment potential [6].

2.3. Computer Vision-Based Non-Contact Fall Detection Technology

2.3.1. YOLOv8.

YOLOv8 is the latest iteration in the YOLO series released by Ultralytics, capable of real-time object detection, segmentation, pose estimation, tracking, and classification. It offers a balance of accuracy and speed and can be widely deployed on heterogeneous computing platforms from edge to cloud.

2.3.2. HRNET.

HRNET is a human pose estimation and detection network that consistently maintains high-resolution features. Proposed by Microsoft Research in 2019, it offers high accuracy, strong robustness against interference, broad applicability, and easy integration compared to traditional networks.

2.3.3. Specific Research.

Shi Huan, Wang Xiaopeng et al. proposed a top-down fall detection algorithm using "enhanced HRNet + YOLO.". First, an improved YOLOv8 (ShuffleNetV2 lightweight backbone + Shuffle-Attention hybrid attention) locates human bounding boxes in camera footage. Then, the BAM-HRNet high-resolution network extracts 17 skeletal keypoints. It then calculates in real-time the human body's height-to-width ratio, center-of-mass descent velocity, and torso-to-ground angle change rate. A fall is detected when all three metrics simultaneously exceed thresholds ($A<1$, $V>350$px/s, $\Omega>75°$), triggering a voice-based secondary confirmation. This study innovatively integrates ShuffleNetV2 with Shuffle-Attention into the YOLOv8 neck detection subnetwork, achieving lightweight processing and occlusion robustness. Simultaneously, Bottleneck Attention Modules (BAM) are introduced at the end of each HRNet branch to fuse multi-scale channel-spatial attention, suppressing redundant information from high resolution. Furthermore, the “Keypoint Random Erasure” (KRE/I-KRE) data augmentation strategy significantly enhances keypoint localization accuracy in occlusion scenarios. This technology offers high precision and robustness, enabling direct deployment on edge devices [7].

Liu Dong proposed the “YOLOv8-DGD” human fall detection system, constructing a 7,782-image “standing/falling” dataset partitioned 6:2:2 for training/validation/testing. The study then integrated DCNv2 deformable convolutions, Dynamic Head attention, and the self-developed GCA global coordinate attention into the YOLOv8 backbone and detection head to achieve end-to-end real-time object detection. By calculating the height-to-width ratio of the human bounding box and the y-axis velocity of the center of gravity across consecutive frames, a fall is detected when the ratio exceeds 1 and velocity increases abruptly, thereby distinguishing falls from static behaviors like lying down. This study innovatively introduces DCNv2 (deformable convolution with weight learning) into YOLOv8 for the first time, enabling adaptive sampling of convolution kernels during pose deformation. It proposes GCA attention, which integrates global channel information on top of CA to achieve dual “global-coordinate” feature enhancement. It combines Dynamic Head's “scale-space-task” triple attention with DCNv2 to build a lightweight, high-precision detection head, suppressing false positives from lying down through motion feature suppression in consecutive frames. The model runs in real-time on an RTX4090 with only 7M parameters, delivering high accuracy, strong occlusion robustness, and deployment flexibility [8].

Nizar Zaghden et al. proposed a novel fall detection model, R-CNN-GAN+YOLOv8, which innovatively integrates YOLOv8, Faster R-CNN, and Generative Adversarial Networks (GAN)for the first time. This fusion enhances detection accuracy and robustness while introducing multiple attention mechanisms (channel attention, spatial attention, squeeze-and-excitation) to strengthen feature extraction in critical regions. GANs are employed to generate synthetic fall data, alleviating the issue of insufficient training data. Experimental results demonstrate the model's outstanding performance on the DiverseFALL10500 and CAUCAFall datasets, achieving mAP scores of 0.9507 and 0.996, respectively, with recall rates reaching 0.929 and 0.9993. With an inference time of only 43ms per frame, significantly outperforming existing mainstream methods and making it suitable for real-time monitoring scenarios [9].

2.4. WiFi (CSI) and RGB Image Multimodal Fusion-Based Fall Detection

Fujia Zhou et al. proposed a “CSI+RGB bimodal human activity recognition framework.” First, two computers equipped with Intel 5300 network cards collected 30 subcarrier 2560-point CSI data, while simultaneously recording 200 frames of RGB video at 20 fps using a camera. CSI data was fed into a CRCF residual network, while RGB data entered a SlowFast-3D network to generate feature/decision tensors. Activity labels for 13 categories were then output via early fusion (TFN, Auto-Fusion, concatenation) or late fusion (mean voting, lightweight DNN), validated using the CRIHAD public dataset. This approach innovatively achieved deep integration of wireless CSI and visual RGB for human activity recognition (HAR) for the first time. The authors designed a two-stage "modality-aware fusion" general framework that supports plug-and-play compatibility with any single-modality network. They also proposed a synchronous collection protocol and a sliding window slicing strategy to resolve inconsistencies between CSI and visual frame rates. Additionally, the study verified the stable performance of CSI in no-light and large-angle scenarios. The system requires only five lightweight models for real-time operation, offering significant advantages in privacy protection, low cost, and ease of deployment [10].

3. Technology Comparison

WiFi CSI technology offers the advantage of requiring no additional hardware, directly leveraging existing WiFi, low cost, and support for data augmentation and semi-supervised learning, making it suitable for large-scale deployment. It is ideal for high-privacy and low-cost scenarios but is sensitive to environmental changes, requires transfer learning, suffers from multipath interference, has difficult feature extraction, and performs poorly in multi-person scenarios. Millimeter-wave radar technology is unaffected by light, offers good privacy protection, high recognition accuracy, and low latency, making it suitable for nighttime or private settings. However, it requires external radar deployment, has moderate cost, and is not adept at simultaneous multi-person recognition. Computer vision technology offers strong real-time performance, requires no wearable devices, and is suitable for open spaces and scenarios demanding high real-time performance. Its drawbacks include dependence on light, poor performance at night or in backlighting, privacy concerns with cameras, and strict line-of-sight requirements without occlusion. CSI+RGB fusion technology benefits from wireless and visual complementarity, offering the highest recognition accuracy and strongest robustness, adaptable to complex environments. It is suitable for high-precision scenarios but involves complex deployment, requires synchronous CSI and image collection, has high cost, and carries privacy leakage risks.

4. Conclusion

With the acceleration of societal aging, falls have become a primary cause of physical injury among the elderly in modern society, thereby placing higher demands on the automatic detection of indoor falls. Traditional wearable devices for fall detection fail to meet the needs of the elderly population, making the research and application of non-contact fall detection technologies critically important. This paper investigates several prominent non-contact fall detection technologies, examining their fundamental principles, basic experimental workflows, and comparative advantages/disadvantages to explore which techniques suit different scenarios. Although current fall detection technologies have made significant progress in accuracy, real-time performance, and deployment convenience, they still face multidimensional challenges, primarily in three areas. First, existing deep learning methods heavily rely on large-scale labeled data, leading to difficulties in cross-environment transfer. Future research may explore unsupervised domain adaptation and federated learning frameworks. By leveraging spatio-temporal consistency in unlabeled data, these approaches could enable rapid model adaptation to new environments while reducing manual annotation costs. Concurrently, techniques like knowledge distillation and neural network architecture search could compress complex models for edge devices, meeting low-power, low-latency real-time detection requirements and facilitating algorithm deployment on platforms such as Raspberry Pi and embedded routers. Second, single-modal approaches have inherent limitations: visual methods suffer from lighting and occlusion constraints, wireless signal methods are susceptible to multipath interference, and radar methods lack fine-grained semantic information. Future work should develop a three-tier collaborative architecture integrating radar, WiFi, and vision. Third, existing datasets primarily simulate falls by young adults, differing from real-world elderly fall patterns (e.g., slow slides, leaning falls). Future efforts should establish open datasets encompassing diverse age groups and pathological conditions (e.g., Parkinson's disease, sarcopenia), while integrating human biomechanical models to enhance the system's ability to recognize atypical falls.

References

[1] Ma, Yongsen, Zhou, Gang, Wang, Shuangquan. WiFi Sensing with Channel State Information: A Survey. ACM COMPUTING SURVEYS, 2019, 52 (03).

[2] Neena Damodaran, Jorg Schafer. Device Free Human Activity Recognition using WiFi Channel State Information. 2019: 1069–1074.

[3] Yue Zihao. Research on Indoor Fall Detection Technology Based on Wireless Signals. University of Electronic Science and Technology of China, 2022.

[4] Ahn, S., Choi, M., Lee, J., Kim, J., Chung, S. Non-Contact Fall Detection System Using 4D Imaging Radar for Elderly Safety Based on a CNN Model. Sensors, 2025, 25: 3452.

[5] Yuan Zhi'an, Zhou Xiaoyu, Liu Xinpu, et al. Millimeter-Wave Radar Human Fall Detection Method Based on RDSNet. Journal of Radars, 2021, 10 (04): 656–664.

[6] Yu, Y.S., Wie, S., Lee, H., Lee, J., Kim, N.H. Long Short-Term Memory-Based Fall Detection by Frequency-Modulated Continuous Wave Millimeter-Wave Radar Sensor for Seniors Living Alone. Appl. Sci., 2025, 15: 8381.

[7] Shi, H., Wang, X., Shi, J. Fall Detection Algorithm Using Enhanced HRNet Combined with YOLO. Sensors, 2025, 25: 4128.

[8] Liu Dong. Research on Human Fall Detection Based on YOLOv8. Wuhan Polytechnic University, 2024.

[9] Nizar Zaghden, Emad Ibrahim, Mukaram Safaldin, et al. Integrating Attention Mechanisms in YOLOv8 for Improved Fall Detection Performance. Computers, Materials & Continua, 2025, 83 (04): 1117-1147.

[10] Fujia Zhou, Guangxu Zhu, Xiaoyang Li, Hang Li, Qingjiang Shi. Towards Pervasive Sensing: A multimodal approach via CSI and RGB image modalities fusion. In Proceedings of the 3rd ACM MobiCom Workshop on Integrated Sensing and Communications Systems (ISACom '23). Association for Computing Machinery, New York, NY, USA, 2023: 25–30.