

# P2MFDS: A Privacy-Preserving Multimodal Fall Detection System for Elderly People in Bathroom Environments

Haitian Wang<sup>1,2</sup>, Yiren Wang<sup>2</sup>, Xinyu Wang<sup>2</sup>, Yumeng Miao<sup>2</sup>, Yuliang Zhang<sup>2</sup>, Yu Zhang<sup>1,\*</sup>, Atif Mansoor<sup>2</sup>

**Abstract**—By 2050, people aged 65 and over are projected to make up 16% of the global population. As aging is closely associated with increased fall risk, particularly in wet and confined environments such as bathrooms where over 80% of falls occur. Although recent research has increasingly focused on non-intrusive, privacy-preserving approaches that do not rely on wearable devices or video-based monitoring, these efforts have not fully overcome the limitations of existing unimodal systems (e.g., WiFi-, infrared-, or mmWave-based), which are prone to reduced accuracy in complex environments. These limitations stem from fundamental constraints in unimodal sensing, including system bias and environmental interference, such as multipath fading in WiFi-based systems and drastic temperature changes in infrared-based methods. To address these challenges, we propose a Privacy-Preserving Multimodal Fall Detection System for Elderly People in Bathroom Environments. First, we develop a sensor evaluation framework to select and fuse millimeter-wave radar with 3D vibration sensing, and use it to construct and preprocess a large-scale, privacy-preserving multimodal dataset in real bathroom settings, which will be released upon publication. Second, we introduce P2MFDS, a dual-stream network combining a CNN-BiLSTM-Attention branch for radar motion dynamics with a multi-scale CNN-SEBlock-Self-Attention branch for vibration impact detection. By uniting macro- and micro-scale features, P2MFDS delivers significant gains in accuracy and recall over state-of-the-art approaches. Code and pretrained models are available at [P2MFDS Github Repository](#).

## I. INTRODUCTION

By 2050, people aged 65 and above are projected to constitute 16% of the global population, up from 10% in 2022 [1]. A growing number of elderly people prefer to live independently, yet this increases their vulnerability to medical emergencies such as falls, which remain one of the leading causes of injury-related morbidity and mortality among elderly adults [2] [3]. Bathrooms, with their slippery surfaces, are particularly hazardous; studies indicate that over 80% of elderly falls occur in these environments [4]. Given that falls can cause severe injuries or deaths, a real-time and accurate solution to detect falls in the bathroom is essential for the elderly [5] [6].

Existing fall detection solutions for elderly people can be categorized into three primary approaches. The first involves in-home caregivers who provide continuous monitoring and immediate assistance in the event of a fall [7] [8]. Although effective, this solution is cost-prohibitive and lacks scalability [9]. The second approach utilizes wearable devices, such as

smart watches and electrocardiogram (ECG) monitors. These devices continuously track physiological parameters such as heart rate, blood pressure, and vibration [10]. However, these devices face adoption challenges, particularly among elderly users who may find them uncomfortable, forget to wear them, or struggle to use them in wet environments such as bathrooms [11]. The third approach integrates smart home technologies, including video-based and audio-based fall detection systems [12] [13]. While these systems can achieve high detection accuracy, their deployment in private spaces, such as bathrooms, raises significant privacy concerns, limiting user acceptance. Given these limitations, there is a critical need for a non-intrusive, real-time fall detection system that ensures both accuracy and privacy in bathroom environments [14].

Recent research has explored non-intrusive fall detection methods, focusing on privacy-preserving solutions that do not require wearable devices or video-based monitoring [15] [16]. Palipana [12] et al. proposed FallDeFi, a WiFi-based system that detects falls by analyzing signal disturbances. While effective in controlled environments, its accuracy degrades in complex spaces like bathrooms due to multipath interference and environmental noise. Similarly, Zigel et al. [17] developed a vibration- and sound-based detection system that addresses privacy concerns but is highly sensitive to background noise, leading to false positives from non-fall events such as dropped objects. Akash et al. [6] explored millimeter-wave (mmWave) radar for fall detection, utilizing its high-resolution motion tracking capabilities.

However, its effectiveness depends on precise calibration and can be influenced by bathroom wall tiles and flooring [18] [19]. These approaches predominantly rely on unimodal sensor data, which undergoes preprocessing, feature extraction, and classification to detect falls [20] [21]. However, unimodal systems face fundamental limitations: the performance is approaching theoretical bounds, leaving minimal room for improvement through feature engineering alone [22] [23]. Moreover, single-sensor data is highly susceptible to system bias and environmental interference. For instance, infrared sensors are affected by temperature variations, while vibration sensors can misinterpret background noise or minor disturbances as falls, leading to increased false alarm rates. These challenges underscore the need for a multimodal, sensor-fusion approach to enhance accuracy and robustness in real-world applications [24] [25].

To solve this issue, we propose a Privacy-Preserving Multimodal Fall Detection System for Elderly People in Bathroom Environments. This system comprises two parts

\*Corresponding Author: zhangyu@nwpu.edu.cn

<sup>1</sup>School of Computer Science, Northwestern Polytechnical University, No. 1 Dongxiang Road, Chang'an District, Xi'an 710129, China

<sup>2</sup>Department of Computer Science and Software Engineering, The University of Western Australia, 35 Stirling Hwy, Crawley, WA 6009, Australia



Fig. 1. System overview of the P2MFDS pipeline. First, a multidimensional evaluation framework scores candidate sensing modalities (e.g., mmWave radar, vibration sensor, RGB and thermal cameras) to select the optimal combination. Next, multimodal data are collected in a realistic bathroom setup. The P2MFDS Network then fuses long-term motion features extracted by a 1D CNN–BiLSTM–Attention stream with short-term impact signatures captured by a Multi-Scale CNN–SEBlock–Self-Attention stream. Finally, the fused representation is classified into fall or non-fall events.

TABLE I

EVALUATION OF 14 COMMON SENSING MODALITIES BASED ON EIGHT KEY FACTORS (E.G., ENERGY EFFICIENCY, RECALL RATE, AND DEPLOYABILITY), CALCULATION THEIR OVERALL SUITABILITY FOR PRIVACY-PRESERVING FALL DETECTION IN BATHROOM SETTINGS.

Method	Target Relevance	Cost Effectiveness	Recall	Energy Efficiency	Deployability	Non-Intrusive	Availability	Usability
Wi-Fi Detection	3	3	4	2 (AC/DC)	3	5	3	61
Infrared Detection	3	2	3	3 (AC/DC)	4	5	1	61
Bluetooth Detection	3	4	4	2 (AC/DC)	3	5	2	64
<b>MMW Detection</b>	<b>3</b>	<b>3</b>	<b>5</b>	<b>4 (AC/DC)</b>	<b>4</b>	<b>5</b>	<b>3</b>	<b>73</b>
Ultrasonic Detection	3	3	4	2 (AC/DC)	4	4	3	62
Audio Detection	3	4	4	3 (AC/DC)	4	2	4	65
Thermal Imaging	3	1	5	5 (AC)	3	4	1	62
<b>3D vibration</b>	<b>3</b>	<b>4</b>	<b>4</b>	<b>4 (AC)</b>	<b>3</b>	<b>5</b>	<b>4</b>	<b>68</b>
Light Sensors	1	5	4	1 (AC)	1	5	5	47
Door Sensors	1	5	4	2 (DC)	5	5	4	62
Furniture Sensors	1	5	2	1 (DC)	5	5	3	51
Weight Sensors	1	5	3	1 (AC/DC)	5	5	4	57
Electricity Usage	1	2	1	1 (AC)	2	5	4	63
Water Usage	1	2	3	1 (AC)	3	5	4	54

(as shown in Fig. 1)). The first part is a systematic sensor evaluation framework that guides the selection and fusion of millimeter-wave radar with 3D vibration sensing, enabling the collection and preprocessing (smoothing, filtering, fusion) of a large-scale, privacy-preserving multimodal dataset in real bathroom settings; this dataset will be released upon publication. The second part is P2MFDS, a dual-stream network combining a CNN–BiLSTM–Attention branch for capturing long-term motion dynamics from radar point clouds and a multi-scale CNN–SEBlock–Self-Attention branch for detecting short-term impact signatures in vibration data. By uniting macro- and micro-scale features, P2MFDS achieves 95.0% accuracy, 87.9% recall, and a 91.3% F1-score, substantially outperforming state-of-the-art unimodal approaches. The complete implementation and pretrained models are available at [P2MFDS Github Repository](#).

## II. OFFLINE MULTIMODAL SENSORS SELECTION

To develop a robust non-intrusive fall detection system suitable for bathroom environments, we propose a comprehensive Multimodal Sensor Evaluation Framework (as shown in Table II). This framework systematically assesses various sensing modalities based on their feasibility, performance, and deployability in real-world conditions. The evaluation criteria include accuracy, non-intrusiveness, energy efficiency, computational complexity, deployability, recall

TABLE II  
PROPOSED EVALUATION FRAMEWORK FOR SELECTING OPTIMAL SENSING MODALITIES IN PRIVACY-PRESERVING FALL DETECTION.

Selection Criteria	Description	Rating System	Weight
Target Relevance	The relevance for fall down detection	1–3 scale	3
Non-Intrusiveness	Level of privacy preservation	1–5 scale	2
Energy Efficiency	Power consumption requirements (AC/DC)	1–5 scale	3
Comp Complexity	Processing demands on the system	1–5 scale	2
Deployability	Ease of integration	1–5 scale	2
Recall Rate	The ability to detect all fall events	1–5 scale	3
Availability	Operational reliability in real-world conditions	1–5 scale	2
Cost-Effectiveness	Economic feasibility of implementation	1–5 scale	1

rate, availability, and cost-effectiveness.

Each sensing modality is assigned a weighted score for the eight evaluation dimensions, contributing to an overall usability score ( $OU$ ) using the following equation:  $OU_m = \sum_{i=1}^N w_i \times S_{m,i}$ , where  $OU_m$  represents the overall usability of modality  $m$ ,  $S_{m,i}$  denotes the score for modality  $m$  in the  $i$ -th dimension, and  $w_i$  is the weight assigned to the  $i$ -th evaluation criterion. Higher  $OU_m$  values indicate superior suitability for bathroom fall detection.

Using the Multimodal Sensor Evaluation Framework, we

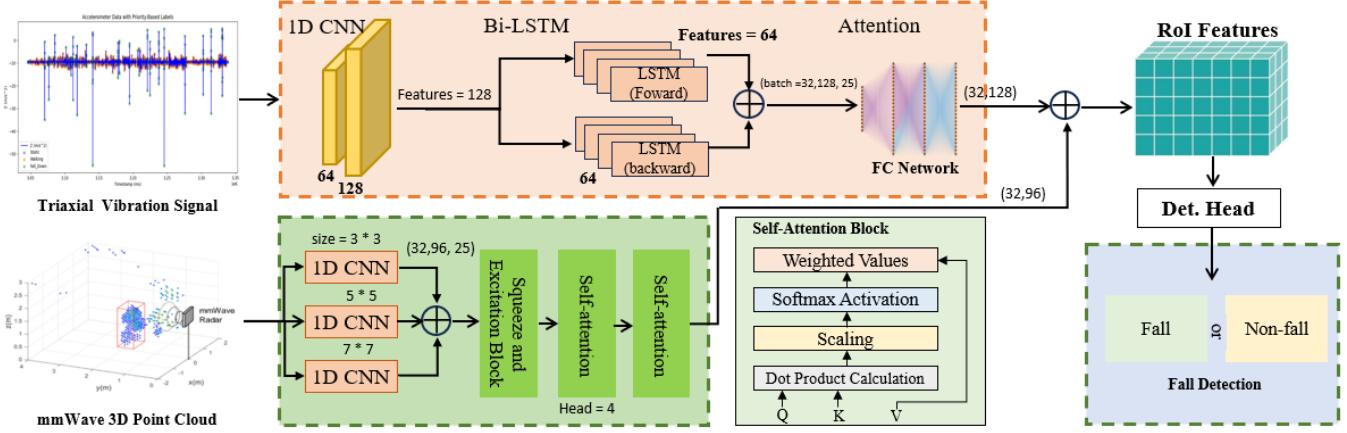


Fig. 2. Overview of the P2MFDS Network architecture. The upper pipeline employs a 1D CNN-BiLSTM-Attention sequence to extract global motion features from mmWave 3D point clouds, while the lower pipeline utilizes Multi-Scale CNN, SEBlock, and Self-Attention to capture localized vibration signals. Both streams are fused to enable robust, privacy-preserving fall detection in bathroom environments.

evaluate 14 widely used sensing modalities across multiple categories. The evaluation results (as shown in Table I) indicate that mmWave radar and a three-axis vibration achieve the highest usability scores, making them the most suitable choices for privacy-preserving fall detection in bathroom environments. The mmWave radar offers high-precision motion tracking that remains effective in varying lighting conditions while maintaining privacy, whereas the three-axis vibration provides real-time vibration sensing with low power consumption, ensuring reliable detection. The combination of these two modalities enables a complementary approach, where the mmWave radar captures overall motion patterns, and the vibration sensor detects localized ground vibrations. This multimodal integration enhances detection accuracy, minimizes false positives, and improves robustness against environmental noise, making the system well-suited for real-world deployment.

### III. PRIVACY-PRESERVING MULTIMODAL FALL DETECTION SYSTEM NETWORK

Privacy-Preserving Multimodal Fall Detection System (P2MFDS) network fuses mmWave radar and triaxial vibration streams to exploit their complementary sensing characteristics. After targeted low-pass filtering to remove high-frequency noise while preserving fall-related signatures, the radar branch captures macro-scale motion attributes (e.g., velocity, distance, signal energy) and the vibration branch extracts micro-scale impact signatures. Two independent processing streams—one employing a CNN-BiLSTM-Attention pipeline for radar point clouds and the other using a Multi-Scale CNN-SEBlock-Self-Attention pipeline for vibration data—then distill salient features. Finally, these embeddings are concatenated and fed into a lightweight classifier for fall vs. non-fall prediction (Fig. 2).

#### A. Signal Preprocessing

Prior to feature extraction, each modality undergoes targeted low-pass filtering to suppress high-frequency noise while preserving fall-related signatures. For vibration data, a Moving Average Filter (MAF) smooths transient disturbance

bances by computing

$$y_{\text{vib}}(t) = \frac{1}{N} \sum_{i=0}^{N-1} x_{\text{vib}}(t-i), \quad (1)$$

where  $N$  is chosen to balance noise reduction against responsiveness. For radar data, a first-order Exponential Low-Pass Filter (ELPF) adapts to rapid motion changes:

$$y_{\text{radar}}(t) = \alpha x_{\text{radar}}(t) + (1 - \alpha) y_{\text{radar}}(t-1), \quad (2)$$

with  $\alpha \in (0, 1)$  controlling the trade-off between stability and sensitivity. This dual-filtering strategy mitigates environmental artifacts (e.g., electrical interference, multipath reflections) and enhances the robustness of subsequent feature extraction.

#### B. 3D Vibration-Based Impact Feature Extraction

To capture fine-grained temporal dynamics and transient impacts related to fall events, the triaxial vibration signal initially undergoes feature extraction through a 1D Convolutional Neural Network (1D CNN) and a Bidirectional Long Short-Term Memory (Bi-LSTM) network. Specifically, the input vibration data is first convolved by multiple CNN kernels to produce intermediate temporal-spatial embeddings, which are subsequently fed into a Bi-LSTM composed of forward and backward sequences. The Bi-LSTM effectively integrates long-term dependencies from historical and future contexts, yielding comprehensive temporal embeddings as follows:

$$\vec{h}_t = \text{LSTM}_{\text{fw}}(X_t, \vec{h}_{t-1}), \quad \overleftarrow{h}_t = \text{LSTM}_{\text{bw}}(X_t, \overleftarrow{h}_{t+1}) \quad (3)$$

where  $\vec{h}_t$  and  $\overleftarrow{h}_t$  represent forward and backward hidden states at time step  $t$ , respectively. These embeddings are concatenated to form the final hidden state  $h_t = [\vec{h}_t, \overleftarrow{h}_t]$ , effectively capturing bidirectional temporal information.

To highlight critical temporal segments indicative of potential falls, an attention mechanism is subsequently applied. Given the Bi-LSTM output  $H = \{h_1, h_2, \dots, h_T\}$ , the attention weights  $\alpha_t$  are calculated by:

$$\alpha_t = \frac{\exp(h_t W_a)}{\sum_{i=1}^T \exp(h_i W_a)}, \quad F_{\text{vibration}} = \sum_{t=1}^T \alpha_t h_t \quad (4)$$

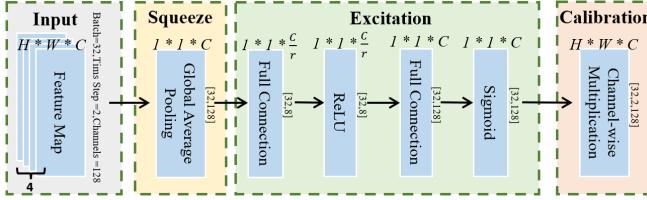


Fig. 3. Architecture of the SE Block. Channel-wise statistics are extracted via global average pooling, passed through two fully connected layers with a bottleneck structure, and used to generate attention weights via Sigmoid activation. The input feature map is then recalibrated through channel-wise multiplication.

where  $W_a$  denotes learnable parameters, and  $F_{\text{vibration}}$  represents the refined impact feature embedding emphasizing the most informative segments of vibration data.

### C. mmWave Radar-Based Motion Feature Extraction

The mmWave radar stream focuses on capturing macro-scale motion patterns within the 3D point cloud data through a multi-scale convolutional strategy enhanced by channel-wise and temporal attention mechanisms. Initially, radar signals are processed by parallel 1D CNN branches with varying kernel sizes (3, 5, and 7) to capture diverse motion scales. These multi-scale features are merged into a unified embedding for subsequent recalibration.

To further enhance critical channel-level characteristics and suppress redundant features, we incorporate a Squeeze-and-Excitation (SE) block [26]. The SE block contributes to improved generalization by reducing the risk of overfitting through dynamic channel reweighting, which is particularly beneficial under data-limited training conditions. The SE block first applies global average pooling (GAP) to aggregate global spatial information across each channel, forming a channel descriptor  $z_c$ :

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W x_c(i, j) \quad (5)$$

where  $x_c(i, j)$  denotes the feature map value at position  $(i, j)$  of channel  $c$  (as shown in Fig. 3). Subsequently, channel-wise recalibration weights are learned through fully connected layers with a bottleneck architecture:

$$s_c = \sigma(W_2 \text{ReLU}(W_1 z_c)), \quad \hat{x}_c = s_c \cdot x_c \quad (6)$$

where  $W_1$  and  $W_2$  represent learnable weights of fully connected layers,  $\sigma$  denotes the sigmoid activation function, and  $\hat{x}_c$  is the recalibrated channel-wise feature map. This recalibration adaptively emphasizes significant motion-related features across channels.

Finally, a self-attention mechanism utilizes these recalibrated features to enhance temporal sensitivity and motion dynamics (as shown in Fig. 2). Query ( $Q$ ), Key ( $K$ ), and Value ( $V$ ) matrices are constructed, and the self-attention output is computed as:

$$F_{\text{radar}} = \text{softmax} \left( \frac{QK^T}{\sqrt{d_k}} \right) V \quad (7)$$

Here,  $d_k$  is the dimensionality of keys, and the resulting embedding  $F_{\text{radar}}$  captures the most salient temporal motion characteristics relevant to accurate fall detection.

### D. Multimodal Feature Fusion and Classification

After the mmWave 3D point cloud and vibration streams independently extract motion and impact features, respectively, these two embeddings are concatenated to form a joint representation:

$$F_{\text{fusion}} = \sigma(W_f[F_{\text{radar}}, F_{\text{acc}}] + b_f), \quad (8)$$

where  $W_f$  and  $b_f$  are learnable parameters, and  $\sigma$  is the activation function. This fused feature map (labeled “RoI Features” in Fig. 2) is then fed into a lightweight *detection head* (Det Head)—a final classification module comprising fully connected layers and a confidence function. The Det Head transforms the high-level fused representation into discrete predictions (fall vs. non-fall) [27] [28].

## IV. EXPERIMENT

This section presents the experimental design and setup used to evaluate the proposed P2MFDS Network in real-world bathroom scenarios. We first describe the testing environment and sensor placement, followed by the methodology for collecting multimodal data under various fall and non-fall conditions. Finally, we present the metrics and benchmarking procedures used to validate our system’s effectiveness and robustness.

### A. Experiment Setup

To ensure the reliability, reproducibility, and real-world applicability of our proposed fall detection system, we conducted experiments in a controlled environment designed to replicate a typical residential bathroom (Fig. 4 shows the Top-down floor plan). The experimental space measure 2.5m (L)  $\times$  1.1m (W)  $\times$  2.2m (H), forming a 6.05m<sup>2</sup> enclosed testing area. The bathroom features ceramic tiled walls, along with a glass partition and a shower curtain that resemble a typical residential bathroom. The floor consists of anti-slip ceramic tiles, with standard bathroom utilities including a showerhead, drain, and storage shelves containing toiletries such as soap and shampoo (as shown in Fig. 5).

To achieve optimal sensor coverage while minimizing occlusion and environmental interference, two sensor nodes

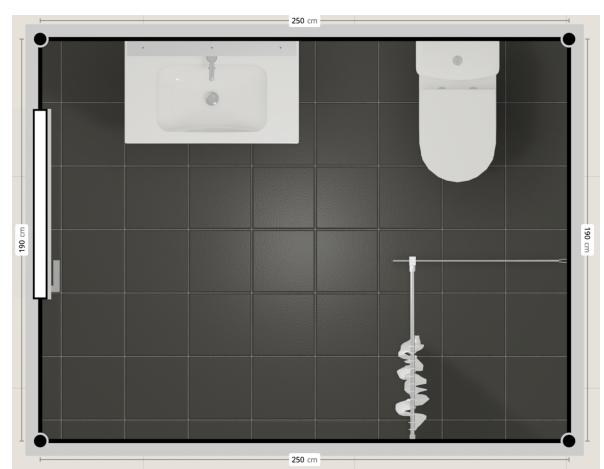


Fig. 4. Top-down floor plan of the experimental bathroom (2.50 m  $\times$  1.90 m), showing the mmWave radar and vibration sensor mounting positions, annotated coverage areas, and key dimensions.

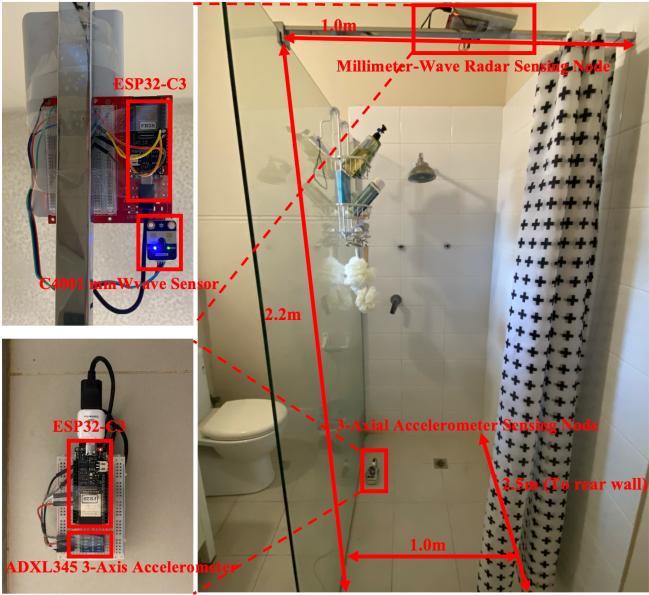


Fig. 5. Illustration of the controlled bathroom environment and sensor placements. The mmWave radar node is mounted at 2.2 m for broad coverage, while the ADXL345 vibration sensor is installed near the shower area to capture ground impacts. This setup closely replicates typical residential bathroom conditions for realistic fall detection experiments.

were strategically deployed within the test area. The triaxial vibration sensor node comprising an ADXL345 3-axis vibration, an ESP32-C3 microcontroller, and a rechargeable battery was securely placed 0.5m from the wall beneath the showerhead, a location frequently occupied during bathing activities. A waterproof case was used to protect the sensor from moisture exposure. The mmWave radar node consisting of a C4001 mmWave radar sensor, an ESP32-C3 MCU, and a battery unit was mounted on a metal beam at a height of 2.2m, ensuring unobstructed coverage of the entire bathroom space.

### B. Experimental Design

The evaluation protocol comprises nine representative bathroom scenarios (Fig. 6) chosen to span a spectrum of non-fall and fall events, namely: (1) empty room (no motion), (2) light object drop (e.g., soap), (3) heavy object drop (e.g., mop), (4) normal walking, (5) bent-posture walking, (6) wall-supported walking, (7) static standing, (8) squatting, and (9) intentional falls. These scenarios were selected to challenge the system with variations in motion intensity, impact signature, and postural change, thereby testing P2MFDS's ability to distinguish true falls from everyday disturbances and fall-like activities.

Each scenario was conducted in multiple trials by all participants, resulting in a dataset comprising 20 minutes per scenario, with a total duration of 3 hours. The mmWave radar operated at a sampling rate of 10 Hz, generating 3D point clouds at 10 frames per second, capturing motion dynamics over time. Simultaneously, the triaxial vibration recorded 3D vibration data at 100 Hz, ensuring high temporal resolution for detecting impact forces associated with falls. The dataset included over 120,000 vibration data points and 18,000 mmWave frames, allowing for robust multimodal



Fig. 6. Overview of the eight experimental scenarios simulating typical bathroom conditions: (a) empty room (no motion), (b) lightweight object drop, (c) heavy object drop, (d) normal walking, (e) bent-posture walking, (f) assisted walking with wall support, (g) static standing or squatting, and (h) fall. These scenarios collectively capture a wide range of non-fall and fall-like events for robust real-world detection analysis.

feature extraction.

This experimental design ensures the system can effectively distinguish genuine falls from environmental disturbances, such as object drops, while accounting for variations in human movement patterns.

### C. Experimental Participants

To ensure safety, ethical integrity, and the reproducibility of our findings, all experiments were conducted with the voluntary participation of 48 healthy adults (24 males and 24 females), aged between 20 and 30 years. All participants provided informed consent for both data collection and subsequent publication. The experimental procedures, including intentional fall simulations, cautious walking, assisted walking, static standing, and squatting, were carried out under strict safety protocols within a controlled environment. Given the inherent risks associated with fall-related activities, elderly people were not involved in the study. Instead, all fall-like behaviors and typical elderly movement patterns were carefully emulated by the participants. The entire experiment was conducted under the supervision of Dr. Atif Mansoor and was approved by the Ethics and Academic Committee of the University of Western Australia (UWA). The collected multimodal dataset will be made publicly available for research purposes, supporting open access and reproducibility in the

TABLE III

P2MFDS NETWORK PERFORMANCE ACROSS EIGHT REPRESENTATIVE BATHROOM SCENARIOS, REPORTING BOTH NON-FALL AND FALL METRICS (ACCURACY, RECALL, PRECISION, AND F1 SCORE).

Model	Scenario	Non-Fall Metrics (%)				Fall Metrics (%)			
		Accuracy	Recall	Precision	F1 Score	Accuracy	Recall	Precision	F1 Score
P2MFDS Network	Empty Bathroom	96.0	96.1	96.1	96.1	96.0	95.9	95.9	95.9
	Light Object Drop	95.0	94.3	96.2	95.2	95.0	95.7	93.8	94.7
	Heavy Object Drop	90.4	91.4	89.8	90.6	90.4	89.5	91.1	90.1
	Normal Walking	94.7	95.7	93.8	94.7	94.7	93.8	95.7	94.8
	Bent Posture Walk	89.2	81.8	88.2	84.9	89.2	93.6	89.7	91.6
	Wall-Supported Walk	85.3	89.1	84.5	86.7	85.3	80.9	86.4	83.5
	Static Standing	92.6	94.6	92.9	93.7	92.6	90.0	92.3	91.1
	Squatting	97.0	96.2	98.1	97.1	97.0	97.9	95.8	96.8
	Total	95.0	98.0	95.1	96.5	95.0	94.6	87.8	91.3

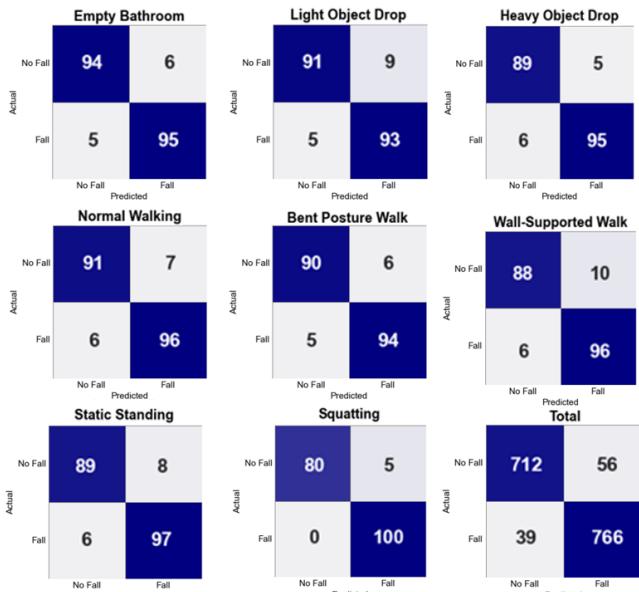


Fig. 7. Confusion matrices of the P2MFDS Network across eight experimental scenarios and the overall total. Each matrix shows classification counts for fall and non-fall events, illustrating high true positive and true negative rates with minimal misclassification.

field of fall detection.

#### D. Experimental Results

The P2MFDS Network achieves an overall accuracy of 95.0%, precision of 94.6%, recall of 87.8%, and F1-score of 91.3% across eight representative bathroom scenarios (Table III). Scenario-wise F1-scores exceed 90% in six activities, peaking at 97.9% for normal walking and 97.3% for squatting. The confusion matrices in Fig. 7 confirm a balanced distribution of true positives and true negatives, with squatting yielding zero false negatives and static standing maintaining a 93.9% F1-score despite potential environmental noise.

By fusing macro-scale motion features from the mmWave radar with micro-impact signatures from 3D vibration sensing, P2MFDS effectively suppresses false alarms in non-fall events (e.g., 97.5% F1 in an empty bathroom) while retaining high fall detection recall under complex disturbances (e.g., 94.6% recall in light object drops). This robust performance across both quantitative metrics and confusion-matrix analyses underscores the method's resilience to noise and its

practical viability for privacy-preserving fall monitoring in confined bathroom environments.

To validate the effectiveness of our proposed Privacy-Preserving Multimodal Fall Detection System (P2MFDS) Network, we conducted a comprehensive performance comparison with 16 state-of-the-art fall detection models, as summarized in Table IV. These methods span a diverse range of sensing modalities—including mmWave radar, UWB radar, Wi-Fi CSI, vibration sensors, and multimodal fusion—and model architectures such as CNNs, LSTMs, attention mechanisms, and SVMs.

#### E. Comparison

Among radar-only methods, Rezaei et al. [29], He et al. [33], and Yao et al. [31] employed traditional CNN and RBF-based models, achieving moderate recall (82.8%–88.9%) but relatively lower precision, particularly in cluttered environments. Transformer-based [36] and Bi-LSTM+CNN [39] architectures showed improved performance due to better temporal modeling, yet still relied on a single sensing modality.

Several UWB radar-based solutions [30], [37], [38] demonstrated high recall (up to 98.8%) but suffered from limited precision or overall accuracy, likely due to lower spatial resolution and vulnerability to occlusion. Similarly, vision-assisted fusion [40] and Wi-Fi CSI-based fusion [42] achieved high recall but may raise privacy concerns in sensitive environments like bathrooms.

Compared to these approaches, P2MFDS achieves competitive accuracy (95.0%), superior precision (94.6%), and robust recall (87.8%) through effective multimodal integration of mmWave radar and 3D vibration sensing. This fusion enables the model to simultaneously capture macro-scale motion dynamics and micro-scale impact signals, enhancing robustness against noise and confounding activities. Notably, our method outperforms other multimodal fusion networks such as Sun et al. [41] and Zhang et al. [44], both in terms of accuracy and precision. These results underscore the superiority of the P2MFDS architecture in balancing detection performance with privacy preservation in challenging, confined environments such as residential bathrooms.

#### F. Ablation Study

To assess the impact of individual components in the Privacy-Preserving Multimodal Fall Detection System

TABLE IV  
COMPARISON OF OUR PROPOSED P2MFDS NETWORK AGAINST EXISTING FALL DETECTION METHODS, COMPARING MAJOR PERFORMANCE INDICATORS.

Author	Model	Detection Method	Accuracy (%)	Precision (%)	Recall (%)
Rezaei et al. [29]	CNN	mmWave Radar	88.7	69.9	85.5
Maitre et al. [30]	CNN-LSTM	UWB Radars	77.0	86.0	89.0
Yao et al. [31]	Range-vertical angle map classifier	mmWave Radar	86.7	84.7	88.9
Li et al. [32]	CNN-LSTM	mmWave Radar	<b>96.8</b>	84.6	89.6
He et al. [33]	RBF neural network	Radar Sensor	85.9	88.3	82.8
Clemente et al. [34]	multiclass SVM	Smart Seismic Sensing	91.2	74.2	73.7
Hanif et al. [35]	SVM	CW Doppler Radar	89.0	88.0	87.0
Wang et al. [36]	Transformer	mmWave Radar	94.5	94.2	86.7
Sadreazami et al. [37]	CNN	UWB Radar	82.9	92.8	81.2
Swarubini et al. [38]	ResNet-50	UWB Radar	67.2	50.6	<b>98.8</b>
Yang et al. [39]	Bi-LSTM+CNN	mmWave Radar	93.1	93.8	91.3
Dai et al. [40]	Improved YOLOX	mmWave Radar + Vision	93.4	94.6	95.6
Sun et al. [41]	Attention-GRU	Multimodal Sensor Fusion	94.2	92.4	83.8
Meng et al. [42]	CNN-LSTM-Attention	Wi-Fi CSI + Accelerometer	93.1	94.3	91.2
Alkhaldi et al. [43]	ResNet-18	mmWave Radar	93.0	91.4	90.2
Zhang et al. [44]	Att-CNN-LSTM	FMCW Radar	93.5	90.2	86.7
<b>Ours</b>	<b>P2MFDS Network</b>	<b>mmWave Radar + vibration</b>	<b>95.0</b>	<b>94.6</b>	<b>87.8</b>

TABLE V

ABLATION STUDY SHOWCASING THE IMPACT OF COMBINING DIFFERENT MODEL COMPONENTS (CNN, LSTM, AND ATTENTION) AND SENSOR MODALITIES (RADAR AND VIBRATION) ON FALL DETECTION PERFORMANCE. METRICS ARE REPORTED IN PERCENTAGE (%).

Attention	CNN+LSTM	CNN	Sensor Modalities		Metrics (%)				
			vibration	Radar	Accuracy	Recall	Precision	F1 Score	AUC-ROC
✓	✓		✓		92.3	80.6	83.4	81.9	86.2
✓	✓			✓	87.2	88.8	64.3	74.5	75.8
✓	✓		✓	✓	98.3	76.7	96.4	85.4	97.5
✓		✓	✓	✓	84.5	51.2	70.0	59.1	72.1
✓		✓		✓	80.2	84.8	66.3	74.4	68.5
✓		✓	✓	✓	93.4	83.1	91.4	87.1	87.4
✓	✓	✓	✓		95.9	91.4	93.7	92.5	88.3
✓	✓	✓		✓	87.2	89.3	77.7	83.1	78.3
✓	✓	✓	✓	✓	95.0	94.6	87.8	91.1	96.4

(P2MFDS) Network, we conducted an ablation study, as summarized in Table V. The results demonstrate that multimodal sensor fusion significantly improves fall detection performance. Using only the vibration, the CNN+LSTM model achieves 92.3% accuracy, while mmWave radar alone yields 87.2% accuracy. When both modalities are combined, accuracy increases to 98.3%, confirming the advantages of multimodal integration.

In terms of model architecture, the full P2MFDS Network (CNN+LSTM+Attention) achieves the highest performance, outperforming CNN- or LSTM-only configurations. The results highlight the complementary strengths of CNN for spatial feature extraction, LSTM for temporal modeling, and Attention for adaptive weighting, leading to robust and privacy-preserving fall detection.

## V. CONCLUSION

Our experimental results confirm that the Privacy-Preserving Multimodal Fall Detection System (P2MFDS) Network effectively fuses mmWave 3D point cloud and vibration data, achieving high accuracy under various activities such as object drops, squatting, and bent-posture walking. This fusion addresses limitations found in unimodal systems—particularly noise susceptibility and poor recall in cluttered bathroom settings. Furthermore, the attention mechanisms within the CNN-BiLSTM-Attention and Multi-Scale

CNN-SEBlock-Self-Attention modules enhance the model's adaptability to different motion patterns while preserving user privacy.

Despite these promising outcomes, environmental factors—such as temperature fluctuations or partial occlusions—may still pose challenges. Additionally, large-scale, long-term deployments in real homes with elderly people are needed to further validate robustness and usability. Future work will focus on exploring adaptive calibration, extending the dataset across diverse bathroom layouts, and improving noise reduction to strengthen overall fall detection performance.

## VI. ACKNOWLEDGEMENT

This work was supported by the UWA PMC CSSE Engineering Research Project Funding under the “Non-Invasive Fall Detection System for Elderly in Bathroom” project. We also thank all participants for their contributions to the experimental evaluation.

## REFERENCES

- [1] D. Leong, K. Teo, S. Rangarajan, P. Lopez-Jaramillo, A. Avezum Jr, A. Orlandini *et al.*, “World population prospects 2019. department of economic and social affairs population dynamics,” *World*, vol. 73, no. 7, 2018.
- [2] V. Strini, R. Schiavolin, and A. Prendin, “Fall risk assessment scales: A systematic literature review,” *Nursing Reports*, vol. 11, no. 2, pp. 430–443, 2021.

- [3] X. Wang, J. Ellul, and G. Azzopardi, "Elderly fall detection systems: A literature survey," *Frontiers in Robotics and AI*, vol. 7, p. 71, 2020.
- [4] A. Singh, S. U. Rehman, S. Yongchareon, and P. H. J. Chong, "Sensor technologies for fall detection systems: A review," *IEEE Sensors Journal*, vol. 20, no. 13, pp. 6889–6919, 2020.
- [5] A. Yazar, F. Erden, and A. E. Cetin, "Multi-sensor ambient assisted living system for fall detection," in *Proceedings of the IEEE international conference on acoustics, speech, and signal processing (ICASSP'14)*, 2014, pp. 1–3.
- [6] M. S. R. Akash, M. N. Shahria, M. A. Morshed, S. S. Rodsee, N. Hannan, and M. H. Imam, "Elderly patient monitoring and fall detection using mmwave fmcw radar system," in *2023 26th International Conference on Computer and Information Technology (ICCIT)*. IEEE, 2023, pp. 1–6.
- [7] W. Chen, Z. Jiang, H. Guo, and X. Ni, "Fall detection based on key points of human-skeleton using openpose," *Symmetry*, vol. 12, no. 5, p. 744, 2020.
- [8] S. Y. Y. Tun, S. Madanian, and F. Mirza, "Internet of things (iot) applications for elderly care: a reflective review," *Aging clinical and experimental research*, vol. 33, pp. 855–867, 2021.
- [9] G. P. Cardoso, D. G. Damaceno, M. F. S. Alarcon, and M. J. S. Marin, "Care needs of the elderly who live alone: an intersectoral perception," *Rev. Rene*, vol. 21, pp. 44 395–44 395, 2020.
- [10] J. L. Guiñón, E. Ortega, J. García-Antón, and V. Pérez-Herranz, "Moving average and savitzki-golay smoothing filters using mathcad," *Papers ICEE*, vol. 2007, pp. 1–4, 2007.
- [11] T. Perumal *et al.*, "A review on fall detection systems in bathrooms: challenges and opportunities," *Multimedia Tools and Applications*, pp. 1–29, 2024.
- [12] S. Palipana, D. Rojas, P. Agrawal, and D. Pesch, "Falldefi: Ubiquitous fall detection using commodity wi-fi devices," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 4, pp. 1–25, 2018.
- [13] F. Riquelme, C. Espinoza, T. Rodenas, J.-G. Minonzio, and C. Tarasacco, "ehomeseniors dataset: An infrared thermal sensor dataset for automatic fall detection research," *Sensors*, vol. 19, no. 20, p. 4565, 2019.
- [14] J. Wagner, P. Mazurek, and R. Z. Morawski, *Non-invasive monitoring of elderly persons: Systems based on impulse-radar sensors and depth sensors*. Springer, 2022.
- [15] A. Mavaddat, S. H. M. Armaki, and A. R. Erfanian, "Millimeter-wave energy harvesting using 4 x 4 microstrip patch antenna array," *IEEE Antennas and wireless propagation letters*, vol. 14, pp. 515–518, 2014.
- [16] M. Gerstmair, A. Melzer, A. Onic, R. Stuhlberger, and M. Huemer, "Highly efficient environment for fmcw radar phase noise simulations in if domain," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 65, no. 5, pp. 582–586, 2018.
- [17] Y. Zigel, D. Litvak, and I. Gannot, "A method for automatic fall detection of elderly people using floor vibrations and sound—proof of concept on human mimicking doll falls," *IEEE transactions on biomedical engineering*, vol. 56, no. 12, pp. 2858–2867, 2009.
- [18] C. Neipp, A. Hernndez, J. Rodes, A. Mrquez, T. Belndez, and A. Belndez, "An analysis of the classical doppler effect," *European journal of physics*, vol. 24, no. 5, p. 497, 2003.
- [19] A. Taha, "Intelligent ensemble learning approach for phishing website detection based on weighted soft voting," *Mathematics*, vol. 9, no. 21, p. 2799, 2021.
- [20] Y. Chen, R. Du, K. Luo, and Y. Xiao, "Fall detection system based on real-time pose estimation and svm," in *2021 IEEE 2nd international conference on big data, artificial intelligence and internet of things engineering (ICBAIE)*. IEEE, 2021, pp. 990–993.
- [21] H. A. Hashim, S. L. Mohammed, and S. K. Gharghan, "Accurate fall detection for patients with parkinson's disease based on a data event algorithm and wireless sensor nodes," *Measurement*, vol. 156, p. 107573, 2020.
- [22] M. K. Appeadu and B. Bordoni, "Falls and fall prevention in the elderly," in *StatPearls [Internet]*. StatPearls Publishing, 2023.
- [23] J. Too, A. R. Abdullah, T. T. Zawawi, N. M. Saad, and H. Musa, "Classification of emg signal based on time domain and frequency domain features," *International Journal of Human and Technology Interaction (IJHaTI)*, vol. 1, no. 1, pp. 25–30, 2017.
- [24] S. R. Khan, S. K. Pavuluri, G. Cummins, and M. P. Desmulliez, "Wireless power transfer techniques for implantable medical devices: A review," *Sensors*, vol. 20, no. 12, p. 3487, 2020.
- [25] J. Kim, W. Lee, and S. H. Lee, "A systematic review of the guidelines and delphi study for the multifactorial fall risk assessment of community-dwelling elderly," *International journal of environmental research and public health*, vol. 17, no. 17, p. 6097, 2020.
- [26] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132–7141.
- [27] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1d convolutional neural networks and applications: A survey," *Mechanical systems and signal processing*, vol. 151, p. 107398, 2021.
- [28] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: concepts, cnn architectures, challenges, applications, future directions," *Journal of big Data*, vol. 8, pp. 1–74, 2021.
- [29] A. Rezaei, A. Mascheroni, M. C. Stevens, R. Argha, M. Papandrea, A. Puiatti, and N. H. Lovell, "Unobtrusive human fall detection system using mmwave radar and data driven methods," *IEEE Sensors Journal*, vol. 23, no. 7, pp. 7968–7976, 2023.
- [30] J. Maitre, K. Bouchard, and S. Gaboury, "Fall detection with uwb radars and cnn-lstm architecture," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 4, pp. 1273–1283, 2021.
- [31] Y. Yao, C. Liu, H. Zhang, B. Yan, P. Jian, P. Wang, L. Du, X. Chen, B. Han, and Z. Fang, "Fall detection system using millimeter-wave radar based on neural network and information fusion," *IEEE Internet of Things Journal*, vol. 9, no. 21, pp. 21 038–21 050, 2022.
- [32] W. Li, D. Zhang, Y. Li, Z. Wu, J. Chen, D. Zhang, Y. Hu, Q. Sun, and Y. Chen, "Real-time fall detection using mmwave radar," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 16–20.
- [33] C. He, S. Liu, G. Zhong, H. Wu, L. Cheng, G. Yan, and Y. Wen, "A noncontact fall detection method for bedside application with a mems infrared sensor and a radar sensor," *IEEE Internet of Things Journal*, vol. 10, no. 14, pp. 12 577–12 589, 2023.
- [34] J. Clemente, F. Li, M. Valero, and W. Song, "Smart seismic sensing for indoor fall detection, location, and notification," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 2, pp. 524–532, 2020.
- [35] K. Hanifi and M. E. Karsligil, "Elderly fall detection with vital signs monitoring using cw doppler radar," *IEEE Sensors Journal*, vol. 21, no. 15, pp. 16 969–16 978, 2021.
- [36] J. Wang, L.-H. Lin, M. Wang, and W. Xu, "Non-contact fall detection system based on mmwave radar and transformer," *IEEE Sensors Journal*, vol. 23, no. 20, pp. 24 002–24 009, 2023.
- [37] H. Sadreazami, M. Bolic, and S. Rajan, "Compressed domain contactless fall incident detection using uwb radar signals," *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 1, pp. 244–256, 2022.
- [38] P. Swarubini and N. Ganapathy, "Radar-based elderly fall detection using spwvd and resnet network," *Current Directions in Biomedical Engineering*, vol. 10, no. 4, pp. 498–501, 2024.
- [39] B. Yang, L. Guo, X. Zhang, J. Yang, Y. Ren, and X. Li, "An intelligent fall detection method based on millimeter-wave radar and hybrid deep learning model," *Sensors*, vol. 22, no. 21, p. 8176, 2022.
- [40] J. Dai, Z. Ji, Z. Xie, C. Wang, and J. Wu, "Multimodal fusion for indoor fall detection using millimeter-wave radar and vision sensor," *Expert Systems with Applications*, vol. 231, p. 120770, 2023.
- [41] Y. Sun, G. Li, Z. Zhu, C. Huang, X. Hou, and Y. He, "Attention-gru based multimodal fusion for elderly fall detection," *Sensors*, vol. 21, no. 2, pp. 1–16, 2021.
- [42] X. Meng, Y. Liu, B. Zhou, P. Wang, J. Song, and D. Zhang, "Multimodal sensor fusion for elderly fall detection using wi-fi csi and accelerometer data," *IEEE Internet of Things Journal*, vol. 9, no. 21, pp. 21 247–21 257, 2022.
- [43] H. Alkhaldi and M. A. A. Faruque, "Fall detection using radar and deep learning for elderly safety," in *Proc. IEEE Sensors Applications Symposium (SAS)*. IEEE, 2022, pp. 1–6.
- [44] Y. Zhang, L. Wang, M. Li, and X. Liu, "A privacy-preserving radar-based system for elderly fall detection using attention-enhanced cnn-lstm architecture," in *Proceedings of the 45th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2023, pp. 1234–1238.