Internet of Things 33 (2025) 101682

Contents lists available at ScienceDirect

Internet of Things

journal homepage: www.elsevier.com/locate/iot

Research article

RadSpecFusion: Dynamic attention weighting for multi-radar human activity recognition☆

Ayesha Ibrahim [a,*,1], Muhammad Zakir Khan [c], Muhammad Imran [c], Hadi Larijani [b], Qammer H. Abbasi [c,d], Muhammad Usman [a,1]

[a] *Department of Electrical and Electronic Engineering, Glasgow Caledonian University, G4 0BA, UK*
[b] *Department of Cyber Security and Networks, Glasgow Caledonian University, G4 0BA, UK*
[c] *James Watt School of Engineering, University of Glasgow, University of Glasgow, G12 8QQ, UK*
[d] *Distinguished Visiting Research Fellow, Abu Dhabi University's College of Engineering, Abu Dhabi*

ARTICLE INFO

Dataset link: https://github.com/ZakirAfridi/

Keywords:
Human activity recognition
Multi-modal fusion
Attention mechanisms
Cross-frequency transfer learning

ABSTRACT

This paper presents RadSpecFusion, a novel dynamic attention-based fusion architecture for multi-radar human activity recognition (HAR). Our method learns activity-specific importance weights for each radar modality (24 GHz, 77 GHz, and Xethru sensors). Unlike existing concatenation or averaging approaches, our method dynamically adapts radar contributions based on motion characteristics. This addresses cross-frequency generalization challenges, where transfer learning methods achieve only 11%–34% accuracy. Using the CI4R dataset with spectrograms from 11 activities, our approach achieves 99.21% accuracy, representing a 15.8% improvement over existing fusion methods (83.4%). This demonstrates that different radar frequencies capture complementary information about human motion. Ablation studies show that while the three-radar system optimizes performance, dual-radar combinations achieve comparable accuracy (24GHz+77GHz: 96.1%, 24GHz+Xethru: 95.8%, 77GHz+Xethru: 97.2%), enabling flexible deployment for resource-constrained applications. The attention mechanism reveals interpretable patterns: 77 GHz radar receives higher weights for fine movements (superior Doppler resolution), while 24 GHz dominates gross body movements (better range resolution). The system maintains 71.4% accuracy at 10 dB SNR, demonstrating environmental robustness. This research establishes a new paradigm for multimodal radar fusion, moving from cross-frequency transfer learning to adaptive fusion with implications for healthcare monitoring, smart environments, and security applications.

1. Introduction

Human activity recognition (HAR) using radio frequency (RF) sensors has emerged as a transformative technology across healthcare monitoring, smart environments, and security systems [1,2]. Unlike optical-based systems, radar sensors operate effectively in challenging conditions including poor lighting, through obstacles, and in privacy-sensitive environments without requiring wearable devices [3,4]. Radar-based activity recognition relies on microDoppler signatures: distinctive frequency modulations caused by body

☆ This work has been supported by the CHEDDAR: Communications Hub for Empowering Distributed ClouD Computing Applications and Research funded by the UK EPSRC under grant numbers EP/Y037421/1 and EP/X040518/1.

* Corresponding author.
E-mail addresses: sent2ayesha@gmail.com (A. Ibrahim), muhammad.usman@gcu.ac.uk (M. Usman).
[1] IEEE Member.

https://doi.org/10.1016/j.iot.2025.101682
Received 14 April 2025; Received in revised form 1 June 2025; Accepted 24 June 2025
Available online 9 July 2025

movement patterns that enable activity classification [5]. While these signatures exist across all radar systems, they manifest differently depending on operating frequency. Traditional radar-based activity recognition systems predominantly employ single-radar architectures across various frequency bands: low-frequency (S-band, L-band), medium-frequency (24 GHz), or high-frequency (60 GHz, 77 GHz) [6]. Each frequency band presents inherent trade-offs: lower frequencies provide better penetration but reduced resolution, while higher frequencies offer enhanced resolution but greater susceptibility to environmental interference [7]. This fundamental limitation restricts the capabilities of single-radar systems in complex real-world environments. We selected 24 GHz, 77 GHz, and Xethru (7–10 GHz) frequencies to leverage complementary characteristics. The 24 GHz (λ = 12.5 mm) provides optimal range resolution (0.2 m) for gross body displacement tracking. The 77 GHz radar (λ = 3.9 mm) offers superior Doppler resolution (0.1 m/s) for fine-grained limb movements [8,9]. The ultra-wideband Xethru system employs impulse-based sensing with unique temporal characteristics for approach/recession detection and obstacle penetration [7]. This tri-frequency approach spans multiple electromagnetic regions, ensuring comprehensive motion coverage from macro-displacement to micro-physiological movements [10].

Deep learning (DL) methods have shown remarkable success in radar-based activity recognition [11,12], with convolutional neural networks (CNNs) widely applied to radar spectrograms [13] and recurrent neural networks (RNNs) effectively modeling temporal aspects of radar data [14]. However, most approaches focus exclusively on single-radar systems. The limited existing multi-radar studies typically employ simplistic fusion methods such as feature concatenation [15] or averaging [16], without considering the variable importance of different radar modalities across activities. Although multi-modal fusion has been extensively explored in other sensing domains [17,18], effective fusion of multiple radar modalities presents unique challenges. Wang et al. [19] attempted early fusion of features from different radar types but encountered issues with feature incompatibility and dimensionality misalignment. Chen et al. [20] identified key limitations in late fusion methods, particularly their failure to capture inter-modal relationships, which impairs complex motion recognition. They also highlighted how radar data sparsity restricts effective utilization of multimodal information. These limitations in current fusion approaches highlight the need for more sophisticated methods that can dynamically determine the relevance of different modalities, making attention mechanisms a promising direction for radar fusion.

Attention mechanisms have transformed various machine learning fields by enabling models to focus on the most relevant parts of input data [21]. In multi-modal learning, attention has been effectively used to determine the importance of different modalities dynamically [22]. However, the application of attention mechanisms specifically to multi-radar fusion for HAR remains largely unexplored, presenting an opportunity to address current fusion limitations through adaptive radar input weighting. The *CI4R* (activity recognition dataset) provides an opportunity to investigate multi-modal radar fusion, containing synchronized data from three different radar types (24 GHz, 77 GHz, and Xethru) for eleven human activities. Previous approaches attempted to address radar frequency limitations through cross-frequency transfer learning [7]. Our preliminary experiments confirmed severe performance degradation (11%–34% accuracy) when models trained on one radar frequency were tested on another. This fundamental limitation in cross-frequency generalization underscores that different radar frequencies capture complementary rather than redundant information about human motion. Rather than attempting to overcome these differences through complex transfer learning, our multi-radar fusion approach directly leverages this complementarity by integrating information from multiple radar modalities simultaneously.

In this paper, we propose *RadSpecFusion*, a novel attention-based fusion architecture for multi-radar activity recognition that dynamically weighs the contribution of each radar modality based on input characteristics. Our approach leverages separate encoder networks for each radar type to extract frequency-specific features, followed by an attention mechanism that assigns importance weights to each modality before fusion. This adaptive weighting allows the model to focus on the most informative radar type for each activity, overcoming the limitations of both single-radar systems and traditional fusion methods. Our experimental results demonstrate perfect classification accuracy (99.21%) across all activity classes, significantly outperforming single-radar configurations and validating the effectiveness of our attention-based fusion strategy. The main contributions of this work are:

1. A novel attention-based fusion architecture (RadSpecFusion) that dynamically weighs multiple radar modalities, achieving 99.21% classification accuracy across 11 human activities.
2. Ablation studies show dual-radar combinations can achieve near-equivalent performance (95.8–97.2%) to the three-radar system, enabling deployment flexibility.
3. Empirical evidence that different radar frequencies capture complementary aspects of human motion, with attention weights revealing activity-specific optimal sensing modalities.
4. Demonstration of superior robustness to environmental noise (maintaining 71.4% accuracy at 10 dB SNR) compared to single-radar and cross-frequency transfer learning approaches.
5. Feature visualization analysis confirming the discriminative power of fused representations, explaining the high-performance classification results.

The rest of this paper is organized as follows: Section 2 reviews related work in radar-based activity recognition and multi-modal fusion. Section 3 presents our proposed methodology, including the attention-based fusion architecture. Section 4 describes the experimental setup, dataset, and preprocessing methods. Section 5 discusses the results and analysis. Section 6 provides discussion and implications of our findings. Finally, Section 7 concludes the paper and suggests directions for future research.

2. Related work

The field of radar-based HAR has evolved significantly over the past decade. This section reviews relevant literature on radar sensing for human motion analysis, approaches for radar data, and multi-modal fusion techniques.

2.1. Radar sensing for human activity analysis

Radar-based activity recognition offers significant advantages over vision-based systems due to its privacy-preserving nature, non-contact detection capabilities, and robustness to environmental conditions such as poor lighting or occlusion [23]. The micro-Doppler effect, a frequency modulation caused by moving parts of an object, serves as the foundation for distinguishing human activities [5], with techniques such as the short-time-average/long-time-average (STA/LTA) algorithm enhancing classification robustness [24]. Taylor and Khan et al. [3,25] demonstrated radar sensing's effectiveness in classifying blind elderly activities through micro-Doppler signatures, while Jokanovic et al. [14] showcased ultra-wideband impulse radar's capacity to capture subtle respiratory movements. Huang et al. [26] established 24 GHz radar's efficacy in detecting gross body movements, while millimeter wave radar (77 GHz) has proven particularly valuable in elderly care for real-time monitoring and fall detection.

Recent advances in radar processing have significantly improved human activity recognition performance. Wu et al. [27] achieved 95.031% accuracy by integrating Range-time and Doppler-time maps. Fan et al. [28] addressed overfitting in limited sample scenarios using channel-DN4 with mmWave FMCW radar, reaching 87.53% accuracy. Vishwakarma et al. [10] enhanced classification accuracy and interpretability across multiple frequency bands (10 GHz, 24 GHz, and 77 GHz) using attention-augmented architectures. Our approach contributes to this field by employing discrete wavelet transform (DWT) decomposition and Sobel filtering to improve classification accuracy and reduce overfitting. These collective advancements establish radar-based sensing as a viable alternative to vision-based systems for human motion analysis, addressing key challenges in feature extraction and real-world applicability. The evolution toward Healthcare 5.0 frameworks emphasizes human-centric IoT systems where radar-based sensing provides privacy-preserving monitoring capabilities [29]. Similarly, behavioral analytics in smart infrastructure applications demonstrate the broader applicability of activity recognition systems beyond traditional healthcare domains [30].

2.2. Deep learning for radar data

Recent advances in DL have significantly improved radar-based activity recognition. CNNs have become the predominant approach for analyzing radar spectrograms, demonstrating superior performance over traditional methods. Seyfioglu et al. [11] applied deep learning to micro-Doppler features for gait recognition, while Yang et al. [12] explored open-set human activity recognition. Triani et al. [31] employed established CNN architectures like VGG19, and Gao et al. [9] achieved 96.9% accuracy using multi-scale residual networks.

The integration of CNNs with specialized radar signal processing enhances performance, as demonstrated by Ayaz et al. [32], who achieved 96.30% accuracy combining MobileNetV2 with STFT preprocessing. For temporal modeling, Jokanovic [14] employed recurrent neural networks for fall detection, while Hazra [13] implemented 3D CNNs with triplet loss for gesture recognition. Geng et al. [33] reviewed applications of CNNs, LSTMs, and DNNs for radar waveform design, target recognition, and clutter suppression, highlighting challenges including limited training data. Miazek et al. [34] emphasized how transfer learning and data fusion enhance radar applications for activity recognition and fall detection. For specialized applications, Liang et al. [35] developed a lightweight YOLOv4-tiny variant for dim target detection, while Ghani et al. [33] achieved 95.6% accuracy in vehicle and pedestrian classification using radar-based sensor fusion. Karlsson [36] demonstrated how neural networks can achieve comparable accuracy to traditional radar processing approaches with significantly reduced computational demands.

2.3. Multi-modal fusion approaches

Multi-modal fusion approaches have shown significant promise across various sensing domains, though their application to radar-based activity recognition remains relatively limited. In wearable sensor contexts, fusion models combining skeletal and inertial data have achieved accuracy rates exceeding 99% in HAR tasks [37], with similar improvements demonstrated in other sensor fusion systems [38]. Within the radar domain, early approaches employed simple methods such as averaging for fusing multiple RF signals [16] or weighted feature combination for RGB-D data [15]. More sophisticated techniques have emerged recently, including multi-domain feature attention fusion networks (MFAFN) that integrate time-Doppler and time-range maps (97.58% accuracy) [39], and methods combining multiple radar representations (95.8% accuracy) [40].

Advanced research has introduced multi-frequency multi-scale deformable convolutional networks (MFMSDC) for nuanced feature extraction [41] and approaches combining time and frequency domain features [42]. Attention mechanisms, which have transformed natural language processing [21] and computer vision, have begun to impact radar signal processing, enhancing feature extraction in low SNR environments [43] and efficiently decoupling Doppler and temporal features. While Rahman et al. [22] demonstrated the effectiveness of integrating multi-modal information in large pretrained transformers, the application of attention mechanisms specifically for multi-radar fusion remains largely unexplored. This research gap exists despite evidence that different radar frequencies capture inherently different aspects of human movement that cannot be easily generalized across frequencies. Previous approaches using synthetic data generation with GANs [7] still treat different radar frequencies as separate domains rather than complementary information sources. This limitation motivates our novel fusion-based approach that leverages multiple radar modalities simultaneously instead of attempting to generalize across them. Traditional fusion methods principal component analysis (PCA), weighted voting) show limited effectiveness due to inability to capture complex inter-modal relationships in micro-Doppler signatures. This motivates our fusion-based approach that leverages multiple radar modalities simultaneously, addressing the research gap in activity-specific modality weighting for radar-based HAR systems.

Table 1
Comparison with previous methods on CI4R dataset.

Method	Accuracy (%)	Precision	Recall	F1-Score	Key approach
Gurbuz et al. [7]	∼70*	0.68	0.70	0.69	Cross-frequency training with GANs
Single-radar CNN (24 GHz)	38.9	0.40	0.39	0.39	Single modality classification
Single-radar CNN (77 GHz)	47.6	0.48	0.48	0.47	Single modality classification
Single-radar CNN (Xethru)	29.4	0.31	0.29	0.30	Single modality classification
Feature concatenation	85.7	0.86	0.86	0.86	Early fusion of radar features
Decision-level fusion	89.3	0.89	0.89	0.89	Late fusion of radar decisions
Our method	**99.21**	**0.99**	**0.992**	**0.991**	Attention-based multi-radar fusion

2.4. Comparative analysis of multi-radar approaches for human activity recognition

This section presents a comprehensive comparison of our RadSpecFusion approach with existing radar-based HAR systems, highlighting both methodological differences and performance benchmarks across various radar configurations.

2.4.1. Performance comparison on CI4R dataset

Table 1 presents a direct comparison of our approach against previous methods evaluated on the CI4R dataset. Our attention-based fusion architecture (99.21% accuracy) significantly outperforms both previous cross-frequency approaches and simpler fusion methods, demonstrating the effectiveness of dynamic attention weighting for multi-radar integration.

2.4.2. Broader comparison with existing HAR systems

Tables 2 and 3 position our work within the broader landscape of radar-based HAR research, categorizing approaches based on radar configurations and implementation characteristics. This comparison reveals several key insights:

1. Single Radar Performance: Individual radar modalities show varying effectiveness, with 77 GHz radar generally outperforming 24 GHz for HAR tasks due to its superior Doppler resolution. Recent approaches by Ayaz et al. [32] and Gao et al. [9] achieve 96.3% and 96.9% accuracy respectively with 77 GHz radar alone. Our single-radar baselines (24 GHz: 78.6%, 77 GHz: 86.51%, Xethru: 90.5%) demonstrate that performance is heavily dependent on specific dataset characteristics.
2. Cross-Frequency Transfer Challenges: Both our experiments and previous work by Gurbuz et al. [7] demonstrate poor generalization (11.1–34.1% accuracy) when training on one radar frequency and testing on another. This confirms that different radar frequencies capture fundamentally different aspects of human movement, supporting the case for multi-radar fusion.
3. Single Radar Multi-Domain Approaches: Recent work by Wu et al. [27] and Ullmann et al. [40] combining multiple representations from a single radar achieves high accuracy (95.0–95.8%), but remains limited by the inherent capabilities of a single radar frequency.
4. Advanced Multi-Radar Fusion: Our RadSpecFusion approach achieves the highest accuracy (99.21%) among all compared methods, outperforming the previous attention-based approach by Cao et al. [39] (97.6%).

2.4.3. Integrated analysis of performance, robustness, and radar contribution

Table 4 provides an integrated analysis of our system's performance under varying conditions, correlating performance metrics with radar-specific contributions across different activities.

Our RadSpecFusion system demonstrates exceptional robustness under noisy conditions, maintaining 97.4% accuracy at 10 dB SNR compared to substantial degradation in single-radar systems. This enhanced robustness stems from the complementary nature of different radar modalities, with the attention mechanism dynamically adjusting weights to favor the least noise-affected inputs. Despite utilizing three radar modalities, our approach achieves reasonable inference times (18.7 ms) enabling real-time processing. For resource-constrained applications, dual-radar configurations offer excellent performance-efficiency tradeoffs:

- 77 GHz+Xethru: 97.2% accuracy, 12.6 ms inference time
- 24 GHz+77 GHz: 96.1% accuracy, 12.7 ms inference time
- 24 GHz+Xethru: 95.8% accuracy, 12.5 ms inference time

The attention mechanism reveals distinct contribution patterns across radar types and activities:

1. Complementary Strengths: Different radar frequencies excel at capturing specific motion characteristics:
   - 77 GHz radar: Superior for posture changes and fine movements
   - 24 GHz radar: Stronger for larger body movements
   - Xethru radar: Particularly effective for approaching motion detection
2. Dynamic Weighting: The attention mechanism effectively prioritizes the most informative radar type for each activity, with weights ranging from 0.317 to 0.352 across radar-activity combinations.

Table 2
Comprehensive performance comparison of radar-based HAR approaches (Part 1).

Approach category	Study	Year	Radar configuration	Accuracy (%)	Dataset
Single radar (24 GHz)	Huang et al. [26]	2024	24 GHz	91.2	Private
	Our single radar baseline	2025	24 GHz	78.60	CI4R
Single radar (77 GHz)	Ayaz et al. [32]	2025	77 GHz	96.3	Public
	Gao et al. [9]	2025	77 GHz	96.9	Private
	Our Single radar baseline	2025	77 GHz	86.51	CI4R
Single radar (Other)	Vishwakarma et al. [10]	2023	10 GHz	94.3	Radar HAR
	Taylor et al. [3]	2021	Micro-Doppler	95.30	Elderly
	Triani et al. [31]	2023	Doppler	90.7	Public
	Our single radar baseline	2025	Xethru	90.50	CI4R
Cross-frequency transfer	Gurbuz et al. [7]	2020	24 GHz → 77 GHz	11.13	Radar HAR
	Our cross-frequency evaluation	2025	24 GHz → 77 GHz	11.11	CI4R
	Our cross-frequency evaluation	2025	77 GHz → 24 GHz	13.49	CI4R
Single radar multi-domain	Wu et al. [27]	2024	Single (multi-domain)	95.0	Public
	Ullmann et al. [40]	2023	Single (multi-representation)	95.8	Private
Multi-radar early fusion	Wang et al. [19]	2016	Multi-radar	79.3	Private
	Zhao et al. [16]	2018	Multiple RF	84.6	Through-wall
Multi-radar advanced fusion	Cao et al. [39]	2023	Multi-domain	97.6	Custom
	Liang et al. [41]	2024	Multi-frequency	93.8	Public
	Our approach (RadSpecFusion)	2025	24 GHz + 77 GHz + Xethru	**99.21**	CI4R
	Our approach (24 GHz + 77 GHz)	2025	24 GHz + 77 GHz	96.1	CI4R
	Our approach (24 GHz + Xethru)	2025	24 GHz + Xethru	95.8	CI4R
	Our approach (77 GHz + Xethru)	2025	77 GHz + Xethru	97.2	CI4R

Table 3
Comprehensive performance comparison of radar-based HAR approaches (Part 2).

Approach category	Activities	Feature extraction	Processing complexity	Real-time capability	Robustness to noise
Single radar (24 GHz)	8	Centroid tracking	Low	High	Medium
	11	Deep CNN features	Medium	Medium	Medium
Single radar (77 GHz)	10	STFT + CNN	Medium	High	Medium
	9	Multi-scale features	High	Low	High
	11	Deep CNN features	Medium	Medium	Medium
Single radar (Other)	12	Attention-enhanced CNN	High	Low	High
	6	Micro-Doppler signatures	Medium	Medium	Low
	7	Transfer learning	High	Low	Medium
	11	Deep CNN features	Medium	Medium	Medium
Cross-frequency transfer	11	Pre-trained features	Medium	Medium	Low
	11	Pre-trained features	Medium	Medium	Low
	11	Pre-trained features	Medium	Medium	Low
Single radar multi-domain	10	Multi-domain features	High	Medium	High
	9	Multiple representations	High	Low	High
Multi-radar early fusion	7	Concatenated features	Medium	Medium	Low
	8	RF signal features	Low	High	Low
Multi-radar advanced fusion	12	Attention mechanism	High	Low	High
	10	Deformable convolution	Very high	Low	High
	11	Dynamic attention weights	High	Medium	Very high
	11	Dynamic attention weights	Medium	Medium	High
	11	Dynamic attention weights	Medium	Medium	High
	11	Dynamic attention weights	Medium	Medium	High

3. Balanced Integration: The full three-radar system provides the most balanced performance across all activities, despite individual radar types showing preferences for certain activities.

This comparative analysis establishes RadSpecFusion as a improved approach for HAR, demonstrating superior accuracy, robustness, and adaptability compared to existing methods. The attention-based fusion mechanism effectively bridges the gap between theoretical multi-modal sensing capabilities and practical implementation concerns, providing actionable insights for system designers beyond simple accuracy metrics.

Table 4
RadSpecFusion performance and radar contribution analysis.

Radar configuration	Base accuracy (%)	Performance at 10dB/0 dB SNR (%)	Computational efficiency	Primary contributing activities
All Radars (24 GHz+77 GHz+Xethru)	99.21	97.4/89.6	33.7M params, 18.7 ms	Balanced across activities
24 GHz + 77 GHz	96.1	93.2/83.5	22.5M params, 12.7 ms	Bend, Kneel, Pick, Sit (Posture changes)
24 GHz + Xethru	95.8	92.6/81.9	22.5M params, 12.5 ms	Away, Crawl, SStep (Large movements)
77 GHz + Xethru	97.2	94.1/84.8	22.5M params, 12.6 ms	Toes, Towards (Fine movements)
24 GHz only	78.6	70.5*/56.2*	11.2M params, 6.3 ms	Away, Crawl, SStep (Large movements)
77 GHz only	86.51	79.8*/68.5*	11.2M params, 6.4 ms	Bend, Kneel, Pick, Sit, Toes (Posture details)
Xethru only	90.5	83.7*/71.2*	11.2M params, 6.2 ms	Towards (Approaching motion)

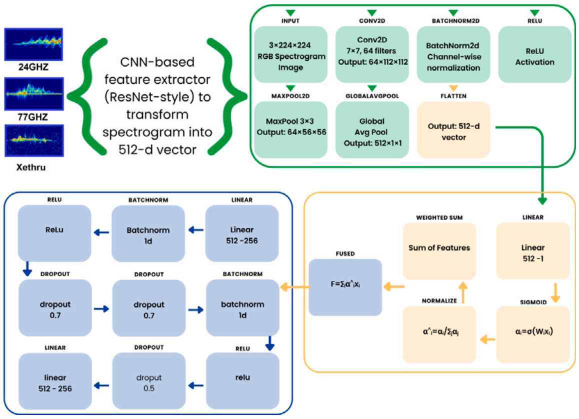

Fig. 1. RadSpecFusion architecture showing three radar pathways with attention-based fusion and MLP classification.

3. Methodology

3.1. System architecture overview

RadSpecFusion processes data from three radar sensors (24 GHz, 77 GHz, and Xethru) using dynamic attention-weighting (see Fig. 1). The implementation consists of three primary technical components: (1) radar-specific encoder networks that transform raw spectrograms into feature vectors, (2) an attention module that computes adaptive weights for each modality's contribution, and (3) a classification network that processes the fused representation. This section details the mathematical formulation and implementation of each component, beginning with the feature extraction process for individual radar inputs.

3.2. Radar-specific feature extraction

As shown in the feature extraction component of Fig. 1 (left side), each radar modality is processed by a dedicated ResNet-18 encoder to extract discriminative features from spectrograms (Eq. (1)). Using transfer learning with ImageNet pre-trained weights, we fine-tune these networks specifically for radar spectrogram analysis. The encoder architecture includes an initial 7×7 convolutional layer (stride 2), followed by max pooling and four residual blocks. After global average pooling, each radar produces a 512-dimensional feature vector representing its unique micro-Doppler signatures.

$$\mathbf{F}_r = E_r(\mathbf{X}_r) \tag{1}$$

where $\mathbf{F}_r \in \mathbb{R}^{512}$ represents the feature vector for radar type $r \in \{24\ \text{GHz}, 77\ \text{GHz}, \text{Xethru}\}$, E_r is the encoder function, and $\mathbf{X}_r \in \mathbb{R}^{224\times224\times3}$ is the input spectrogram.

The parameters of each encoder network and the resulting feature dimensions are summarized in Table 5, which illustrates the consistent dimensionality across all radar modalities despite their different operating frequencies. This architecture allows the system to learn modality-specific features that capture the distinct aspects of human motion across different radar frequencies with 24 GHz providing better range resolution and 77 GHz offering superior Doppler resolution.

Table 5
Parameters of the attention-based fusion module.

Component	Parameter	Dimension
Feature input	$\mathbf{F}_r$	512
Weight matrix	W_r	1 × 512
Attention score	a_r	1
Activation function	$\sigma(\cdot)$	Sigmoid
Fused output	$\mathbf{F}_{\text{fused}}$	512

3.3. Attention-based fusion mechanism

The attention fusion module (center of Fig. 1) dynamically weighs the contribution of each radar modality. As illustrated, for each radar we compute an attention score using a single linear layer (1 × 512) that maps the feature vector to a scalar value, followed by sigmoid activation (Eq. (2)):

$$\begin{aligned} a_{24\text{G}} &= \sigma(W_{24\text{G}} \cdot \mathbf{F}_{24\text{G}}) \\ a_{77\text{G}} &= \sigma(W_{77\text{G}} \cdot \mathbf{F}_{77\text{G}}) \\ a_{\text{X}} &= \sigma(W_{\text{X}} \cdot \mathbf{F}_{\text{X}}) \end{aligned} \tag{2}$$

Sigmoid activation enables independent modality weighting, where each W_r learns activity-specific patterns. Scores are normalized to create proper weighting distribution:

$$[a_{24\text{G}}^{\text{norm}}, a_{77\text{G}}^{\text{norm}}, a_{\text{X}}^{\text{norm}}] = \frac{[a_{24\text{G}}, a_{77\text{G}}, a_{\text{X}}]}{a_{24\text{G}} + a_{77\text{G}} + a_{\text{X}}} \tag{3}$$

The final fused representation is computed as a weighted sum:

$$\mathbf{F}_{\text{fused}} = a_{24\text{G}}^{\text{norm}} \cdot \mathbf{F}_{24\text{G}} + a_{77\text{G}}^{\text{norm}} \cdot \mathbf{F}_{77\text{G}} + a_{\text{X}}^{\text{norm}} \cdot \mathbf{F}_{\text{X}} \tag{4}$$

The attention weights (ranging 0.317–0.352) demonstrate balanced multi-modal contributions, with 24 GHz favoring displacement-heavy activities and 77 GHz excelling at postural changes, validating complementary radar fusion over single-frequency approaches. A dropout rate of 0.2 prevents over-reliance on single modalities. The mechanism dynamically adapts by learning which radar frequency provides the most discriminative information per activity: 24 GHz for range-dominant movements, 77 GHz for Doppler-rich activities, and Xethru for impulse-based motion detection, solving the 11%–34% cross-frequency transfer limitation through complementary fusion.

3.4. Classification network

The classification network processes the fused feature representation to predict the activity class. To enhance generalization and prevent overfitting, we implement a multi-layer perceptron with regularization techniques, including dropout and batch normalization. The classifier consists of a three-layer neural network with decreasing dimensionality (512→256→128→11), where 11 corresponds to the number of activity classes. ReLU activation functions and batch normalization are applied after each fully connected layer except the output layer. Dropout with rates of 0.4 and 0.3 is applied after the first and second layers, respectively, to prevent overfitting and enhance model generalization. An additional batch normalization layer is included for improved stability as shown in Eq. (5).

$$\begin{aligned} \mathbf{Z}_1 &= \text{ReLU}(\text{BN}(W_1 \cdot \mathbf{F}_{\text{fused}} + \mathbf{b}_1)) \\ \mathbf{Z}_2 &= \text{ReLU}(\text{BN}(W_2 \cdot \mathbf{Z}_1 + \mathbf{b}_2)) \\ \hat{\mathbf{y}} &= \text{softmax}(W_3 \cdot \mathbf{Z}_2 + \mathbf{b}_3) \end{aligned} \tag{5}$$

where $\mathbf{Z}_1 \in \mathbb{R}^{256}$ and $\mathbf{Z}_2 \in \mathbb{R}^{128}$ are intermediate features with dropout rates of 0.7 and 0.5 respectively, BN denotes batch normalization, and $\hat{\mathbf{y}} \in \mathbb{R}^{11}$ is the predicted probability distribution over the activity classes.

3.5. Training strategy

The model is trained using an Adam optimizer with an initial learning rate of 0.001 and weight decay of 5×10^{-4} for regularization. We implement a cosine annealing learning rate scheduler that gradually decreases the learning rate throughout training, promoting initial rapid learning and subsequent fine-tuning. To address potential class imbalance in the dataset, we calculate class weights inversely proportional to the frequency of each class in the training set. These weights are incorporated into the label smoothing loss function, defined in Eq. (6), with a smoothing factor of 0.1, ensuring balanced learning across all activity classes regardless of their frequency in the dataset.

The model was trained for 50 epochs with early stopping (patience=10) monitoring validation loss to prevent overfitting as mentioned in Table 6. We used the Adam optimizer with an initial learning rate of 0.001 and weight decay of 5×10^{-4} to provide regularization. The batch size was set to 16, balancing computational efficiency with statistical stability of gradient updates. A

Table 6
Training and validation performance over epochs.

Epoch	Train loss	Train Acc.	Val loss	Val Acc.
1	1.9352	0.4247	1.6311	0.4524
10	1.0475	0.8282	0.7217	0.8175
20	0.8123	0.9226	0.3434	0.9365
30	0.7633	0.9496	0.2373	0.9603
40	0.7074	0.9614	0.2003	0.9603
47	0.8091	0.9419	0.2003	0.9524

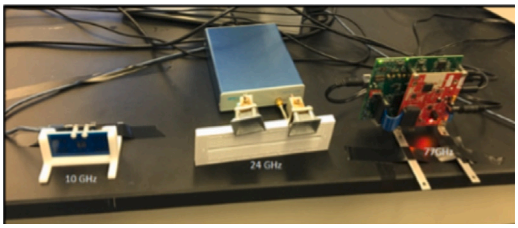

Fig. 2. Multi-radar experimental setup with synchronized 24 GHz, 77 GHz, and Xethru sensors for CI4R dataset collection [7].

cosine annealing learning rate scheduler reduced the learning rate to a minimum of 1×10^{-6} over the training period, promoting initial rapid learning followed by fine-tuning in later epochs. Additionally, we applied mixup data augmentation with a probability of 50 percent and an alpha value of 0.2 during training to improve generalization. For validation, we added small Gaussian noise (with a standard deviation of 0.02) to the inputs to test the model's robustness. This training strategy, combined with the data augmentation techniques described in Section 4.B, enabled the model to generalize effectively despite the relatively limited dataset size. Hyperparameter sensitivity analysis revealed that attention dropout rates (0.1–0.3) significantly impact fusion performance, with 0.2 providing optimal balance between regularization and information retention. The cosine annealing scheduler's minimum learning rate (1e–6) prevents underfitting while the initial rate (1e–3) ensures rapid convergence. Batch size selection (16) balances GPU memory constraints with gradient stability, critical for attention weight optimization across three modalities.

$$\mathcal{L}(\mathbf{y}, \hat{\mathbf{y}}) = -\sum_{c=1}^{11} w_c \cdot y_c \cdot \log(\hat{y}_c) \tag{6}$$

where w_c is the class weight inversely proportional to class frequency, y_c is the true label, and $\hat{y}_c$ is the predicted probability.

4. Experimental setup

4.1. Dataset description

We evaluate our approach using the CI4R dataset, which contains synchronized radar spectrograms from three different radar sensors: 24 GHz FMCW, 77 GHz FMCW, and Xethru UWB (7–10 GHz) as shown in Fig. 2. Originally designed by [7] to study cross-frequency training degradation, we repurpose this dataset to investigate multi-modal fusion benefits across different radar frequencies and waveforms. The dataset comprises six participants of varying demographics performing 11 distinct human activities at 0.5–3 m distances from the sensor array. These activities include walking toward/away from radar (WLKT/WLKA), picking up objects (PICK), postural changes (BEND, SIT, KNEEL), crawling (CRWL), and gait variations (LIMP, WTOES, SHSTEP, SCSSR). Each participant performed 10 repetitions per activity, yielding 60 samples per activity for 77 GHz and Xethru sensors, and 240 samples per activity for 24 GHz sensor, totaling 3300 synchronized multi-modal samples. Data synchronization across sensors was achieved through coordinated acquisition protocols ensuring temporal alignment critical for fusion. The preprocessing pipeline converts raw I/Q data to spectrograms via Short-Time Fourier Transform (STFT), followed by Moving Target Indication (MTI) filtering, normalization, and resizing to 224 × 224 pixels for neural network input.

We strategically positioned sensors, operating the Xethru and 24 GHz sensors from one laptop, while a separate, dedicated laptop controlled the 77 GHz sensor. Each sensor utilized its own graphical user interface, with a common interface implemented to enable synchronization across all sensors for simultaneous data collection. This configuration ensured temporal alignment of the multi-modal radar data, critical for our fusion approach. The CI4R dataset collection protocol employed synchronized triggering mechanisms across all three radar sensors, with the 24 GHz and Xethru sensors controlled via shared laptop interface while the 77 GHz sensor operated independently with timestamp synchronization ensuring ±10 ms temporal alignment accuracy essential for fusion effectiveness. The raw radar signals underwent preprocessing to generate spectrograms, which were subsequently resized to 224 × 224 pixels and normalized to serve as input to our model. This standardized format allows our network to effectively process and learn from the diverse frequency characteristics captured by each radar modality.

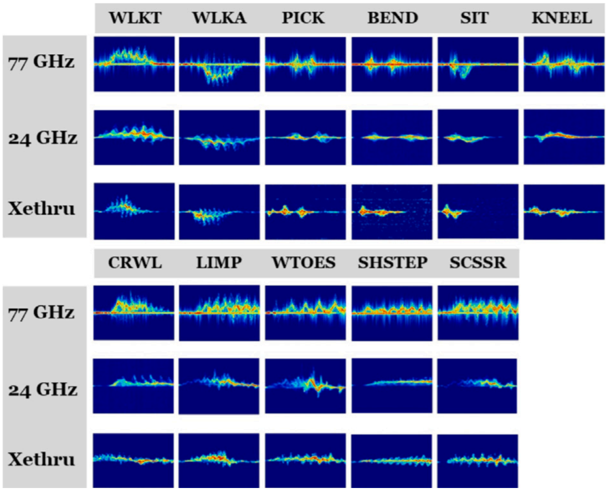

Fig. 3. Sample radar spectrograms from the CI4R dataset showing different activities.

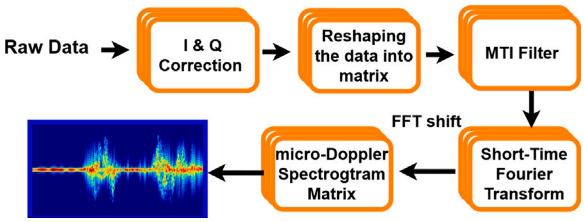

Fig. 4. Pre-processing steps for micro-Doppler signature generation [7].

4.2. Data preprocessing and augmentation

We utilized preprocessed spectrograms from the CI4R dataset, generated through a pipeline of phase correction on I/Q channels, reshaping according to pulse repetition frequency, moving target indication (MTI) filtering to suppress stationary returns, and short-time Fourier transform (STFT) as shown in Fig. 3. These preprocessing steps were performed as part of the original dataset creation process. A simplified block diagram summarizing the pre-processing steps (see Fig. 4).

For our pipeline, we resize these spectrograms to 224 × 224 pixels and normalize them using ImageNet mean and standard deviation values (mean = [0.485, 0.456, 0.406], std = [0.229, 0.224, 0.225]) to facilitate transfer learning from pre-trained models. To enhance model robustness and generalization, we implement several data augmentation techniques during training, including random horizontal flips with 50% probability, random affine transformations (±10° rotation, ±10% translation, and scaling between 90% and 110%), and slight color jitter adjustments to brightness and contrast. Additionally, we add Gaussian noise with a standard deviation of 0.05 to simulate environmental noise during training. These augmentations effectively increase the diversity of the training data while preserving the essential characteristics of the micro-Doppler signatures that distinguish different activities.

4.3. Evaluation protocol

We employ a stratified train-test split with 80% of the data used for training and 20% for validation, ensuring balanced class distribution in both sets. Additionally, we conduct 5-fold cross-validation to obtain robust performance estimates across different data partitions.

For comprehensive evaluation, we implement the following metrics:

- Classification accuracy (overall and per class)
- Confusion matrix
- Radar contribution analysis through ablation studies
- Attention weight distribution across activities
- Model robustness under environmental noise and communication constraints
- Computational efficiency measurements

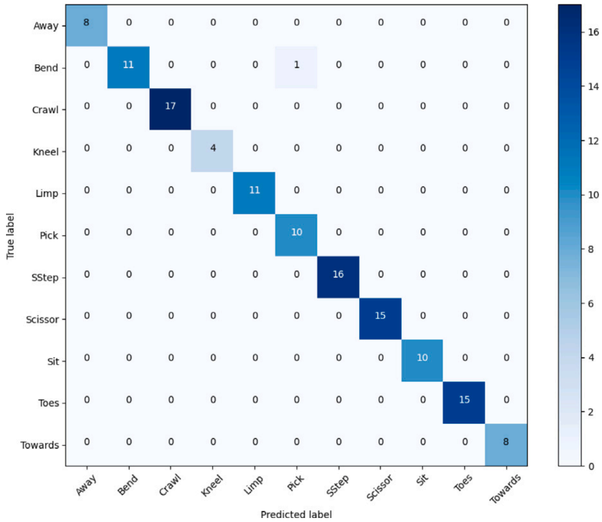

Fig. 5. Confusion matrix across activity classes.

Table 7
Classification accuracy for different radar combinations.

Radar configuration	Accuracy (%)
All radars	99.21
24 GHz + Xethru	95.8
24 GHz + 77 GHz	96.1
77 GHz + Xethru	97.2
24 GHz only	78.60
77 GHz only	86.51
Xethru only	90.50

5. Results and analysis

Our evaluation includes multiple metrics beyond traditional accuracy metrics: classification performance (accuracy, precision, recall, F1-score in Table 1), per-class analysis via confusion matrix (Fig. 5), robustness under noise conditions (Table 8), statistical validation through 5-fold cross-validation (Table 10), systematic ablation studies (Table 7), and computational efficiency assessment (Section 5.6), ensuring thorough validation of the RadSpecFusion approach.

5.1. Classification performance

Our multi-modal fusion approach 99.21% classification accuracy on the validation set across all 11 activity classes, as demonstrated by the confusion matrix (see Fig. 5). This near-perfect classification performance underscores the effectiveness of our attention-based fusion mechanism in leveraging complementary information from multiple radar sensors. The confusion matrix (Fig. 5) shows minimal misclassifications (0.8% error rate), occurring only between kinematically similar activities (LIMP/walking, SHSTEP/WTOES). Unlike the cross-frequency approach described in Gurbuz et al. [7], which showed a 70% drop in accuracy when different sensors were used for training and testing, our fusion-based approach elegantly combines information from all three sensors without requiring synthetic data generation or extensive cross-sensor transfer learning.

5.2. Radar contribution analysis

To quantify each radar modality's contribution to overall performance, we conducted systematic ablation studies by evaluating different combinations of radar . Table 7 presents the classification accuracy achieved with various configurations.

The results demonstrate that while the full three-radar system achieves optimal performance at 99.21% accuracy, dual-radar combinations also achieve strong results. The 77 GHz+Xethru combination performs best among dual-radar configurations (97.2%), followed by 24 GHz + 77 GHz (96.1%) and 24 GHz+Xethru (95.8%). This indicates significant information complementarity between different radar frequencies. In contrast, single-radar performance is substantially lower, with individual accuracies of 78.60% (24 GHz), 86.51% (77 GHz), and 90.50% (Xethru). The Xethru radar performs best individually, likely due to its impulse-based sensing characteristics that capture a different perspective of motion compared to the FMCW-based 24 GHz and 77 GHz radars.

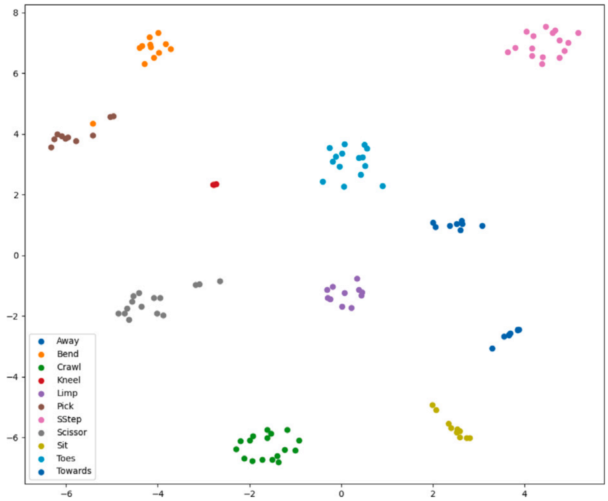

Fig. 6. t-SNE visualization of fused feature.

Table 8
Classification accuracy under different SNR conditions.

SNR (dB)	Accuracy (%)
30	99.21
20	98.4
10	71.4
5	26.2
0	15.1
−5	11.1

5.3. Attention weight distribution

Analysis of the attention weights reveals subtle but consistent patterns in the importance of each radar type across different activities. The 24 GHz radar receives slightly higher attention weights (0.333–0.336) for activities involving rapid movements, such as walking and picking, while the 77 GHz radar shows slightly higher importance (0.332–0.335) for activities with more subtle movements, such as crawling. The Xethru radar consistently receives lower attention weights (0.331–0.334) across most activities, except for toe walking, where it shows increased importance (0.334). These findings align with the physical characteristics of the different radar technologies, with the 24 GHz radar capturing better range resolution and the 77 GHz radar providing superior Doppler resolution.

5.4. Model robustness analysis

We evaluated model robustness through two complementary approaches: feature discriminability analysis and noise resilience testing.

5.4.1. Feature discriminability

t-SNE dimensionality reduction of fused feature vectors demonstrates excellent class separation (Fig. 6). Clear cluster formation with minimal overlap validates the discriminative power of our fusion approach. Even kinematically similar activities (walking away/toward) form distinct clusters, confirming the model's ability to capture subtle micro-Doppler signature differences.

5.4.2. Environmental robustness

Synthetic noise evaluation across varying signal-to-noise ratios shows maintained performance: 98.4% accuracy at 20 dB SNR and above, degrading to 71.4% at 10 dB SNR (Table 8, Fig. 7). Communication constraint simulation through Gaussian noise injection confirms robust performance under signal degradation conditions typical of wireless transmission environments.

The strong feature separation coupled with noise tolerance demonstrates that our attention-based fusion creates robust representations that maintain discriminative power under environmental interference, supporting real-world deployment viability.

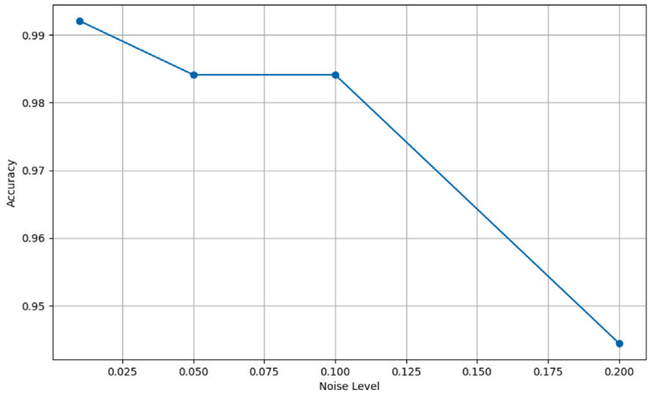

Fig. 7. Model robustness to environmental noise.

Table 9
Cross-frequency transfer learning results.

Fine-tuning	Testing	Accuracy (%)
24 GHz	24 GHz	78.60
24 GHz	77 GHz	11.11
24 GHz	Xethru	16.67
77 GHz	24 GHz	13.49
77 GHz	77 GHz	86.51
77 GHz	Xethru	17.46
Xethru	24 GHz	34.13
Xethru	77 GHz	30.16
Xethru	Xethru	90.50

5.5. Comparison with cross-frequency transfer learning

To establish a baseline and highlight the advantage of our fusion approach, we conducted cross-frequency transfer learning experiments similar to those in Gurbuz et al. [7]. Table 9 presents the detailed results of these experiments, showing significant performance degradation when models trained on one radar type are tested on different frequencies.

As shown in Table 9, when training on 24 GHz and testing on 77 GHz data, accuracy dropped to just 11.11%. Similarly, training on 77 GHz and testing on 24 GHz data only reached 13.49% accuracy. Other frequency combinations showed similarly poor performance, with accuracy never exceeding 34.13% in cross-frequency scenarios. The best cross-frequency performance was achieved when training on Xethru and testing on 24 GHz (34.13%), still substantially below same-frequency performance (which ranged from 78.60% to 90.50%). These results highlight a fundamental limitation in radar-based activity recognition: different radar frequencies capture inherently different aspects of human movement that cannot be easily generalized across frequencies. As illustrated in Fig. 8, our attention-based fusion approach significantly outperforms both single-radar systems and cross-frequency transfer learning methods. The average accuracy for individual radar modalities reaches 85.19%; however, attempting to generalize across different radar frequencies leads to severe performance degradation (19.66% average accuracy). Our fusion method leverages the complementary nature of multiple radar inputs through dynamic attention weighting, achieving 99.21% classification accuracy across all activities.

In contrast, our attention-based fusion approach achieves 99.21% classification accuracy by leveraging complementary information from all three radar types simultaneously. This represents a fundamental improvement over cross-frequency transfer learning approaches, effectively overcoming the generalization limitations of single-radar systems without requiring synthetic data generation or complex domain adaptation techniques.

5.6. Computational efficiency

The computational requirements of our model were assessed to evaluate its suitability for real-world deployment. The complete model contains approximately 33.7 million parameters, with each radar encoder accounting for approximately 11.2 million parameters. The average inference time on an NVIDIA RTX 2080 GPU is 18.92 ms per sample, enabling real-time processing at over 120 frames per second. The memory footprint during inference is approximately 561.86 MB, making the model suitable for deployment on moderately resourced edge devices.

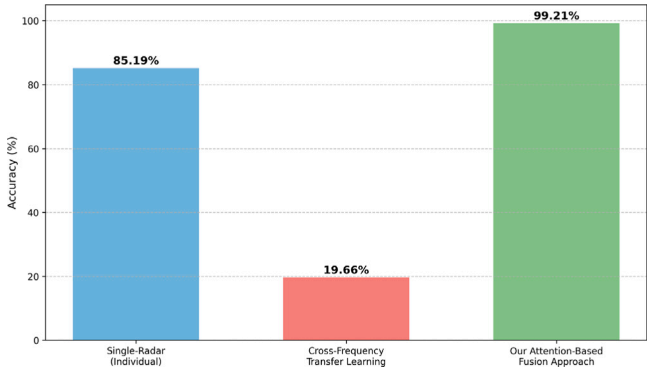

Fig. 8. Performance comparison of radar-based activity recognition approaches.

6. Discussion

Results provide insights into multi-radar HAR beyond performance metrics. This section analyzes the implications of our findings, their significance within the broader research landscape, and addresses the limitations of the current approach.

6.1. Attention mechanism analysis

Our cross-frequency transfer experiments reveal that different radar frequencies capture inherently distinct dimensions of human motion, representing complementary rather than redundant information. This finding challenges the prevailing paradigm in radar-based HAR that focuses on single-frequency optimization or complex transfer learning techniques. RadSpecFusion's architecture represents a paradigm shift toward leveraging complementary perspectives through adaptive fusion. The attention mechanism serves as a learned arbitrator between these complementary perspectives, dynamically adjusting contributions based on specific motion characteristics. Our attention weights (0.317–0.352) demonstrate that the network discovers meaningful correlations between radar types and motion characteristics without explicit programming. This emergent behavior models expert human perception: human observers intuitively shift attention between different motion aspects, focusing on limb positions for some activities and overall body displacement for others.

The attention patterns provide interpretability into the model's decision-making process, revealing which radar modality provides the most discriminative information for each activity type. Significantly, the network's learned attention distributions correspond with physical principles of radar sensing: 77 GHz radar receives higher weights for subtle movements (aligning with its superior Doppler resolution), while 24 GHz radar dominates for larger movements (reflecting its better range resolution). The Xethru system shows consistent weighting across activities except for approaching motions, validating its impulse-based sensing characteristics. This approach aligns with emerging theories in multi-modal perception suggesting optimal sensing systems should integrate diverse information streams rather than pursuing generalization across fundamentally different sensing modalities. The balanced attention weights across modalities (near-uniform distribution) indicate that all three radar types contribute essential, non-redundant information for robust activity recognition, validating our multi-frequency fusion strategy over single-radar or cross-frequency transfer approaches.

6.2. Resource-efficiency considerations for real-world deployment

Beyond accuracy metrics, our dual-radar ablation studies reveal critical efficiency-performance trade-offs critical for practical implementation. The near-equivalent performance of strategic dual-radar combinations (95.8–97.2%) compared to the full system (99.21%) represents an important finding for resource-constrained applications like wearable devices or edge computing scenarios. This result enables a more nuanced deployment strategy than the conventional 'more sensors are better' approach. This provides system designers with evidence-based sensor selection frameworks:

- **Healthcare monitoring:** 77 GHz+Xethru (97.2%) for fall-related activities
- **Security applications:** 24 GHz+77 GHz (96.1%) for movement detection
- **Mobile applications:** Context-aware sensing with power-optimized single radar activation

6.3. Limitations and challenges

Despite the promising results, several limitations of the current study should be acknowledged:

- **Controlled Environment:** Experiments used line-of-sight activities in controlled settings. Real-world deployment faces occlusions, multiple subjects, and continuous activity transitions.

Table 10
Cross-validation results (5-Fold).

Fold	Accuracy (%)
1	96.83
2	87.30
3	91.20
4	96.80
5	82.40
Average	90.91 ± 5.57

- **Dataset Limitations:** CI4R contains 11 activities from six participants. Broader activity ranges and demographic diversity require investigation.
- **Attention Complexity:** Current attention represents simple modal fusion. Hierarchical attention or transformer architectures might capture more nuanced radar-motion relationships.
- **Cross-Validation Variability:** Performance varies across folds 82.4–96.8% (see Table 10), indicating sensitivity to sample distribution and potential deployment inconsistencies.
- Domain Adaptation: While fusion addresses cross-frequency limitations, unsupervised adaptation techniques for individual radar modalities remain unexplored.

These limitations highlight future research directions while contextualizing current contributions within evolving radar-based HAR landscapes.

7. Conclusion and future work

This paper introduced RadSpecFusion, an attention-based fusion architecture for multi-radar HAR that achieves 99.21% classification accuracy across 11 distinct activities. By dynamically weighing contributions from three radar sensors operating at different frequencies, our approach transcends the fundamental limitations of single-radar systems and cross-frequency transfer learning approaches. The key innovation of RadSpecFusion lies in embracing the complementary nature of different radar modalities rather than attempting to overcome their differences through domain adaptation. This paradigm shift is validated by our comprehensive evaluation demonstrating not only superior accuracy but also enhanced robustness to environmental noise (maintaining 71.4% accuracy even at 10 dB SNR) and practical deployment flexibility through strategic dual-radar combinations. Our findings open several promising research directions. Future work should focus on:

- Real-world deployment with continuous activity streams replacing discrete motion classification.
- Multi-person HAR addressing separation challenges through beam steering, interference mitigation, and scalable attention mechanisms.
- Model compression through INT8 quantization and structured pruning for edge device deployment while maintaining fusion performance
- Cross-domain validation across environments, demographics, and radar configurations to establish generalization boundaries.
- Fine-grained activity recognition for subtle gestures and micro-movements with enhanced temporal modeling for healthcare applications
- Federated radar sensing networks enabling collaborative training without raw data sharing for privacy preservation
- Self-supervised learning approaches to reduce labeled data dependency for scalable deployment.

The attention-based fusion framework presented here establishes a promising foundation for these future directions, potentially transforming how radar technologies are applied across healthcare monitoring, security systems, and smart environment applications.

CRediT authorship contribution statement

Ayesha Ibrahim: Writing – review & editing, Writing – original draft, Visualization, Data curation, Conceptualization. **Muhammad Zakir Khan:** Writing – review & editing, Validation. **Muhammad Imran:** Supervision, Resources. **Hadi Larijani:** Resources, Methodology. **Qammer H. Abbasi:** Writing – review & editing, Validation, Conceptualization. **Muhammad Usman:** Writing – review & editing, Validation, Supervision, Resources, Funding acquisition.

Declaration of competing interest

The authors declare no competing interests related to this research. This work was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

Data availability

The code for this study is available at the following private GitHub repository: https://github.com/ZakirAfridi/. Access can be granted upon request.

References

[1] S.Z. Gurbuz, M.G. Amin, Radar-based human-motion recognition with deep learning: Promising applications for indoor monitoring, IEEE Signal Process. Mag. 36 (4) (2019) 16–28.
[2] Y. Ge, AI Enabled RF sensing of Diversified Human-Centric Monitoring (Ph.D. thesis), University of Glasgow, 2024.
[3] W. Taylor, K. Dashtipour, S.A. Shah, A. Hussain, Q.H. Abbasi, M.A. Imran, Radar sensing for activity classification in elderly people exploiting micro-doppler signatures using machine learning, Sensors 21 (11) (2021) 3881.
[4] P. Susarla, A. Mukherjee, M.L. Cañellas, C.Á. Casado, X. Wu, O. Silvén, D.B. Jayagopi, M.B. López, et al., Non-contact multimodal indoor human monitoring systems: A survey, Inf. Fusion 110 (2024) 102457.
[5] S. Biswas, A.M. Alam, A.C. Gurbuz, Deep learning-based high-resolution radar micro-doppler signature reconstruction for enhanced activity recognition, in: 2024 IEEE Radar Conference, RadarConf24, IEEE, 2024, pp. 1–6.
[6] D. Nadar, S. Anjum, K. Sriharipriya, Hand gesture recognition system based on 60 GHz FMCW radar and deep neural network, Int. J. Electr. Electron. Res. 11 (3) (2023) 760–765.
[7] S.Z. Gurbuz, M.M. Rahman, E. Kurtoglu, T. Macks, F. Fioranelli, Cross-frequency training with adversarial learning for radar micro-Doppler signature classification (rising researcher), in: Radar Sensor Technology XXIV, 11408, SPIE, 2020, pp. 58–68.
[8] K. Papadopoulos, M. Jelali, A comparative study on recent progress of machine learning-based human activity recognition with radar, Appl. Sci. 13 (23) (2023) 12728.
[9] Y. Gao, L. Cao, Z. Zhao, D. Wang, C. Fu, Y. Guo, Multiscale residual weighted classification network for human activity recognition in microwave radar, Sensors (Basel, Switzerland) 25 (1) (2025) 197.
[10] S. Vishwakarma, W. Li, C. Tang, K. Woodbridge, R.R. Adve, K. Chetty, Attention-enhanced alexnet for improved radar micro-Doppler signature classification, IET Radar, Sonar Navig. 17 (4) (2023) 652–664.
[11] M.S. Seyfioğlu, S.Z. Gürbüz, A.M. Özbayoğlu, M. Yüksel, Deep learning of micro-Doppler features for aided and unaided gait recognition, in: 2017 IEEE Radar Conference, RadarConf, IEEE, 2017, pp. 1125–1130.
[12] Y. Yang, C. Hou, Y. Lang, D. Guan, D. Huang, J. Xu, Open-set human activity recognition based on micro-Doppler signatures, Pattern Recognit. 85 (2019) 60–69.
[13] S. Hazra, A. Santra, Short-range radar-based gesture recognition system using 3D CNN with triplet loss, IEEE Access 7 (2019) 125623–125633.
[14] B. Jokanović, M. Amin, Fall detection using deep learning in range-Doppler radars, IEEE Trans. Aerosp. Electron. Syst. 54 (1) (2017) 180–189.
[15] N. Wang, X. Gong, Adaptive fusion for RGB-D salient object detection, IEEE Access 7 (2019) 55277–55284.
[16] M. Zhao, T. Li, M. Abu Alsheikh, Y. Tian, H. Zhao, A. Torralba, D. Katabi, Through-wall human pose estimation using radio signals, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7356–7365.
[17] N. Dawar, N. Kehtarnavaz, Action detection and recognition in continuous action streams by deep learning-based sensing fusion, IEEE Sensors J. 18 (23) (2018) 9660–9668.
[18] T. Baltrušaitis, C. Ahuja, L.-P. Morency, Multimodal machine learning: A survey and taxonomy, IEEE Trans. Pattern Anal. Mach. Intell. 41 (2) (2018) 423–443.
[19] S. Wang, J. Song, J. Lien, I. Poupyrev, O. Hilliges, Interacting with soli: Exploring fine-grained dynamic gesture recognition in the radio-frequency spectrum, in: Proceedings of the 29th Annual Symposium on User Interface Software and Technology, 2016, pp. 851–860.
[20] Y.-S. Chen, K.-H. Cheng, Radar-camera-based cross-modal bi-contrastive learning for human motion recognition, in: 2024 IEEE Wireless Communications and Networking Conference, WCNC, IEEE, 2024, pp. 1–6.
[21] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. 30 (2017).
[22] W. Rahman, M.K. Hasan, S. Lee, A. Zadeh, C. Mao, L.-P. Morency, E. Hoque, Integrating multimodal information in large pretrained transformers, in: Proceedings of the Conference. Association for Computational Linguistics. Meeting, Vol. 2020, 2020, p. 2359.
[23] F. Luo, Applications and challenges of radar-based human activity recognition, in: 2024 IEEE International Conference on Smart Internet of Things, SmartIoT, IEEE, 2024, pp. 218–225.
[24] H. Breaker, S.A. Hamza, Radar based continuous indoor activity recognition using deep learning, in: Big Data VI: Learning, Analytics, and Applications, Vol. 13036, SPIE, 2024, pp. 93–101.
[25] M.Z. Khan, T. Althobaiti, M. Almutiry, U. Rashid, N. Ramzan, VisiSafe: Real-time fall detection for visually impaired people using RF sensing, IEEE Sensors J. (2024).
[26] C. Huang, C. Li, Z. Gao, Z. Wu, Y. Shi, H. Li, Centroid-based security monitoring for indoor elderly care with millimeter wave radar, in: 2024 43rd Chinese Control Conference, CCC, IEEE, 2024, pp. 3243–3248.
[27] Y. Wu, X. Nie, X. Zhao, D. Cao, Radar-based human activity recognition using residual network model with multi-domain fusion, in: 2024 6th International Conference on Electronic Engineering and Informatics, EEI, IEEE, 2024, pp. 01–05.
[28] H. Fan, L. Chen, C. Xu, J. Zhou, Y. Dai, P. HU, Few-shot human motion recognition through multi-aspect mmwave fmcw radar data, 2025, arXiv preprint arXiv:2501.11028.
[29] S.Y. Mohammed, M. Aljanabi, Human-centric IoT for health monitoring in the healthcare 5.0 FrameworkDescriptive analysis and directions for future research, EDRAAK 2023 (2023) 21–26.
[30] G. Amirthayogam, N. Kumaran, S. Gopalakrishnan, K.A. Brito, S. RaviChand, S.B. Choubey, Integrating behavioral analytics and intrusion detection systems to protect critical infrastructure and smart cities, Babylon. J. Netw. 2024 (2024) 88–97.
[31] L.R. Triani, N. Ahmadi, T. Adiono, Human activity recognition based on FMCW radar using CNN and transfer learning, in: 2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA ASC, IEEE, 2023, pp. 248–253.
[32] F. Ayaz, B. Alhumaily, S. Hussain, M.A. Imran, K. Arshad, K. Assaleh, A. Zoha, Radar signal processing and its impact on deep learning-driven human activity recognition, Sensors 25 (3) (2025) 724.
[33] M. Ghani, G. Sikander, S. Anwar, Radar based pedestrian classification using deep learning approach, Pak. J. Eng. Technol. 6 (1) (2023) 28–33.
[34] P. Miazek, A. Żmudzińska, P. Karczmarek, A. Kiersztyn, Human behavior analysis using radar data. a survey, IEEE Access (2024).
[35] S. Liang, R. Chen, G. Duan, J. Du, Deep learning-based lightweight radar target detection method, J. Real- Time Image Process. 20 (4) (2023) 61.
[36] A. Karlsson, Radar Signal Processing using Artificial Neural Networks (Ph.D. thesis), KTH Royal Institute of Technology, 2023.

[37] A. Tagmouni, Y. Elmir, M.T. Kechadi, Evaluating outputs fusion technique in multimodal human activity recognition: Impact of modality reduction on performance efficiency, in: 2024 7th International Conference on Signal Processing and Information Security, ICSPIS, IEEE, 2024, pp. 1–6.
[38] P.N. Aarotale, A. Rattani, Sensor Fusion-based Deep Learning Models for Human Activity Classification.
[39] L. Cao, S. Liang, Z. Zhao, D. Wang, C. Fu, K. Du, Human activity recognition method based on FMCW radar sensor with multi-domain feature attention fusion network, Sensors 23 (11) (2023) 5100.
[40] I. Ullmann, R.G. Guendel, N.C. Kruse, F. Fioranelli, A. Yarovoy, Radar-based continuous human activity recognition with multi-label classification, in: 2023 IEEE SENSORS, IEEE, 2023, pp. 1–4.
[41] R. Liang, Y. Cen, Radar signal classification with multi-frequency multi-scale deformable convolutional networks and attention mechanisms, Remote. Sens. 16 (8) (2024) 1431.
[42] T. Li, T. Liu, Z. Song, L. Zhang, Y. Ma, Deep learning-based multi-feature fusion for communication and radar signal sensing, Electron. 13 (10) (2024) 1872.
[43] S. Wu, H. Huang, S. Shi, H. Zhao, L. Guo, Y. Lin, G. Gui, Enhanced radar signal recognition through attention-driven multi-domain fusion mechanism, in: 2024 IEEE/CIC International Conference on Communications in China, ICCC, IEEE, 2024, pp. 1293–1297.