Received: 31 August 2025 | Revised: 13 March 2026 | Accepted: 25 June 2026 | Published online: 10 July 2026

Artificial Intelligence and Applications
2026, Vol. 00(00) 1–30
DOI: 10.47852/bonviewAIA62027517

REVIEW

WiFi-Based Human Activity Recognition and Fall Detection with Taxonomy, Benchmarks, and Future Directions: A Narrative Review

Kok Chung Chua[1], Kai Liang Lew[2], Chean Khim Toa[1,*], Matthew Alexander Paudianto[1], Ivan Nathanael[1] and Abel Nathanael Hutapea[1]

[1] *School of Computing and Data Science, Xiamen University Malaysia, Malaysia*

[2] *Faculty of Engineering & Technology, Multimedia University, Malaysia*

Abstract: WiFi-based human activity recognition (HAR) and fall detection have emerged as promising alternatives to vision- and wearable-based systems for modern assisted living and clinical monitoring for the elderly. These approaches exploit variations in Channel State Information, capturing amplitude and phase fluctuations across subcarriers to characterize how human motion perturbs these patterns through multipath propagation. This narrative review synthesizes 26 peer-reviewed studies published between 2017 and 2025, focusing on systems evaluated in real-world, non-laboratory environments. While recent deep learning models demonstrate near-perfect benchmark accuracy on fine-grained datasets, cross-environment assessments consistently reveal a severe "reality gap." Specifically, several reviewed studies reported accuracy declines exceeding 20% when deployed in previously unseen buildings, while multi-user interference reduced recognition accuracy by 21–46%. Furthermore, dataset availability and reproducibility are critical considerations; only a limited number of reviewed studies utilized publicly available datasets, and none provided a standardized benchmarking guideline. To address these gaps, this review presents a comprehensive taxonomy of post-2017 architectures, a comparative synthesis of laboratory versus deployment performance, and a critical analysis of key implementation challenges. We conclude by proposing a set of recommended reporting practices and outlining actionable future research. In doing so, the paper outlines key recommendations for developing adaptive, location-independent models, emphasizing the need for domain-transfer techniques, cross-building validation, and expanding open datasets to enhance generalizability.

Keywords: Channel State Information (CSI), deep learning, domain adaptation, human activity recognition (HAR), WiFi

1. Introduction

Fall remains one of the most critical health risks [1–5] for the elderly, often resulting in disability, bone fractures, and even mortality. Because of this, developing robust human activity recognition (HAR) [6–14] and fall detection systems is essential for modern assisted living and clinical monitoring for the elderly. While traditional HAR relies on a camera or a wearable [15], WiFi-based sensing has become an alternative in recent 5 years. It is based on wireless infrastructure and provides non-intrusive monitoring. It also preserves user privacy by eliminating line-of-sight video recording by the camera-based sensing and bypasses the common issues with wearables, where users often find them uncomfortable, forget to charge them, or simply refuse to wear them [16].

The path of WiFi-based HAR systems starts with the use of Received Signal Strength Indicator (RSSI) measurements. Although RSSI can detect human activity, it is inherently limited as it only provides a low-resolution view of signal changes. Besides, it mainly captures large-scale body movements. Afterward, the system pivoted toward micro-Doppler shifting signatures [16–20] to capture the subtle kinematic details [21]. But the real breakthrough came with the move to Channel State Information (CSI). By tapping into individual subcarriers of standard wireless networks, CSI offers a far more granular "map" of human motion. It tracks exactly how multipath components shift in amplitude and phase, effectively solving the resolution problems that made RSSI tracking unreliable in complex spaces.

Previous research predominantly leveraged handcrafted feature extraction techniques, utilizing machine learning such as Support Vector Machines (SVMs), k-Nearest Neighbors, or Dynamic Time Warping applied to the CSI for HAR. For example, Wang et al. [7] proposed CSI based Human Activity Recognition and Monitoring system (CARM), a CSI-speed model for correlating signal dynamics with human activity. Similarly, the WiFall model, proposed by Wang et al. [1], demonstrates the effectiveness of SVMs for fall detection. These existing proposed systems rely on manual feature extraction with a model trained in a room, familiar with a layout and furniture. When tested in a new environment with a different layout and

*Corresponding author: Chean Khim Toa, School of Computing and Data Science, Xiamen University Malaysia, Malaysia. Email: cheankhim.toa@xmu.edu.my

furniture, the model's performance will degrade. To enable models to adapt to various environments, research after 2020 has shifted to deep learning architectures. Common architectures such as Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) are able to learn patterns directly from raw CSI signals, without the need for tedious manual engineering. For example, Wang et al. [22] and Yang et al. [23] both achieved near-perfect accuracy on curated datasets, with ExtRe [22] achieving 100% and ERPANet [23] 99.6% on the UT-HAR and NTU-HAR datasets.

Although the WiFi field has emerged in 2019–2025, existing review articles have not adequately highlighted the deployment limitations. Previous studies from 2017 to 2025 focus primarily on controlled laboratory testing or provide generic overviews without systematically evaluating cross-environment real-world robustness. Of the 26 studies surveyed in this review [2, 5, 21, 22, 24–45], half of the studies have reported an accuracy decrease of more than 20% in unseen buildings. Additionally, we found that environmental interference, such as metallic furniture and complex room layouts, can increase false-positive rates by 17–28%, whereas multi-person scenarios can decrease accuracy by 21–46%. These metrics indicate substantial evidence that current models lack cross-environment generalization. We also identified that another underexplored area is dataset availability and reproducibility. Based on our findings, only seven reviewed studies [39–43, 45, 46] used publicly available datasets, and the lack of commonly shared datasets further makes fair comparisons and benchmarking highly challenging.

This paper provides a comprehensive review of published research on WiFi-based HAR and fall detection from 2017 to 2025 with a focus on cross-environment robustness and real-world implementations. The review pursues three objectives, including:

1) To systematically examine and classify influential studies that evaluate systems beyond controlled laboratory settings, establishing a detailed taxonomy of current datasets, signal representations, modeling, and evaluation metrics.
2) To benchmark performance and identify critical deployment barriers, specifically analyzing the performance degradation caused by cross-environment, data availability, and edge hardware constraints.
3) To address the central research question: "How can WiFi-based HAR systems transition from environment-specific systems to robust, location-agnostic ones?"

By addressing the research questions, this review first contributes to a comprehensive taxonomy and comparative benchmark of post-2017 state-of-the-art (SOTA) WiFi-based HAR systems. It explicitly highlights the contrast between laboratory successes and cross-environment degradation. The next contribution is to provide a critical synthesis of real-world deployment challenges by detailing the impacts of multi-user interference, public dataset limitations, and computational bottlenecks. Subsequent contributions proposed a set of recommended reporting practices alongside actionable future research directions. Together, these contributions provide a structured roadmap for achieving adaptive, standardized, and reproducible WiFi sensing models.

The remainder of this paper is organized as follows. Section 2 presents the survey methodology, detailing the search strategy, inclusion and exclusion criteria, and classification framework used to select the 26 studies. Section 3 reviews the literature and develops the taxonomy. Section 4 discusses benchmark results and deployment-related challenges. Section 5 outlines future research directions. Section 6 concludes the review.

2. Literature Search and Selection Approach

This article is a narrative review of WiFi-based HAR and fall detection studies published between 2017 and 2025, based on a structured search of major scholarly databases. This review aims to provide an interpretive synthesis of major technical developments, benchmark practices, deployment challenges, and future research directions rather than an exhaustive systematic review.

2.1. Search strategy and data sources

The search covered papers published between 2017 and 2025 on WiFi-based HAR and fall detection. The search was conducted in IEEE Xplore, ACM Digital Library, ScienceDirect, SpringerLink, and Google Scholar to cover both journal and conference publications. The keywords used included WiFi-based HAR, fall detection, CSI, Doppler frequency shift, RSSI, and deep learning. Boolean operators were used to combine search terms and capture common variations in terminology across the literature.

2.2. Selection criteria

Only English-language papers that analyzed WiFi signal, such as CSI amplitude and phase, RSSI, and Doppler or time-frequency features, and reported HAR or fall detection are considered in this review. Moreover, papers are included if they were published between 2017 and 2025 in bibliographic databases, and the conducted evaluations should contain at least one experimental condition not corresponding to a studio or controlled lab, for example, cross-room or cross-building, cross-device, and/or cross-person deployment.

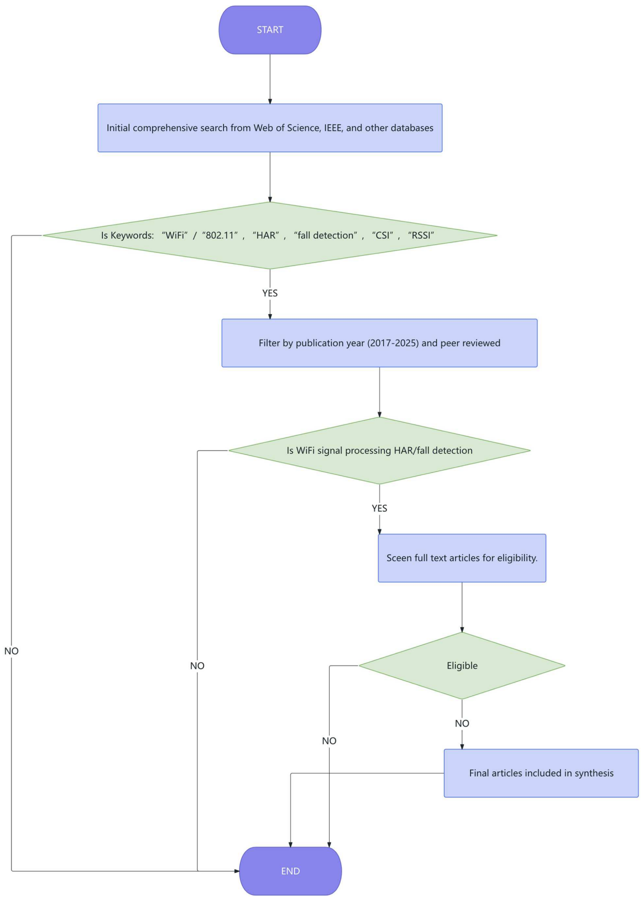
Papers were excluded if WiFi was not the primary sensing modality, if the results were simulation-only without hardware experiments, if essential methodological details were not reported, or if the study focused only on camera-based or wearable-based sensing. If more than one version of a similar work existed, the most complete version was kept. Therefore, the final set of 26 studies for the extended analysis was obtained by applying these criteria.

2.3. Screening process

There were two stages in this screening process. In the first stage, records from IEEE Xplore, the ACM Digital Library, ScienceDirect, SpringerLink, and Google Scholar were initially retrieved (Section 2.1). Metadata and full texts were collected, and duplicates were removed. At this stage, non-English-language manuscripts as well as publications that were evidently not relevant to WiFi-based HAR and fall detection were removed. The second stage involved a more in-depth review of relevance and suitability. The corpus was limited to studies on WiFi-based HAR or fall detection, with publication dates between 2017 and 2025. Following that, full-text checks were conducted to confirm that the studies were relevant to the review topic and included deployment-related evaluation beyond a controlled laboratory environment, such as cross-room, cross-building, or cross-device scenarios. Searches for relevant literature were expanded in terms of backward and forward referencing to identify additional relevant studies that met the defined inclusion criteria.

The overall identification and screening process is presented in Figure 1. In total, 26 studies were retained for detailed comparative analysis in Sections 3 and 4, while additional background and survey references were cited where needed for context.

Figure 1
Literature search and study selection process



2.4. Classification framework

To facilitate consistent comparison, we analyzed each study based on four major dimensions: sensing and signals, feature representation, model and learning paradigm, and task and scenario.

The first dimension, sensing and signals, specifies the WiFi measurements used (RSSI, CSI amplitude, CSI phase, or amplitude–phase fusion) and signal processing methods used. Special focus is placed on phase information handling (raw phase, reconstructed phase, or interference-robust extraction), as this heavily affects stability with respect to environmental change. This dimension also covers studies extracting Doppler or time-frequency signatures based on WiFi signals.

The second dimension, feature representation, describes the way studies represent WiFi signals as input for learning. Methods range from standardized raw time series, time-frequency or spectrogram images, and graph or correlation structures across the subcarriers to compact handcrafted statistics. We distinguish between single-stream and multi-stream fusion strategies to show whether amplitude and phase are processed jointly or separately.

The third dimension, model and learning paradigm, encompasses both the architectures of models as well as the training strategies used. This dimension includes architectures such as CNN, LSTM, Gated Recurrent Unit (GRU), Graph Neural Network, and attention-based models. Different learning paradigms include fully supervised, domain adaptation, semi-supervision, self-supervised pre-training for related domains, adversarial or cross-adversarial learning, and continual or incremental learning. It also captures the measures used to drive efficiency, such as pruning, lightweight model design, and deployment on edge devices.

Lastly, the task and scenario coverage describes the type of recognition problem (general HAR vs. fall detection, binary vs. multi-class) and operating conditions as well. These conditions can include single-user or multi-user activity, stationary or moving access points, and real-time versus offline inference.

3. Taxonomy

This section presents a taxonomy of the research on WiFi-based HAR, covering literature from 2017 to 2025. Four core areas are categorized in the taxonomy, which are dataset landscape, signal preprocessing, modeling, and evaluation metrics, as illustrated in Figure 2. By following this systematic categorization, a consistent comparison of different methodologies through the years is established.

3.1. Dataset landscape

3.1.1. Public dataset

1) Overview

Public datasets allow researchers to assess the validity of their findings. They act as a common benchmark to determine whether a newly proposed signal processing method and model architecture provide improvements over previous work. Furthermore, privately collecting WiFi sensing datasets requires specialized hardware and a complex setup [47]. Therefore, a public dataset serves as a lower barrier to entry option for students and research groups who are new to the field. Kaggle, IEEE DataPort, and GitHub are common platforms to host these datasets. Table 1 presents a list of publicly available datasets, ranging from daily movements to fine-grained gesture recognition.

2) Experiment setup

The experimental setup is the most important source of variability in the dataset characteristics. Specifically, the experimental setup consists of hardware configurations and environmental conditions.

The choice of hardware can be viewed as a trade-off between applicability, efficiency, and standardization. Commercial off-the-shelf (COTS) WiFi routers are often employed in studies that aim to maximize real-world applicability [22]. In contrast, a low-cost distributed device, for example, ESP32, is utilized to achieve energy-efficient and scalable deployments [34]. Nonetheless, Intel 5300 Network Interface Cards (NICs) with dedicated transmitter–receiver (Tx–Rx) pairs remain as the primary choice for most public datasets, reflecting a more research-oriented configuration.

Environmental conditions further complicate comparisons. Earlier WiFi-based HAR studies were typically conducted in controlled laboratories. Therefore, they often yield optimistic results but fail in real-world deployment. Recent research has begun to collect datasets in heterogeneous environments. For example, performance was compared in offices, dormitories, and laboratories [5]. Variations in room size, furniture layout, and Tx–Rx distance introduce additional degrees of freedom that are seldom

Figure 2
The taxonomy of WiFi-based HAR literature

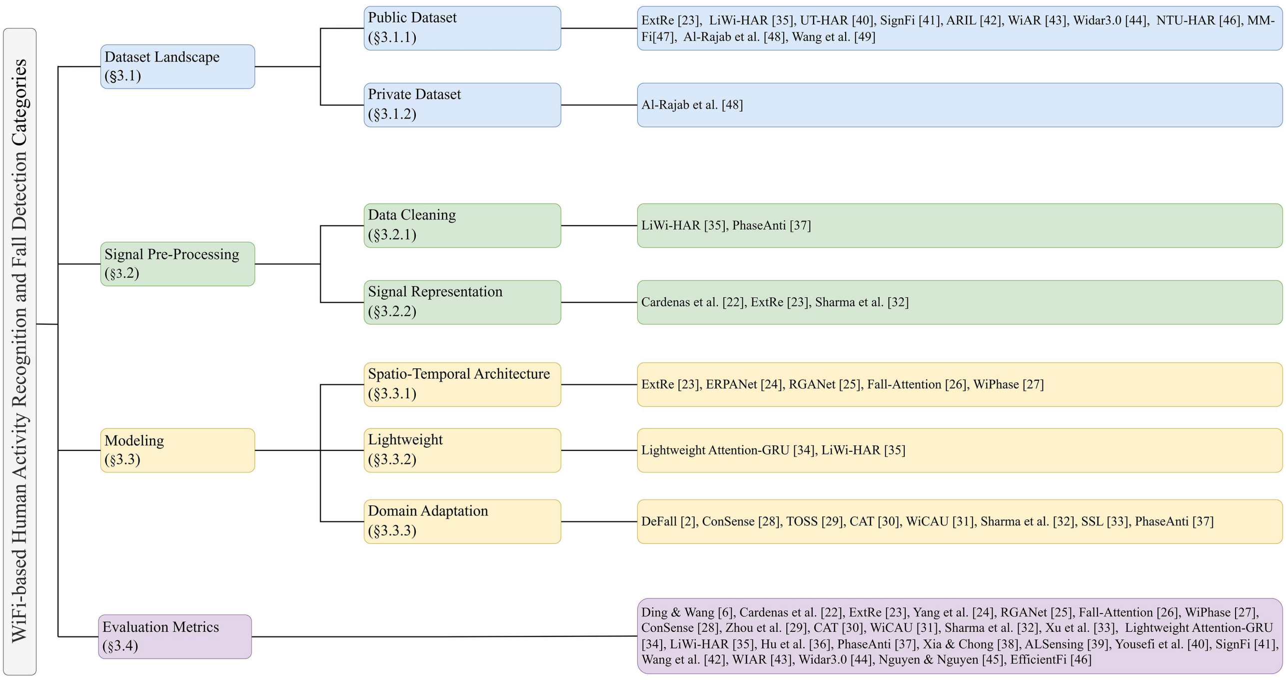


Table 1
Summary of prominent public datasets for WiFi-based HAR

References	Dataset	Modality	Task	Number of classes	Number of instances
[39]	UT-HAR	WiFi CSI	Human behavior recognition in indoor areas	7	720
[45]	NTU-HAR	WiFi CSI	Human behavior recognition in indoor areas	6	2400
[40]	SignFi	WiFi CSI	Sign language recognition	276	14,280
[41]	ARIL	WiFi CSI	Joint activity recognition	6	1394
[42]	WiAR	WiFi CSI, RSSI	Human behavior recognition in indoor areas	16	4800
[46]	MMFi	RGB, Depth, LiDAR, mmWave, WiFi CSI	Daily and rehabilitation actions recognition	27	1080
[43]	Widar 3.0	WiFi CSI, RSSI	Hand gesture recognition	16	17,750

reported in sufficient detail. Table 2 highlights the diversity in the experimental configurations of various representative studies.

A parallel categories diagram in Figure 3 shows the multilevel relationships between datasets, software extraction tools, hardware platforms, and experimental environments. A bias toward the Intel 5300 NIC hardware platform and the Linux 802.11n CSI tool is observed for the standard benchmark datasets UT-HAR and Widar3.0. But the TP-Link N750 is also used in a variety of settings, along with the OpenRF software tools, from laboratory environments to in-home and domestic environments. This variation is relevant for cross-platform WiFi-based sensing.

Figure 4 shows the visual representation of the spatial parameters of existing public WiFi CSI datasets. It is observed that there is a wide range of experimental environments used in the current literature. Standard benchmarks (e.g., UT-HAR and ARIL) use relatively small room environments with Tx–Rx distances typically between 2 and 4 m. However, the SignFi (Lab) dataset covers a much larger room area (see Table 2) and thus exhibits a more complex multipath environment. Furthermore, data points with unspecified dimensions (in the shape of diamonds or squares) indicate a lack of standardized reporting of the physical layout of sensing environments in the early literature.

Early datasets often suffer from domain-shift limitations [48]. This is the condition in which a machine learning algorithm fails to perform well in a new environmental setup and device specification. Therefore, it is important to choose a public dataset that offers a multi-environment setup to better assess the model's generalizability [42].

3.1.2. *Private dataset*

Most WiFi-based HAR studies still utilize private datasets as the primary option for their experiments [47]. Since private datasets offer greater control over experimental design, they allow

Table 2
Equipment and environment from public dataset

References	Dataset	Software	Hardware	Equipment setup	Environment
[39]	UT-HAR	Unspecified	Transmitter (Tx): Router Receiver (Rx): Intel 5300 NIC	Tx and Rx distance: 3 meters under line-of-sight (LOS)	Indoor office area Dimension: Unspecified
[45]	NTU-HAR	Atheros CSI tool	Access points (APs): Two TP-Link N750	Tx and Rx distance: Unspecified (LOS) The two APs act as both receivers and transmitters.	Room Dimension: 5 × 6.5 m
[40]	SignFi	OpenRF, a modified version of the 802.11n CSI tool	AP and station (STA): Laptop with WiFi Link 5300	AP and STA distance: 2.3 m Angle between transmit antenna array and direct path: 40 degrees	Lab Dimension: 13 × 12 m It has a more complex multipath environment.
				AP and STA distance: 1.3 m Angle between transmit antenna array and direct path: 90 degrees	Home Dimension: 4.11 × 3.86 m Characteristic: Less complex multipath environment
[41]	ARIL	GNU Radio	Tx and Rx: PC + Ettus N210 Synchronization hardware: Ettus Clock and synchronization cables Connectivity: Ethernet Cable	Tx and Rx distance: Approximately 2.4 m The Ettus Clock is used to synchronize the two Ettus N210 units.	Room Dimension: Unspecified
[42]	WiAR	Linux 802.11n CSI tool	Tx and Rx: Lenovo ThinkPad T400 with Intel WiFi Wireless Link 5300 802.11n MIMO radios	Tx and Rx distance: 4 m	Empty room Dimension: 6 × 8 m Meeting room Dimension: 6 × 10 m Office Dimension: 6 × 8 ma
[46]	MMFi	Atheros CSI tool	Tx and Rx: TP-Link N750	Tx and Rx distance: 3.75 m	Room Dimension: Unspecified
[43]	Widar3.0	Linux CSI tool	1 Tx and 6 Rx: Off-the-shelf Mini desktops with Intel 5300 wireless NIC	Sensing area: 2 × 2 m 1 Tx and 2 Rx are placed in the corner of the sensing area. 4 Rx are placed randomly outside the sensing area.	Empty classroom Dimension: 4.5 × 5.5 m Hall room Dimension: 4.5 × 2.5 m Office Dimension: 4 × 2.5 m

Figure 3
Equipment and environment configuration of public WiFi CSI datasets

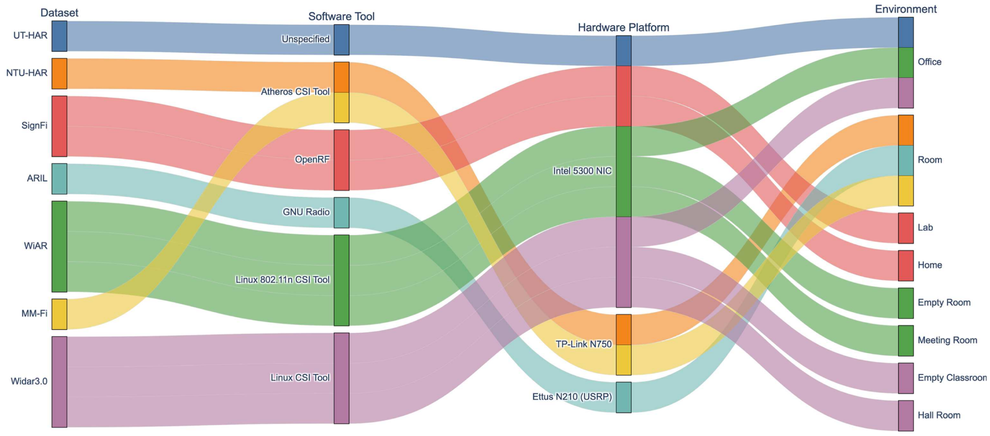
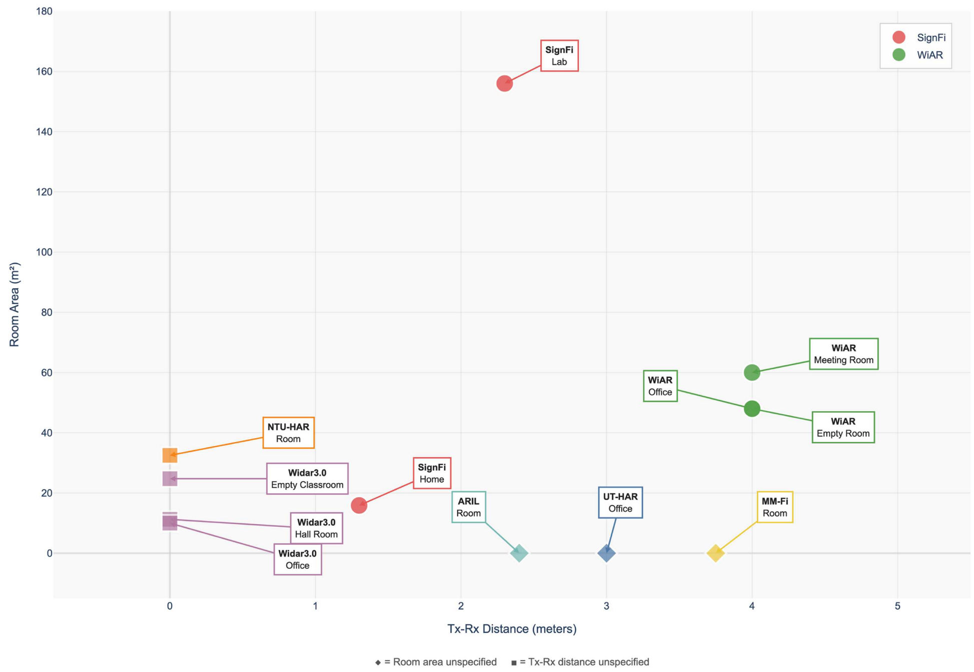


Figure 4
Spatial configuration of public WiFi CSI datasets comparing transmitter–receiver separation distance against room area



Note: Diamonds indicate datasets with unspecified room dimensions; squares indicate datasets with unspecified Tx–Rx distance.

researchers to tackle more specific research objectives. Nevertheless, papers that hinge on a private, undisclosed dataset are subject to a lack of reproducibility. Furthermore, comparing the performance of the proposed technique in such papers is unreliable, as performance differences are more likely to stem from the dataset characteristics rather than algorithmic improvement. Thus, there is an increasing call for either the release of private datasets or, at a minimum, the publication of detailed replication protocols that describe hardware, environment, and preprocessing transparently.

3.2. Signal preprocessing

3.2.1. Data cleaning

Raw WiFi is inherently susceptible to environmental noise, multipath fading, and hardware imperfections. To extract reliable human motion patterns, robust filtering mechanisms must be applied to the raw signal. In LiWi-HAR [34], to clean the raw CSI data, it uses 1-D linear interpolation to reconstruct packets lost during transmission. Then, to filter outliers, Moving Absolute Median Deviation was used. Finally, after the data are cleaned, the cleaned data are passed to PCA to reduce the dimension.

Besides environmental noise, dense Internet of Things deployments suffer from significant co-channel interference from multiple transmitters in the same frequency band. To handle this, PhaseAnti [36] uses a phase component that is called Nonlinear Phase Error Variation (NLPEV), which is also sensitive to human motion. To get NLPEV, first, the raw phase data need to be unwrapped to correct the phase value that is distributed between and to get a continuous linear curve. Then, it is passed through a gradient filter and a Hampel filter to eliminate faulty phase shift and high frequencies. This helps separate motion-related phase information from environmental interference and hardware-induced artifacts.

3.2.2. Signal representation

Cárdenas et al. [21] exploited Doppler frequency diversity to harvest micro-Dopplers generated on falls with 98.1% accuracy per location. In the beginning, the receiver antenna (RX) receives the continuous wave. Rather than using one frequency, they used the frequency diversity technique by transmitting the three subcarriers' signals to the stretched signal. This raw signal has the shape still blurred. The effect of thermal noise and static bounce from objects (like walls or furniture) dominated a lot, which makes the signature from falling activity covered by noise and clutter, as shown in Figure 5.

To process the unprocessed frequency spectrum, they processed it using Short-Time Fourier Transform (STFT) to change the raw signal into a spectrogram that represents time versus frequency. In this spectrogram, the falling activity showed a trail of sharp spikes in the Doppler frequency distribution that are very different in pattern compared to a person just walking normally, which can be seen in Figures 6 and 7.

Another paper that used STFT is Reference [31]. They manipulated the raw amplitude from the CSI signal, which was then denoised using the Discrete Wavelet Transform with the mother wavelet Symlet-4 (sym4), to retain the sudden change in the signal (like a falling event) without distorting the signal. Only after denoising, the signal was executed using STFT.

Raw CSI cannot be directly processed as phase signals are prone to nonlinear distortion due to imperfections in hardware, such as carrier frequency offset and sampling clock offset. Thus, to address this problem, Wang et al. [22] remove phase information while exclusively extracting the magnitude component. Subsequently, as the duration of human motion tends to vary randomly, a frame trim is applied by cutting several constant frames to standardize its temporal length. Although the temporal dimension has been standardized, adjacent frames still contain redundancy. Therefore, the frame fusion step is executed by taking the mean of every m-frame. This process ensures the subsequent deep learning model can extract spatial and temporal signals without losing the directional topology of the original signal. The overall ExtRe data collection and preprocessing procedure described above is summarized in Figure 8.

Figure 5
Un-preprocessed frequency spectrum during a falling event

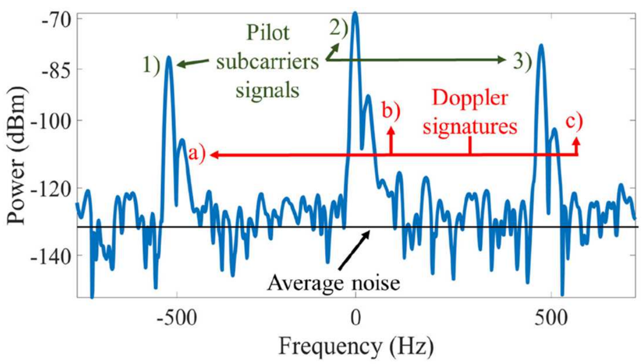


Figure 6
Spectrogram for falling event

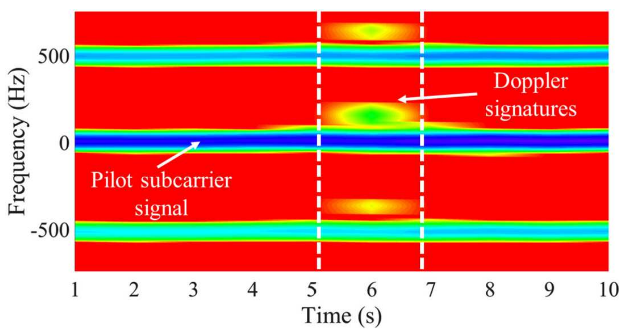
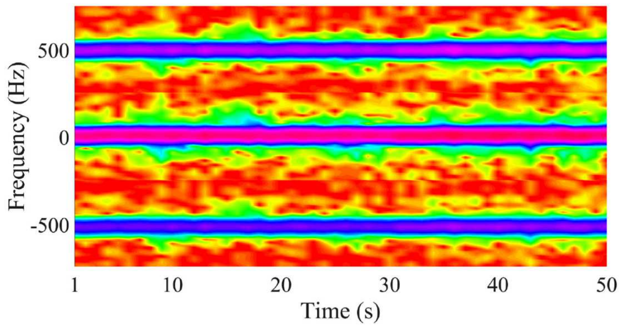


Figure 7
Spectrogram for walking event



3.3. Modeling

In terms of machine learning models, CNN, Recurrent Neural Network (RNN), and more recent transformer-based models have been shown to dominate the modeling space. Most developing trends aim to achieve HAR and fall detection in a noninvasive and real-time manner.

However, there is still a lot of methodological plurality. Other authors resort to end-to-end learning with raw CSI inputs. There is also a wide variance in the methods used to achieve generalization: some authors focus on domain adaptation techniques like cross-adversarial training (CAT), while others use a carefully curated supervised dataset.

As shown in Table 3, research in WiFi-based HAR was driven by three main themes that include fast growth of model

Figure 8
ExtRe data collection and preprocessing step

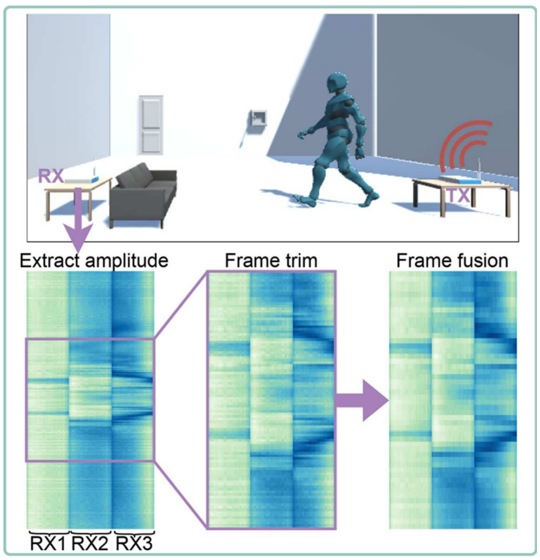


Table 3
Overview of recent technology in WiFi-based human activity recognition and fall detection

References	Purpose of the study	Method	Findings	Technique	Similarities	Uniqueness	Detailed review
[22]	To improve HAR accuracy by processing CSI as a 3D stream for both coarse- and fine-grained activities	CSI, 3D CNN, LSTM	Achieved superior accuracy (100% on fine-grained dataset) with significantly fewer FLOPs than SOTA methods	Spatiotemporal fusion, deep learning	Uses CSI data and a hybrid deep learning architecture for HAR	Novel treatment of CSI data as a 3D stream rather than a 2D image, enhancing dynamic feature capture	Proposes a 3D CNN-LSTM model that improves HAR by processing CSI as a 3D stream to better capture dynamic features
[26]	To improve HAR by fusing reconstructed CSI phase features, addressing ignored subcarrier correlations	CSI, gated pseudo-Siamese network (GPSiam), Graph Attention Network (DRGAT)	Outperforms SOTA methods with a lightweight model that shows good cross-domain generalization	Feature fusion, Graph Neural Networks (GNNs)	Uses WiFi CSI phase data for HAR with a deep learning model	First to use a Graph Attention Network (DRGAT) to model the nonlinear correlations between CSI subcarriers explicitly	Introduces a novel two-stream network that fuses temporal features with subcarrier correlations modeled by a GNN, enhancing cross-domain performance
[24]	To effectively extract both temporal and spatial features from CSI, addressing the common underutilization of spatial information is crucial	CSI, modified ResNet, GRU, attention	Achieved high accuracy (99.4% on UT-HAR) and demonstrated robustness and efficiency	Spatiotemporal fusion, model efficiency	Combines a CNN-like architecture with an RNN for HAR, a typical hybrid approach	Employs a modified, efficient ResNet (with depth-wise separable convolutions) for stronger spatial feature extraction	Combines an efficient, modified ResNet for robust spatial feature extraction with an attention-augmented GRU for temporal analysis, improving HAR accuracy
[23]	To create an end-to-end model for HAR that handles the nonstationary and fluctuating nature of CSI packets	CSI, residual network, packet attention	Outperforms SOTA methods, achieving very high accuracy (99.4% on UT-HAR, 99.6% on NTU-HAR)	Deep learning, attention mechanisms	Employs residual connections and attention, which are common trends in modern deep learning	The proposed efficient packet attention mechanism is specifically designed to encode channel info and long-range dependencies in CSI	Presents a novel residual network with a specialized packet attention mechanism to capture multi-scale features in CSI data effectively

(*Continued*)

Table 3
(*Continued*)

References	Purpose of the study	Method	Findings	Technique	Similarities	Uniqueness	Detailed review
[33]	To develop a lightweight HAR model that maintains high performance while drastically reducing model size	CSI, GRU, attention, pruning, data augmentation	Achieved SOTA accuracy (~98.9%) while reducing model parameters and FLOPs by over 99.9%	Model efficiency, lightweight AI, edge computing	Uses a GRU and attention, which are standard components for time-series analysis	Extreme model compression via pruning, combined with a simple architecture and data augmentation, to achieve SOTA efficiency	Demonstrates that a lightweight GRU with attention, when combined with data augmentation and pruning, can achieve top accuracy with a tiny fraction of the computational cost
[27]	To solve the challenge of exemplar-free class-incremental learning in HAR, allowing models to learn new activities without forgetting old ones	CSI, transformer, continual learning	Outperforms competitive continual learning methods on public datasets with fewer parameters	Continual learning, privacy, transformers	Uses a transformer architecture, which is increasingly popular for sequence data	Novel exemplar-free continual learning framework for WiFi HAR, addressing a key real-world deployment challenge	Proposes a transformer-based framework for continual learning that allows HAR systems to adapt to new activities without catastrophic forgetting or storing user data
[21]	Smart healthcare with smart devices	Radio sensing	Enhanced detection accuracy in diverse environments	Healthcare, signal processing	Use of WiFi-based sensing	Doppler frequency diversity approach	Innovative use of Doppler frequency diversity improves fall detection reliability for healthcare settings
[25]	Walk fall detection	CSI	Improved accuracy by focusing on key motion patterns	Attention mechanism, fall detection	Use of CSI data	Attention-based methodology	Attention mechanism significantly increases fall detection accuracy in walking scenarios
[34]	AIoT with ESP32 devices	CSI	Efficient human activity recognition with low computational overhead	Lightweight algorithms, IoT	WiFi-based HAR	Lightweight, distributed AIoT solution	Lightweight algorithms tailored for distributed IoT systems effectively reduce computational resources

(*Continued*)

Table 3
(*Continued*)

References	Purpose of the study	Method	Findings	Technique	Similarities	Uniqueness	Detailed review
[29]	Cross-adversarial training	CSI	Improved robustness in various environments	Adversarial training, robustness	WiFi-based HAR	Cross-adversarial training approach	Effectively addresses domain discrepancies to enhance robustness in activity recognition
[30]	Negative transfer, partial domain adaptation	CSI	Effective adaptation across diverse environments	Domain adaptation, uncertainty management	WiFi-based HAR adaptation	Uncertainty-aware adaptation approach	Mitigates negative transfer, offering efficient cross-environment model adaptation
[32]	A systematic study to evaluate Self-Supervised Learning (SSL) algorithms for HAR to address labeled data scarcity	CSI, SSL	Provides a comprehensive analysis and practical guidelines, identifying Masked Autoencoders as a promising approach to SSL	SSL, benchmarking	Addresses the common problem in the field of limited labeled data	The first systematic, comparative study of various SSL methods in the context of WiFi HAR	Provides a crucial systematic evaluation of SSL methods for WiFi HAR, highlighting their potential to reduce dependency on labeled data and identifying promising research directions
[37]	DBSCAN clustering, Smart Medical IoT	Nexmon CSI	High reliability and accuracy in medical IoT environments	Medical IoT, clustering algorithms	WiFi-based detection	DBSCAN clustering with Nexmon CSI	DBSCAN effectively clusters data for accurate medical IoT-based fall detection
[36]	NLPEV CSI, anti-interference	CSI	Significant improvements in recognition accuracy despite interference	Anti-interference techniques	CSI-based HAR	Nonlinear Phase Error Variation (NLPEV)	Effectively addresses interference issues to maintain high accuracy in recognition systems
[38]	Active learning	CSI	Maintains accuracy with reduced labeled data	Active learning	WiFi-based HAR	Active learning integration	Successfully reduces dependency on extensive labeled datasets while preserving accuracy

(*Continued*)

Table 3
(*Continued*)

References	Purpose of the study	Method	Findings	Technique	Similarities	Uniqueness	Detailed review
[28]	Target-oriented semi-supervised domain adaptation	CSI	Robust performance with minimal labeled target data	Semi-supervised learning, domain adaptation	WiFi-based HAR adaptation	Target-oriented adaptation method	Provides an effective domain adaptation solution specifically targeting WiFi-based HAR
[2]	Complex online and offline detection model	CSI	Environment-independent accuracy in fall detection	Environment independence, passive detection	WiFi-based fall detection	Online-offline model combination	Demonstrates high reliability in diverse environments, combining real-time and retrospective detection methods
[31]	Convolutional Neural Network (CNN)	CSI	High accuracy across various locations	CNN, robust classification	WiFi-based HAR	Time-frequency CNN approach	Utilizes CNN-based analysis of time-frequency diagrams, robust across different locations
[35]	Precursor to DeFall	CSI	Early insights into passive fall detection methodologies	Preliminary research, fall detection	WiFi-based fall detection	Foundational methods	It provides foundational insights that set the stage for more advanced passive fall detection systems
[44]	Adversarial data augmentation	CSI	Improved robustness via adversarial augmentation	Robustness, data augmentation	WiFi-based fall detection	Adversarial augmentation technique	Enhances system robustness effectively through the use of adversarial techniques
[5]	Recurrent Neural Network (RNN)	CSI	Reliable real-time fall detection	Smart home, real-time detection	WiFi-based fall detection	RNN application	Employs RNN effectively for reliable, real-time fall detection in smart home contexts

architectures for spatiotemporal feature extraction, an asserted effort toward reaching environmental generalization, and the establishment of lightweight models suitable for deployment on-the-edge.

3.3.1. Spatiotemporal architecture

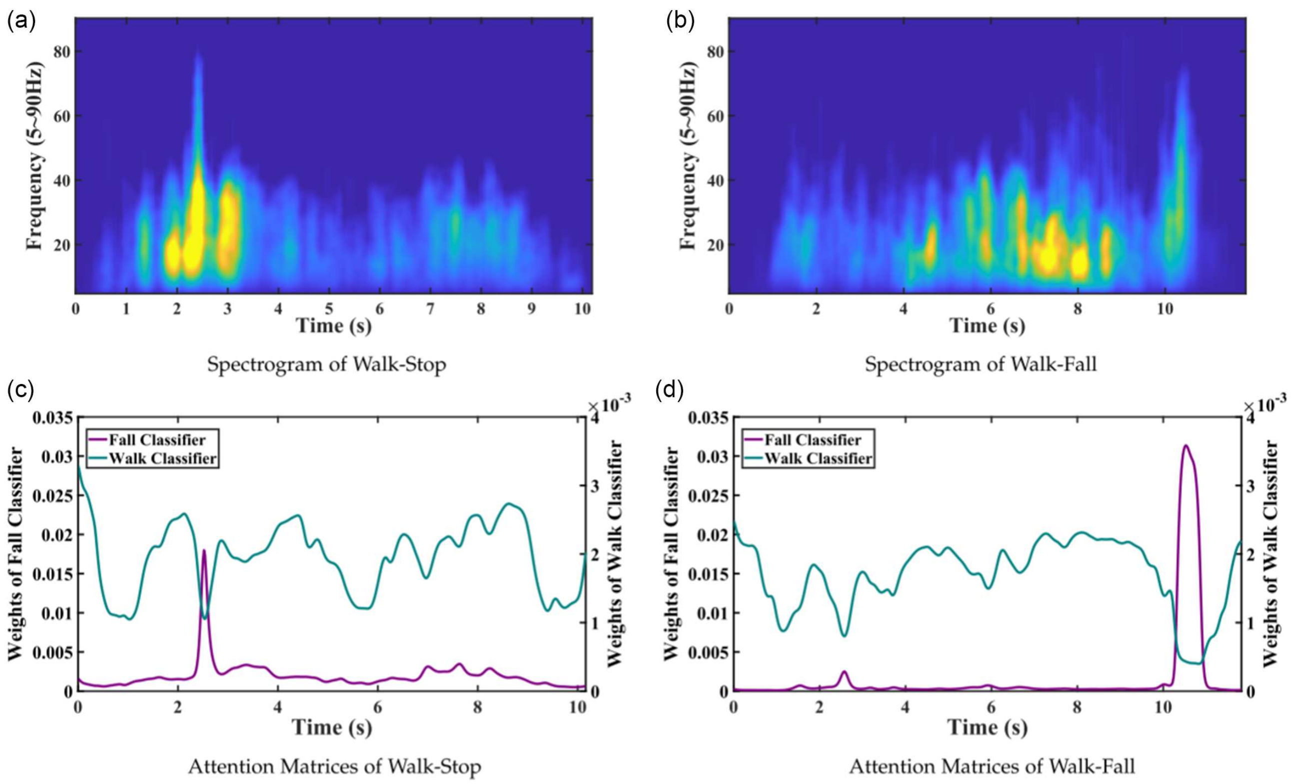
Fusing spatial and temporal CSI information presents a major challenge in WiFi-based HAR. Models at the beginning needed to employ only purely temporal or spatial features. Training the model on both features together shows substantial improvement in accuracy based on recent studies.

To address the first theme of spatiotemporal feature extraction, researchers realized that traditional 2D CNNs are insufficient for capturing the nonstationary dynamics and complex CSI signals. For example, RGANet [24] uses the depth-wise separable convolutional Residual Neural Network (ResNet) integrated with the spatial-channel Squeeze-and-Excitation module. It strengthens spatial feature extraction, followed by the features being passed to a GRU for temporal modeling. Their method has been able to achieve 99.4% and 99.24% accuracy using the UT_HAR dataset and NTU-FI HAR dataset, respectively. For example, ERPANet [23] applied efficient packet attention to mitigate the nonstationary behavior of WiFi signals caused by changing human movement patterns. With down-sampling attention residual (AR) operation and domain-specific AR operation written specifically for encoding channel information as well as long-range correlations of CSI packets, the model can learn parallel multi-scale spatiotemporal features. The model further demonstrated its efficacy with accuracy metrics of 99.4% for the UT-HAR and 99.6% for the NTU-HAR datasets, as indicated in the Benchmark Comparison subsection. A more advanced method referred to as ExtRe [22] represents CSI as a 3D stream and extracts short-term spatial and temporal features with a 3D CNN before modeling longer-term dependencies with LSTM. The method extracts short-term temporal and spatial-channel information from CSI data using a 3D CNN. Using an LSTM network, the features are modeled in a way that learns long-term temporal relationships. This design allows the system to maintain high accuracy for diverse activities of daily living under standard consumer conditions.

While the aforementioned CNN and LSTM architectures are effectively able to capture general spatiotemporal features, they often struggle to isolate overlapping activities, such as classifying a sudden fall from walking. To resolve this interference issue, attention mechanisms have become a fundamental evolution in HAR modeling. For instance, Fall-Attention [25] is an RNN model based on attention. The method solved the problem of interference between nearby activities commonly encountered in life, such as walking and falling. The proposed Fall-Attention method tackles the challenge through an attention-based sentence embedding approach and an RNN to highlight fall-related information while minimizing contributions from unrelated features. It also performs multi-task learning to recognize the fall and other activities, for example, walking at the same time. This allows it to perform better than standard RNNs in recognizing loose and overlapping activities like walking versus falling. The hierarchical model dynamically highlights the most relevant part of an action, as shown in Figure 9.

Thus, Figure 9 shows a comparison of the Walk-Stop event with the Walk-Fall event. This confirms the efficacy of the sentence embedding, where the attentional weights for “fall” and “walk” are focused on various sections along the sequence. The segmentation of these continuous activities can be inferred from the time points that overlap in the attentional weights of different

Figure 9
Distribution of attentional weights of two classifiers



activities. This is because two atomic activities are never executed at the same time [25]. Bottom right graph (d) shows the attention weight over the sequence for a purple-colored line, indicating the "Fall Classifier" and only having a quick spike toward the end of the sequence, exactly where the fall takes place. On the other hand, "Walk Classifier" (green line) keeps constant attention across the walking period. This graph clearly shows how the attention mechanism isolates the features describing the fall: this enables the model to segment and identify much more easily critical events even when they are part of an overall continuous movement.

Going beyond the architectural level, WiPhase [26] investigates correlations across CSI subcarriers that were commonly overlooked in prior work. It uses a DRGAT to learn the relations directly, and it fuses the output with temporal features extracted from a gated pseudo-Siamese network (GPSiam). Then, a network of dendrites (DD) fuses these features and makes the final decision. With this knowledge, WiPhase is able to achieve high accuracy and efficiency with a lightweight model. Finally, migrating toward graph networks represents a better understanding of CSI data's intrinsic structure and greater generalization ability.

3.3.2. Lightweight

As HAR systems advance toward processing on the device to facilitate low latency and maintain user privacy, computational efficiency has emerged as a crucial design constraint. As a result, the investigation of compact models that provide sufficiently high accuracy while requiring fewer resources has drawn attention. The Lightweight Attention-GRU [33] is a typical example of such efforts. A single-layer GRU combined with an embedded attention module achieves 98.92% accuracy on ARIL while using pruning and data augmentation to drop the model parameters by more than half. LiWi-HAR [34] is another tool aiming at low-cost ESP32 nodes. This enables fast inference on constrained hardware, as raw CSI is compressed by PCA into one or two principal components, and classification is performed using a Particle Swarm Optimization (PSO)-optimized Backpropagation Neural Network.

3.3.3. Domain adaptation

Even more than the architectural advances, a major deployment challenge is domain shift, the degradation when a model trained in one environment is deployed in a different one. Addressing this issue has therefore become a parallel research priority.

An important step in this direction is continuous learning, exemplified by ConSense [27]. Here, an approach to exemplar-free class-incremental learning is presented in which a model can incrementally learn new activities over time without losing memory of previous tasks and does not need to store prior data. This is a major step toward the realization of properly adaptive and privacy-preserving HAR systems for dynamic smart home scenarios.

In addition to this idea, domain adaptation techniques are also extremely important for greater generalization. Meta-learning has been successfully integrated into the semi-supervised domain adaptation framework for overcoming the challenges and being combined with methods like uncertainty-guided pseudo-labeling to make use of limited labeled target samples and abundant unlabeled ones [28]. It works well for one-shot and multi-source settings compared to other domain adaptation models. Similarly, DeFall [2] addresses the domain shift by adversarial data augmentation as a way of creating new training environments to span the entire range of possible scenarios through which it might walk in during inference. As a result, this method generalizes better to new locations that have not been seen before and does not require data collection in the specific area of interest. The system was tested in different environments and successfully detected falls, particularly as seen in the benchmark comparison.

Another pioneer method is the CAT framework [29] that aligns the distributions of features over domains using adversarial objectives. In fact, CAT consistently preserves accuracy even with increases in attack intensity over both Original model (ORI), standard adversarial training (SAT) models, as shown across all three device configurations (Figure 10), thus proving it to be robust against digital noise.

As seen in Figure 10, the CAT model (solid orange line) outperforms both ORI and SAT across multiple adversarial attack settings, indicating stronger robustness under cross-domain evaluation. The transfer pairs E2→E1, E3→E6, and E4→E5 denote training and testing across different environments, while the compared curves show that CAT remains more robust than ORI and SAT under adversarial perturbations. Each row corresponds to a set of adversarial attack methods used for evaluating the adversarial robustness, namely, fast gradient sign method (FGSM), projected gradient descent (PGD) (step = 1), and PGD (step = 3), from top to bottom. The evaluated adversarial settings include FGSM and PGD variants, and CAT remains more robust than ORI and SAT across the tested transfer pairs.

To tackle a related problem, WiCAU [30] applies a designed and partial domain adaptation (PDA) method in the scenario where source domain activities are absent from the target. WiCAU employs feature alignment along with source reweighting and uncertainty-aware loss within a wavelet-based hybrid network to combat negative transfer and improve cross-environment robustness.

At a representation level, Sharma et al. [31] proposed transforming CSI as a time-frequency image suitable for CNN learning, correctly achieving more than 90% accuracy by leveraging spectral characteristics that are considerably invariant with respect to environmental changes, as shown in Figure 11. It illustrated that various types of falls generate different signatures in the time-frequency domain. This enables the system to learn and classify visual patterns by representing the raw CSI signal as images. This system trained the model with CNN architectures, which successfully improved the ability to generalize from one environment to another with the correct architecture and features.

Self-Supervised Learning (SSL) has also been explored as a means of reducing dependence on large amounts of labeled CSI data [32]. Among the approaches evaluated, Masked Autoencoders (MAE) exhibit strong transfer performance, and encoder choice was key among a number of paradigms. Since MoCo-MAE captures better intrinsic CSI structure during linear evaluations, it offers advantages over ResNet models when only limited data are available for fine-tuning.

At the signal processing level, PhaseAnti [36] focuses on improving robustness to wireless interference by extracting a phase-based feature known as NLPEV, which is designed to be less sensitive to co-channel interference. The experimental result shows that NLPEV remained relatively stable under varying interference conditions and maintained recognition performance in scenarios where several amplitude-based approaches experienced substantial degradation.

Figure 10
Accuracy comparison of CAT, ORI, and SAT models under varying adversarial attack intensities

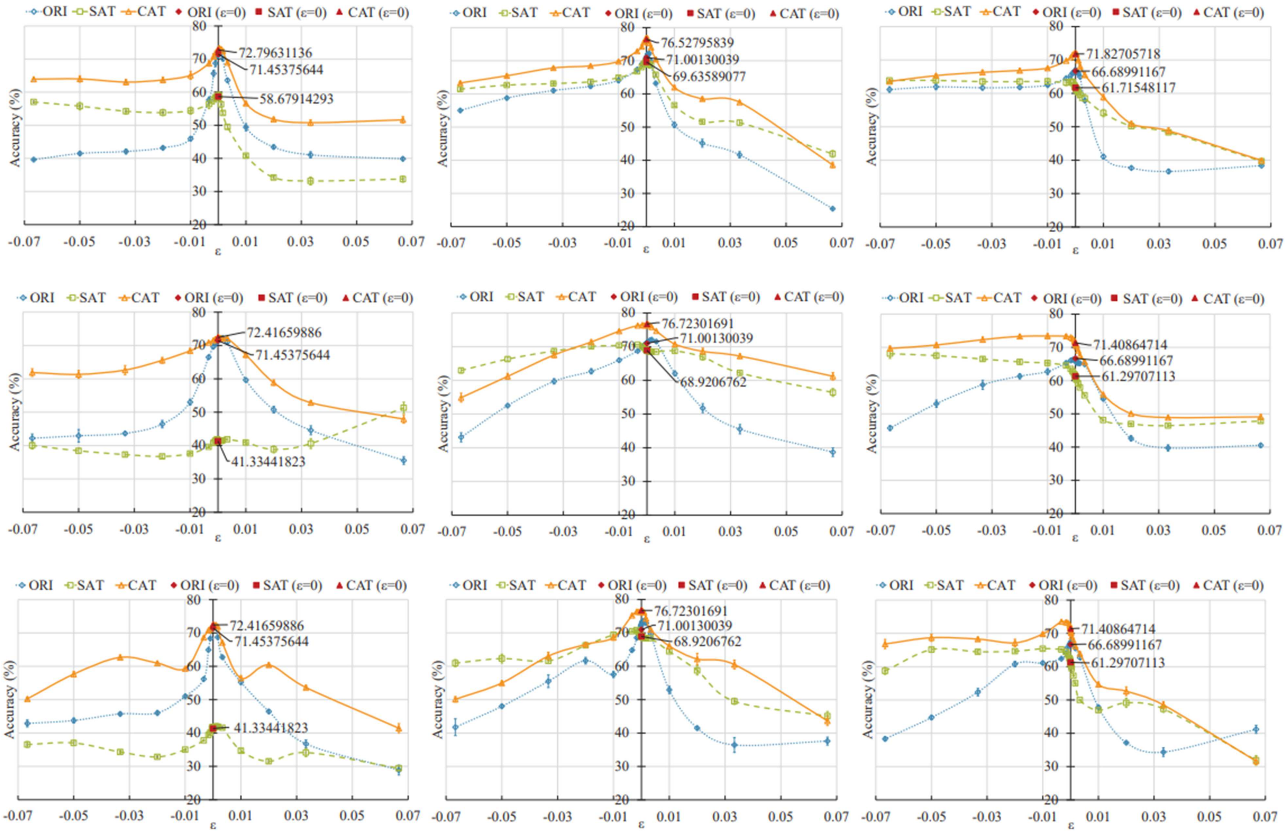
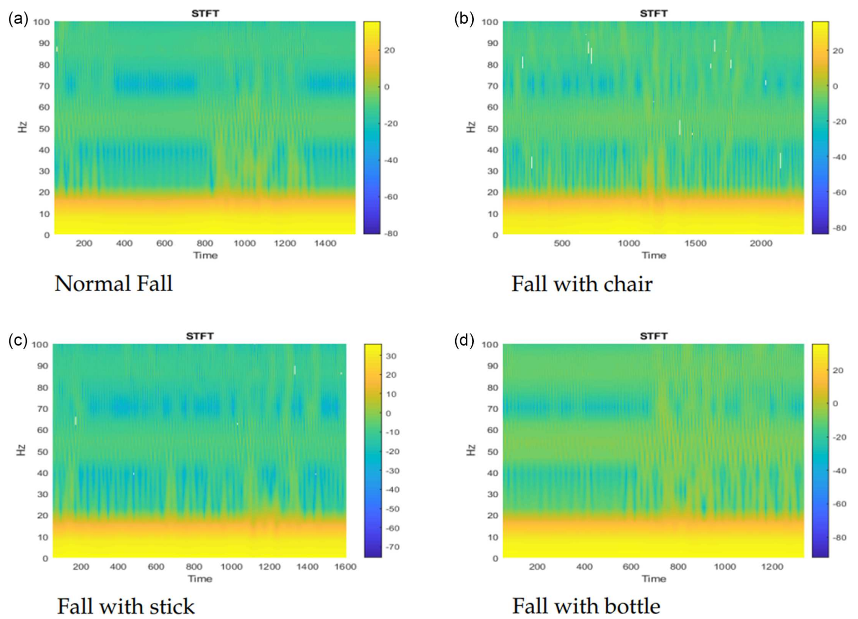


Figure 11
An elderly person falls with different objects



3.4. Evaluation metric

The final aspect of the evaluation protocol relates to the metrics used. Accuracy is the most common metric reported. However, it can be misleading in imbalanced tasks such as fall detection, where false negatives are much more important than false positives. As such, the safety-critical nature of the problem is better characterized by precision, recall, and F1-score.

More recently, authors have also started to report robustness-oriented metrics, which quantify the degree of generalization of the model between domains. Some authors directly report the performance drop between the training and evaluation domains [35]. Additionally, as HAR systems are increasingly deployed to consumer devices, efficiency metrics such as model size, Floating-point Operations (FLOPs), and inference latency are also becoming increasingly relevant. For instance, the Lightweight Attention-GRU [33] reports accuracy in tandem with computational cost.

Table 4 reports the benchmark results for the key papers discussed in this survey. Many architectures report accuracies above 98% on a few public datasets. However, because robustness-oriented metrics are still limited, the generalization problem remains unresolved. Direct comparison across studies is also problematic due to differences in datasets, numbers of participants, and evaluation scenarios. Currently, benchmarking needs to move beyond simply maximizing accuracy and evaluate generalization between domains, robustness, and efficiency.

4. Discussion

While remarkable progress has been shown in the reviewed studies on WiFi-based HAR, a key limitation persists between the performance in a controlled environment and the real world. Our analysis of the literature has concluded several challenges.

4.1. Comparative analysis

As shown in Figure 12, the performance of WiFi CSI-based HAR methods depends on the target application scenario. In HAR benchmarks, such as UT-HAR and NTU-HAR, the latest deep learning methods achieve more than 95% accuracy. Similarly, in fine-grained gesture recognition, methods evaluated on datasets such as SignFi and ARIL achieve high performance. However, the performance degrades in fall detection, where the accuracy lies between 91% and 97%. This difference may partly be due to the use of self-collected datasets for falling scenarios. This illustrates the critical gap in research and the long-standing challenge in achieving high reliability for safety-critical scenarios.

As illustrated in Figure 13, a comparative analysis reveals a significant performance gap between public and self-collected datasets by dataset origin, technique, and performance band. The Sankey diagram visualizes a total of 129 reported results ($N = 129$) extracted from Table 4 across three levels: dataset type (public data vs. self-collected data), technique category (Table 3), and accuracy band, which ranges from excellent (≥95%) to poor (<70%). The flow width corresponds to the number of results, and color encodes the destination performance band. Public benchmarks are largely processed using CNN-based and CNN + RNN hybrid architectures, consistently resulting in "Excellent" accuracy (≥95%). In contrast, self-collected data show a much wider range of results. While domain adaptation is the most commonly applied technique for self-collected data, results still vary wildly, often falling in the bands "Fair" to "Good." This illustrates the "reality gap" in WiFi sensing, where models that perform well on public data often fail to reproduce the high accuracy when used on real-world data.

4.2. Domain shift between training and deployment environments

Environmental variability remains the major challenge in the case of WiFi-based HAR. As WiFi signals are affected by environmental variations, the performance of most models on a single-environment train and test drastically deteriorates when applied to a new environment.

Several papers particularly mentioned this challenge. For instance, Nguyen and Nguyen [44] noted that even with their improvements, the performance was still not suitable for deployment in the real world, highlighting the difficulty in rendering models reliable outside of the lab environment. Another study [5] reported that the model's accuracy varied significantly in a lab, office, and dorm. Some datasets' "data leakage" and "inflated performance metrics" were also highlighted by the paper "RGANet: A Human Activity Recognition Model for Extracting Temporal and Spatial Features from WiFi Channel State Information," which addresses a significant aspect of the domain-shift challenge: poor generalizability [24].

Researchers have employed several solutions to mitigate these challenges, but none are entirely effective. For example, TOSS leveraged a few labeled examples from the new environment to quickly adapt the model. However, this model still depends on collecting some data in the new domain, which is not very realistic [28]. Environmental factors, such as metal furniture and interference from multiple individuals, also deteriorate the accuracy. These problems collectively indicate that making cross-environment performance is still a major problem that must be addressed.

4.3. Hardware limitations and real-time constraints

In the current research stream, the application of high-performance deep learning models is a common approach to yield high accuracy. However, it can be too computationally intensive for real-world deployment. For instance, VGG-16, a well-known CNN architecture, is known to require approximately 14 GB of memory, whereas a common IoT device, such as the Raspberry Pi Model A, typically offers only 56 MB of memory [34].

In response, several researchers primarily focus on lightweight models that perform efficiently without compromising their performance. An approach that focuses on designing a more efficient model using the Lightweight Attention-GRU model by utilizing a network pruning technique to reduce the parameter count by over 99.9% without sacrificing accuracy [33]. However, the real-world deployment of this model poses an issue, as it is incapable of online inference and scalability across different hardware platforms. Additionally, Liang et al. [34] proposed an approach to decentralize the processing by utilizing a distributed system with low-power ESP32 nodes, in combination with a specialized system framework, to process CSI data within the limited memory of Artificial Intelligence of Things (AIoT) devices and perform HAR efficiently. However, this approach struggles with fine-grained activities, such as waving, in an environment with a rich multipath effect. Moreover, ConSense applied a technique that optimizes the training and updating process for models to avoid the high cost of retraining the entire model by applying dynamic model expansion and selective retraining. However, this model requires further testing regarding optimization and real-world implementation [27].

Furthermore, the focus on lightweight models is crucial for real-life applications, such as elder care, which require real-time, continuous monitoring. While most system evaluations are done offline, the application requires immediate data handling and alerts. The system proposed by Ding and Wang [5] addressed this issue by implementing an RNN-based fall detection system to learn temporal sequences, thereby achieving efficient and real-time fall detection in smart homes. Nevertheless, their approach struggles with similar movement between activities, low

Table 4
Summary of reported performance in WiFi-based HAR

References	Performance Metrics	Result		Dataset
[22]	Accuracy over 50 runs with an 80/20 train-test split	Classroom (self-collected)	95.48%	Coarse-grained activities:EfficientFi (NTU-HAR), Office (UT-HAR), Classroom (self-collected)
		EfficientFi (NTU-HAR)	99.91%	
		Office (UT-HAR)	95.80%	
		Home (SignFi)	99.98%	
		Lab (SignFi)	100.00%	
		Home + Lab (SignFi)	99.99%	
	FLOPs	114 subcarriers	157.58 million	Fine-grained sign language activities: Lab, Home, Home + Lab (SignFi)
		56 subcarriers	80 million	
		30 subcarriers	45.22 million	
[26]	Accuracy on 10-fold cross-validation	NTU-HAR	98.75%	Public dataset:
		SD	96.20%	NTU-HAR
	Accuracy on Leave-One-Subject-Out cross-validation	All self-collected datasets	96.19%	Self-collected datasets:
	Accuracy under cross-domain conditions with SD as the training set over 100 runs	SD	~96.2%	Laboratory (source domain, SD),
		CE	~95.6%	Office (cross-environment, CE),
		CL	~95.0%	Laboratory with cross-location (CL),
		CO	~92.6%	Laboratory with cross-orientation (CO),
		CU	~96.0%	Laboratory with cross-user (CU),
		CCD	~90.5%	Office with combined cross-domain conditions (CCD)
	Training Time	SD	105 seconds	
	Inference Speed	SD	1.81 milliseconds/sample	
	Computational Complexity	NTU-HAR	0.49GMac	
	Number of Parameters	NTU-HAR	4.78 million	
[24]	Accuracy	UT-HAR	99.40%	UT-HAR, NTU-FI HAR
		NTU-FI HAR	99.24%	
	Macro Precision	UT-HAR	99.01%	
		NTU-FI HAR	99.28%	

(*Continued*)

Table 4
(*Continued*)

References	Performance Metrics	Result		Dataset
	Macro Recall	UT-HAR	98.93%	
		NTU-FI HAR	99.24%	
	Macro F1-score	UT-HAR	98.91%	
		NTU-FI HAR	99.24%	
	FLOPs	UT HAR	164.61 million	
	Number of Parameters		0.44 million	
	Training Time		18,890.93 seconds	
	Inference Speed		7.60 milliseconds/sample	
[23]	Accuracy using an 80/20 train-test split	UT-HAR	99.40%	UT-HAR, NTU-HAR
		NTU-HAR	99.62%	
	FLOPs	UT-HAR	398.9 million	
		NTU-HAR	434.2 million	
	Inference Speed	UT-HAR	0.435 milliseconds/sample	
		NTU-HAR	0.519 milliseconds/sample	
	Number of Parameters	17.52 million		
[33]	Accuracy	ARIL	98.92%	ARIL, StanWiFi, SignFi, Nexmon HAR
		StanFi	99.33%	
		SignFi (lab)	99.32%	
		SignFi (home)	99.62%	
		SignFi (lab + home)	97.52%	
		SignFi (lab-5)	89.60%	
		Nexmon HAR	100.00%	
	FLOPs	ARIL	10 million	
		StanFi	8.3 million	
		SignFi (lab)	50 million	
		Nexmon HAR	180 million	
	Number of Parameters	ARIL	0.0578 million	
		StanFi	0.0680 million	
		SignFi (lab)	0.2818 million	
		Nexmon HAR	0.1660 million	

(*Continued*)

Table 4
(*Continued*)

References	Performance Metrics	Result		Dataset
[27]	Accuracy using an 80/20 train-test split	WiAR with short task	91.66%	WiAR, MMFi, XRF
		WiAR with a long task	89.85%	
		MMFi with short task	84.42%	
		MMFi with a long task	71.97%	
	Accuracy using an 70/30 train-test split	XRF with short task	66.19%	
		XRF with a long task	65.79%	
	Number of Parameters	WiAR	3.35 million	
		MMFi	1.92 million	
		XRF	1.50 million	
	Forgetting Measure	WiAR with short task	14.31%	
		WiAR with a long task	12.89%	
		MMFi with short task	16.28%	
		MMFi with a long task	17.99%	
		XRF with short task	19.51%	
		XRF with a long task	18.09%	
[21]	Accuracy using an 80/20 train-test split	92.63		Indoor fall detection dataset (self-collected)
	Falling Precision using an 80/20 train-test split	98.10%		
	Falling Recall using an 80/20 train-test split	73.80%		
	Falling Specificity using an 80/20 train-test split	98.10%		
[25]	Accuracy over 20 runs with a 70/30 train-test split	Atomic Activity	97%	Walking-fall sequence dataset (self-collected)
		Compound Activity	93%	
		Unseen Compound Activity	83.80%	
	Precision over 20 runs with a 70/30 train-test split	Atomic Activity	~98%	
		Compound Activity	~96%	
		Unseen Compound Activity	~100%	
[34]	Accuracy using an 80/20 train-test split	91.70%		IoT multi-activity dataset (self-collected)
	Training Time + Testing Time	35.3 seconds		

(*Continued*)

Table 4
(*Continued*)

References	Performance Metrics	Result		Dataset
[30]	Accuracy for the partial domain adaptation (PDA) task	WiHAR	80.80%	WiHAR (self-collected), CSLOS
		CSLOS	52.10%	
[29]	Accuracy on testing set (FGSM attack)	CAT	Outperforms SATs by 10.37% on average	Self-collected dataset: 9 volunteers with varying heights (165 cm to 95 cm) and weights (50 kg to 90 kg), four activities (walking, standing up or sitting down, jumping, and turning around), six different environments (a hall, a corridor, a meeting room, an open platform, and a building entrance)
		SATs	Decrease of 6.37% compared to ORI	
		ORI	39.87%	
	Accuracy on testing set (PGD attack)	CAT	Outperform SATs by 16.33% on average	
		SATs	Decrease of 12.53% compared to ORI	
		ORI	Poor	
	Accuracy on testing set (unattacked)	CAT	Modest improvement compared to ORI	
		SATs	Decreased compared to ORI	
		ORI	71.45%	
[38]	Accuracy of ALSensing	Using 15% unlabelled samples	58.97%	
	Accuracy of BS_Rd	Using 15% unlabelled samples	48.95%	
	Accuracy of BS_All	Using 100% unlabelled samples	62.19%	
[37]	Accuracy after data clustering	DeepConvLSTM Model	91%	Undisclosed self-collected dataset in a laboratory: three volunteers and five activities (fall, boxing, stepping, sit down, and bend down)

(*Continued*)

Table 4
(*Continued*)

References	Performance Metrics	Result		Dataset
[32]	Accuracy on 60/20/20 train-validation-test split for Masked Autoencoders (MAE) with linear classifier	UT-HAR	87.50%	Widar_R2, UT-HAR
		Falldefi	81.57%	
		CSIDA	93.10%	
		Widar_R2	64.29%	
[36]	Recognition Accuracy Rate (RAR)	NLPEV	96.5% on average across different CCI scenarios	Self-collected dataset: eight people, nine activities (standing, walking, running, sitting, push-ups, lifting dumbbells, squats, stoop-down, and sit-ups) and reference state (empty room), these activities are divided into "in-place activities" and "walking activities", collected in three different indoor rooms (a conference room, a student office, and a corridor.)
		Amplitude	Perform poorly in CCI scenarios but perform well in non-interfering scenarios (91.31%)	
		Original Phase	Poor in all scenarios	
		WiFall, WiKey, WiReader	Poor in CCI-scenarios	
		WiAnti	Lower RAR than PhaseAnti in CCI scenarios	
	Recognition Time	PhaseAnti-NLPEV	10.3x faster than WiAnti in CCI scenarios	
		PhaseAnti-Am	0.033s on average in a non-interfering scenario makes it the fastest	
	CCI Detection Accuracy Rate (DAR)	Proposed Algorithm	98% on average and reaches 100% DAR in CCI scenarios	
	Segmentation Accuracy Rate (SAR)	Proposed Algorithm	95.4% on average across all motions	
		Walking	98.20%	
		Running	97.70%	

(*Continued*)

Table 4
(*Continued*)

References	Performance Metrics	Result		Dataset
[28]	Average accuracy in a comprehensive 1-on-1 domain adaptation	TOSS	82.61%	Self-collected dataset: 9 volunteers with various heights (165cm to 195 cm) and various weights (50kg to 90kg), four activities (walking, standing up or sitting down, jumping, and turning around), and six different wireless environments
		TOSS with fine-tuning	82.69%	
		Heterogeneous Domain Adaptation (HDA)	78.27%	
		HDA with fine-tuning	78.17%	
		PACL	78.11%	
		Model-Agnostic Meta-Learning (MAML)	76.85%	
		MAML with fine-tuning	77.27%	
		MatNet	76.35%	
		CNN	74.90%	
		CNN with fine-tuning	74.77%	
		Environment Independent (EI)	74.86%	
		EI with fine-tuning	72.21%	
	Average Accuracy in Multi-source One-shot Domain Adaptation	TOSS	Performs better in some domain adaptation scenarios than state-of-the-art models	
		TOSS with fine-tuning	91.27%	
	Accuracy in the E4 to E6 Domain Adaptation Scenario	TOSS	78.73%	
		TOSS with inner set division	79.87%	
		TOSS with inner set division, and soft label	86.18%	
		TOSS with inner set division, soft label, and hard label	82.17%	
		TOSS with inner set division, and hard label	85.14%	
		TOSS with inner set division, soft label, hard label, and dynamic balance	87.28%	
[45]	Accuracy using 80/20 train-test split	No Compression	98.30%	NTU-HAR
		Compression Rate 67x	98.60%	
		Compression Rate 148x	98.60%	

(*Continued*)

Table 4
(*Continued*)

References	Performance Metrics	Result		Dataset
		Compression Rate 334x	98.30%	
		Compression Rate 763x	98.10%	
		Compression Rate 1781x	98.10%	
	Communication Cost	No Compression	1.368 Mbps	
		Compression Rate 67x	20.418 Kbps	
		Compression Rate 148x	9.243 Kbps	
		Compression Rate 334x	4.096 Kbps	
		Compression Rate 763x	1.792 Kbps	
		Compression Rate 1781x	0.768 Kbps	
[43]	Accuracy on the in-domain testing set using a 90/10 split	Dataset 1	92.70%	Widar3.0
		Dataset 2	92.90%	
	Accuracy on cross-domain testing set for dataset 1	Cross Location	89.70%	
		Cross Orientation	82.60%	
		Cross Environment	92.40%	
		Cross Person	88.90%	
	Area Under ROC Curve (AUROC) for invalid gesture detection	Mean	0.81	
		Variance	3.00E-04	
	Area Under Precision Recall Curve (AUPRC) for invalid gesture detection	Mean	0.79	
		Variance	7.00E-04	
	False Positive Rate at 80% TPR (FPR80) for invalid gesture detection	Mean	0.31	
		Variance	2.00E-03	
	System time overhead	10 Hz BVP frame rate	13 seconds	
		5 Hz BVP frame rate	6.6 seconds	
		3.3 Hz BVP frame rate	6.6 seconds	
		2.5 Hz BVP frame rate	6.6 seconds	
[31]	Accuracy using five-fold cross-validation and testing at the same place as the training	CNN	93.20%	Self-collected dataset (5 volunteers in a conference room with fall, sit-down, walk, and bed activities) and online dataset
		DNN	80.30%	
		SVM	69.50%	

(*Continued*)

Table 4
(*Continued*)

References	Performance Metrics	Result		Dataset
	Accuracy using five-fold cross-validation and testing at a different place from the training	CNN	90.30%	
		DNN	25.30%	
		SVM	21.10%	
[5]	Accuracy using 70/15/15 train-validation-test split	Laboratory	~85%	Smart home activity data (self-collected)
		Office	~87%	
		Dormitory	~89%	
[44]	Accuracy on the testing set	VADA	52.70%	FallDeFi Dataset
		CNN-ADA, ρ=0.001	64.12%	
		CNN-ADA, ρ=0.01	62.45%	
		CNN-ADA, ρ=0.1	59.41%	
		CNN-ADA, ρ=1	56.60%	
		CNN-ADA, ρ=4	41.51%	
		LSTM-ADA, ρ=0.001	40.20%	
		LSTM-ADA, ρ=0.01	40.81%	
		LSTM-ADA, ρ=0.1	51.72%	
		LSTM-ADA, ρ=1	63.15%	
		LSTM-ADA, ρ=4	66.03%	
[41]	Accuracy using 80/20 train-test split	ResNet1D-[1,1,1,1]	88.13%	ARIL
		ResNet1D-[1,1,1,1]+	89.57%	
		ResNet1D-[2,2,2,2]	87.77%	
		ResNet1D-[3,4,6,3]	85.17%	
	Macro Precision	ResNet1D-[1,1,1,1]	89%	
	Macro Recall		88%	
	Macro F1-score		88%	
	Training time + Testing time	ResNet1D-[1,1,1,1]	161 seconds	
		ResNet1D-[2,2,2,2]	240 seconds	
		ResNet1D-[3,4,6,3]	412 seconds	

(*Continued*)

Table 4
(*Continued*)

References	Performance Metrics	Result		Dataset
[42]	Accuracy on the testing set (taken from 3 out of 10 Participants)	K-Nearest Neighbour	90.42%	WiAR
		Naive Bayes	86.74%	
		Random Forest	88.86%	
		Decision Tree	83.28%	
		Support Vector Machine	89.86%	
		CNN	91.12%	
		LSTM	92.35%	
[40]	Accuracy on 5-fold cross-validation	Lab	98.01%	SignFi
		Home	98.91%	
		Lab + Home	94.81%	
	Accuracy on leave-one-subject-out validation	New user with similar data collection dates and settings	76.96%	
		New user with different data collection dates and settings	49.30%	
	Inference speed	0.62ms / sign gesture		
	Training speed	8.28ms / sign gesture		
[39]	Accuracy on the testing set	Random Forest	65%	UT-HAR dataset
		Hidden Markov	73%	
		LSTM	91%	

Figure 12
Performance benchmarking of WiFi CSI-based HAR across varied application scenarios

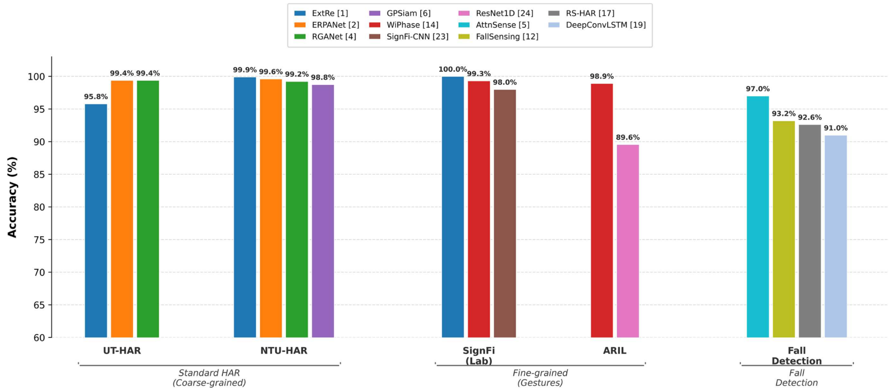
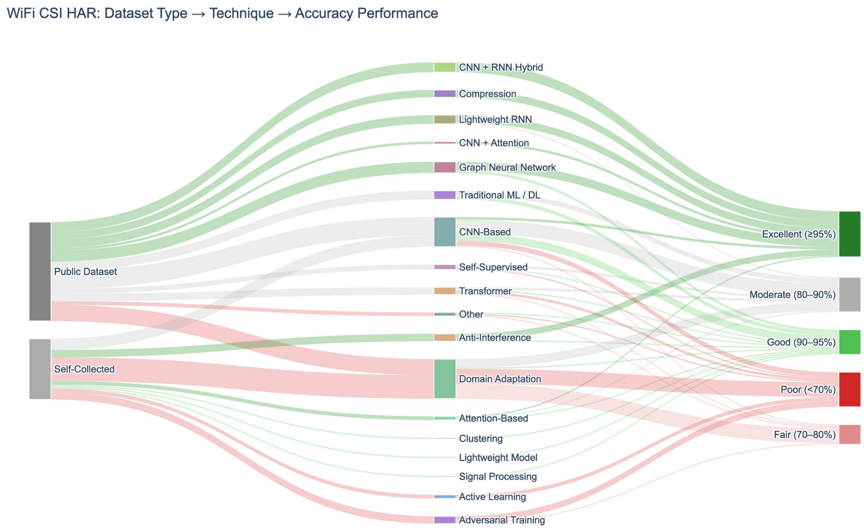


Figure 13
Accuracy distribution flow for WiFi CSI-based HAR



recognition accuracy on bending, and low accuracy in movement when standing.

4.4. Limitations in multi-user recognition

WiFi-based activity recognition in multi-user environments is a challenging task. The reason is that the reflections of radio frequency (RF) signals from multiple users interfere and blend at the receiving antenna. This interference causes the received RF signals to be tangled and hard to separate. Furthermore, the absence of user identification features in commodity WiFi systems makes it challenging to associate RF signals with specific users [49].

Due to the previously mentioned difficulties, most WiFi-based HAR research focuses on designing systems under the assumption of single-user conditions. Prominent datasets, such as Widar3.0, and HAR systems, such as DeFall, conduct their experiments with only one person in the sensing area [2, 43]. Although some studies have attempted to address the multi-user challenge, they often rely on specialized, non-commodity hardware or advanced signal processing that is not yet widely applicable. For instance, MultiTrack utilizes modified hardware to achieve ultra-wideband WiFi (over 600 MHz) for its multi-user activity recognition and joint localization [14]. Therefore, a robust, scalable, and accurate multi-user HAR on commodity WiFi devices remains a significant and open challenge for the research community.

4.5. Public dataset constraints

Publicly available datasets are widely available. However, these datasets are often limited in scale and diversity. As evident from the largest publicly accessible CSI dataset, Widar only comprises 45,170 instances, which is significantly lower than public computer vision datasets [32]. Moreover, current datasets frequently lack diversity in their experimental setups. Even the largest publicly accessible datasets feature a limited number of participants and environments. For instance, there are only 17 users in three rooms from the previously mentioned Widar dataset [32]. The lack of diversity across environmental conditions, user demographics, and action patterns is a major drawback, as it results in trained models that generalize poorly to unseen scenarios [32]. This results in those models that are trained on these datasets performing poorly in practice when deployed in dynamic real-life environments.

To address this, researchers have adopted approaches such as SSL that learn robust features from unlabeled data and allow models to learn from data without depending on labels. However, there is a lack of research on the application of SSL in CSI HAR, and its performance is below the supervised baseline [32]. While the ALSensing approach employs active learning that only annotates the most informative unlabeled samples, the approach still requires human annotation and assistive hardware, which also poses practical and cost challenges [38].

Furthermore, a clear benchmarking gap exists in the surveyed literature, which often does not follow the standardized evaluation protocol. This results in an inability to make objective and consistent comparisons across the different models proposed in the literature. The papers surveyed use a variety of evaluation strategies, including various train-test splits, such as 80/20 [21–25, 27, 34, 41, 45], as well as 70/15/15 [5], and more complex cross-validation methods, such as 10-fold [26], 5-fold [31], and the rigorous Leave-One-Subject-Out [26, 40].

The difficulty of the evaluation and interpretation of the results is significantly impacted by the choice of protocol. For instance, a simple 80/20 split on data collected in a single environment with a few subjects is often the "easiest" test and is highly prone to data leakage, where the test set is overly similar to the training set. Whereas subject-independent tests, such as the Leave-One-Subject-Out (LOSO), are more rigorous for evaluating a model's generalization ability, thus producing a more reliable result. In other words, reaching 90% accuracy with LOSO is arguably more impressive than achieving 98% on a simple 80/20 split. Without a standardized benchmark, results are difficult to compare meaningfully, and the reported numbers lose much of their scientific value.

5. Future Research Directions

5.1. Adaptive domain-transfer methods for location-agnostic models

Strong performance in laboratory conditions often fails when models face a domain shift. Different solutions have been proposed to address this issue and enable the construction of location-agnostic HAR and fall detection systems. For instance, environment-independent fall detection using adversarial augmentation, such as DeFall [2], has improved cross-site robustness. Similarly, data augmentation targeted at environmental variability has enhanced generalization [44]. Cross-adversarial training methods, such as CAT [29], further align feature distributions across domains and achieve higher accuracy under distribution shift.

Semi-supervised domain adaptation has been shown to reduce labeled target data requirements. Uncertainty-guided pseudo-labeling improves one-shot and multi-source adaptation, outperforming baselines [28]. Self-supervised pre-training on unlabeled CSI generates a transferable representation that maintains accuracy even with a few labeled samples [32]. Continual learning with selective retraining incorporates new activities with low forgetting risk and without storing old data (ConSense [27]).

At the same time, signal-level methods have been studied as well. Phase components independent of interference stabilize the input against multipath and cross-device noise [36], which are further enhanced by subcarrier relation modeling on top of phase reconstruction [26]. The time-frequency representations have maintained accuracy between different rooms by focusing on spectral motion signatures [31]. Overall, these findings indicate practical directions to build domain-transfer models: pre-train with SSL [32], simulate environment shift through augmentation [2, 44], align domains during adaptation [28, 29], update compact parts of the model during continual learning [27], and stabilize phase inputs before representation learning [26, 36].

5.2. Standardized benchmarking protocols

Comparison among studies is still limited and challenging. Prior surveys found that datasets, data splits, and metrics are inconsistently used in the literature and have urged common evaluation practices for WiFi sensing with deep learning [49]. Reported accuracies on UT-HAR and NTU-FI HAR using a ResNet-GRU with attention (RGANet) reach 99.4% and 99.24%, respectively [24]. An efficient residual packet attention network (ERPANet) [23] reports accuracies of 99.4% and 99.6% on UT-HAR and NTU-HAR, respectively. An extended temporal-spatial CNN-LSTM (ExtRe) [22] reports 100% on fine-grained datasets and 95.80% in an office-site test. However, these reported accuracies are not directly comparable because the studies differ in task definitions, datasets, and evaluation settings [7, 35].

A minimal set of benchmarks would greatly improve consistency. Tasks should be clearly defined, including whether the study focuses on HAR or fall detection and whether it involves single-user or multi-user activities. Evaluation should include at least one cross-environment test that is distinct from training, such as a shift in room or building, and ideally a device shift as well. Results should report both accuracy and macro-F1, along with a robustness delta between in-room and cross-room or cross-building settings. Claims of edge deployment should also be supported by reporting model size, number of parameters, FLOPs, latency, and energy per inference, as in lightweight real-time studies [5, 33, 34].

Reproducibility can be enhanced by providing fixed train, validation, and test splits, together with preprocessing code and random seeds. Benchmarking should also include comparisons against reference baselines. These baselines should include both Doppler or time-frequency pipelines [21, 31] and modern attention-based CNN and RNN hybrids [23, 24, 49].

To address the critical lack of standardized evaluation, a set of recommended reporting practices was proposed to be adopted by future WiFi-based HAR studies. The heterogeneity of current experimental setups makes it impossible to directly compare results across studies. Thus, future work should explicitly report and control the factors detailed in Table 5. Adopting these reporting guidelines will allow for fair, quantitative comparisons across different architectures and accelerate the transition from laboratory prototypes to reliable real-world deployment.

5.3. Expansion of public datasets

Literature indicates that a lack of large, varied, and standardized publicly available, WiFi-based HAR datasets is preventing fair comparisons and generalizability [49]. Several datasets are commonly used, but high-performing models rely on either private or small datasets [23, 24, 34, 49].

Future datasets should include amplitude and phase data, ideally in raw form, for reconstruction and interference-robust processing [26, 36], multi-user activity traces to evaluate separation in the presence of signal mixing [50], and staged but safe fall events for coverage of safety-critical classes [35]. Environments

Table 5
Recommended reporting practices for WiFi-based HAR studies

Dimension	Reporting requirement	Rationale
Data partitioning	Explicitly distinguish between subject-dependent (random split) and subject-independent (Leave-One-Subject-Out) results	Subject-independent accuracy is a more reliable indicator of real-world generalization
Environmental diversity	Validate models in at least two distinct physical layouts (e.g., furnished vs. empty rooms)	Single-environment results often overestimate performance due to overfitting to specific multipath profiles
Safety-critical metrics	Report false negative rates and recall specifically for fall detection events	In healthcare, a missed fall is significantly more critical than a false alarm
Edge readiness	Report model size (parameters), FLOPs, and inference latency on specified hardware	Essential for assessing practical viability on constrained IoT devices like the ESP32
Signal metadata	Document antenna configuration, sampling rate, channel bandwidth, and room dimensions	Enables controlled ablation studies and facilitates research reproducibility

should reflect deployment conditions with common interference sources, such as metal furniture [36].

Metadata indicating room type and layout, building material type, access point and station models, antenna type, channel bandwidth, sampling rate, and subject count would improve their value further, facilitating controlled ablation studies and cross-project re-use. Dataset providers should also consider hosting a held-out evaluation server with hidden labels, whereby researchers submit predictions to be scored, to avoid overfitting to a single test set. Broader sensing surveys have also emphasized the importance of such practices to foster a stronger benchmarking culture [49].

5.4. Edge-optimized HAR systems

Edge deployment is becoming more practical. Kang et al. [33] showed that an attention-augmented single-layer GRU, pruned after training, was able to maintain high recognition accuracy while decreasing both parameters and latency. Liang et al. [34] showed that an AIoT scenario is supported by distributed ESP32 nodes running lightweight CNNs, showing that low-cost hardware is capable of the task as long as the models are designed to accommodate the resources. Ding and Wang [5] presented real-time fall detection in a smart home scenario using a recurrent network, showing that continuous monitoring is achievable on constrained devices.

Adaptation after installation can be achieved using small and focused updates. This is done by updating normalization layers or inserting small adapter blocks. It avoids the need to fully retrain the model, which mirrors the selective update strategies used in continual WiFi sensing systems such as ConSense [27].

Claims of edge readiness need to be accompanied by thorough reporting. Reports should include end-to-end latency under batch-one and streaming conditions, as well as memory footprint, throughput, and energy per inference on given hardware, in addition to accuracy, and macro-F1 should be provided. Examples of such reporting can be found in prior lightweight and real-time literature [5, 33, 34]. Additional operating information, such as window length, hop size, packet loss rate, and channel settings, should be given such that speed and accuracy can be recreated on comparable devices.

6. Conclusion

This review synthesizes key advancements in WiFi-based HAR and fall detection from 2017 to 2025. While deep learning models have achieved remarkable performance in controlled environments, a critical gap persists. Most of the prior models struggle to generalize beyond their specific training settings. Cross-environment tests show that model accuracy typically declines by around 20% due to domain shifts. This degradation is further exacerbated by environmental interference, such as metal objects and multi-user activity. To bridge this gap, emerging domain adaptation methods (e.g., TOSS and WiCAU), adversarial training, and interference-independent signal preprocessing techniques (e.g., PhaseAnti) offer promising pathways toward location-agnostic systems [28, 30, 36].

As the field transitions toward practical, real-world deployment, addressing hardware limitations through on-device computing is essential. Lightweight models that use compression and pruning (e.g., LiWi-HAR and Lightweight Attention-GRU) demonstrate that computational constraints on edge devices can be managed without severe accuracy penalties [33, 34]. To maintain performance stability despite data scarcity, data efficiency has become a priority. Approaches such as active learning [38] and SSL, particularly MAE [32], have shown promise in learning robust representations from unlabeled data.

Despite these technical strides, the widespread global adoption of WiFi-based HAR is hindered by fragmented evaluation methodologies, small-scale private datasets, and the unresolved complexity of multi-user recognition. Future research must prioritize adaptive domain-transfer techniques, the expansion of diverse public datasets, and strict adherence to standardized reporting practices. By addressing these foundational challenges, the field can successfully transition WiFi sensing from controlled laboratory prototypes to adaptive, location-independent models for healthcare and smart home monitoring.

Funding Support

We acknowledge the financial support of Xiamen University Malaysia Research Fund (XMUMRF) Cycle 14/2024 with the grant number XMUMRF/2024-C14/IECE/0052 for the technical assistance and comparative analysis provided during the preparation of this manuscript.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Author Contribution Statement

Kok Chung Chua: Conceptualization, Methodology, Validation, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. **Kai Liang Lew:** Resources, Writing – original draft, Writing – review & editing, Project administration. **Chean Khim Toa:** Conceptualization, Validation, Resources, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Matthew Alexander Paudianto:** Investigation, Data Curation, Writing– original draft, Writing – review & editing. **Ivan Nathanael:** Investigation, Data curation, Writing – original draft, Writing – review & editing. **Abel Nathanael Hutapea:** Investigation, Data curation, Writing – original draft, Writing – review & editing.

References

[1] Wang, Y., Wu, K., & Ni, L. M. (2017). WiFall: Device-free fall detection by wireless networks. *IEEE Transactions on Mobile Computing*, *16*(2), 581–594. https://doi.org/10.1109/TMC.2016.2557792

[2] Hu, Y., Zhang, F., Wu, C., Wang, B., & Liu, K. J. R. (2022). DeFall: Environment-independent passive fall detection using WiFi. *IEEE Internet of Things Journal*, *9*(11), 8515–8530. https://doi.org/10.1109/JIOT.2021.3116136

[3] Chu, Y., Cumanan, K., Sankarpandi, S. K., Smith, S., & Dobre, O. A. (2023). Deep learning-based fall detection using WiFi channel state information. *IEEE Access*, *11*, 83763–83780. https://doi.org/10.1109/ACCESS.2023.3300726

[4] Chen, T., Li, X., Li, H., & Zhu, G. (2024). Deep learning-based fall detection using commodity Wi-Fi. *Journal of Information and Intelligence*, *2*(4), 355–364. https://doi.org/10.1016/j.jiixd.2024.04.001

[5] Ding, J., & Wang, Y. (2020). A WiFi-based smart home fall detection system using recurrent neural network. *IEEE Transactions on Consumer Electronics*, *66*(4), 308–317. https://doi.org/10.1109/TCE.2020.3021398

[6] Shi, Z., Cheng, Q., Zhang, J. A., & Da Xu, R. Y. (2022). Environment-robust WiFi-based human activity recognition using enhanced CSI and deep learning. *IEEE Internet of Things Journal*, *9*(24), 24643–24654. https://doi.org/10.1109/JIOT.2022.3192973

[7] Wang, W., Liu, A. X., Shahzad, M., Ling, K., & Lu, S. (2015). Understanding and modeling of WiFi signal based human activity recognition. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, 65–76. https://doi.org/10.1145/2789168.2790093

[8] Abuhoureyah, F., Wong, Y. C., & Mohd Isira, A. S. B. (2024). WiFi-based human activity recognition through wall using deep learning. *Engineering Applications of Artificial Intelligence*, *127*, 107171. https://doi.org/10.1016/j.engappai.2023.107171

[9] Gu, Z., He, T., Wang, Z., & Xu, Y. (2022). Device-free human activity recognition based on dual-channel transformer using WiFi signals. *Wireless Communications and Mobile Computing*, *2022*(1), 4598460. https://doi.org/10.1155/2022/4598460

[10] Hussain, A., Chen, Y., Ullah, A., & Zhang, S. (2024). WiSigPro: Transformer for elevating CSI-based human activity recognition through attention mechanisms. *Expert Systems with Applications*, *258*, 124976. https://doi.org/10.1016/j.eswa.2024.124976

[11] Pallaprolu, A., Korany, B., & Mostofi, Y. (2023). Analysis of Keller cones for RF imaging. In *2023 IEEE Radar Conference*, *1–6*. https://doi.org/10.1109/RadarConf2351548.2023.10149785

[12] Li, B., Ren, Y., Wang, Y., & Yang, J. (2024). SpaceBeat: Identity-aware multi-person vital signs monitoring using commodity WiFi. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *8*(3), 1–23. https://doi.org/10.1145/3678590

[13] Zhuang, W., Lu, Y., Shen, Y., & Su, J. (2024). Human daily breathing monitoring via analysis of CSI ratio trajectories for WiFi link pairs on the I/Q plane. *Sensors*, *24*(22), 7352. https://doi.org/10.3390/s24227352

[14] Abuhoureyah, F., Sim, K. S., & Wong, Y. C. (2024). Multi-user human activity recognition through adaptive location-independent WiFi signal characteristics. *IEEE Access*, *12*, 112008–112024. https://doi.org/10.1109/ACCESS.2024.3438871

[15] He, M. H., & Toa, C. K. (2024). A review of sensor-based gait recognition. In *2024 International Conference on Computing Innovation, Intelligence, Technologies and Education*, 1–6. https://doi.org/10.1109/CIITE62244.2024.10987605

[16] Zhao, L., Xiao, R., Liu, J., & Han, J. (2024). One is enough: Enabling one-shot device-free gesture recognition with COTS WiFi. In *IEEE Conference on Computer Communications*, 1231–1240. https://doi.org/10.1109/INFOCOM52122.2024.10621091

[17] Yin, Y., Zhang, Z., Yang, X., Yan, F., & Niu, Q. (2021). Towards fully domain-independent gesture recognition using COTS WiFi device. *Electronics Letters*, *57*(5), 232–234. https://doi.org/10.1049/ell2.12097

[18] Feng, C., Wang, N., Jiang, Y., Zheng, X., Li, K., Wang, Z., & Chen, X. (2022). Wi-learner: Towards one-shot learning for cross-domain Wi-Fi based gesture recognition. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *6*(3), 114. https://doi.org/10.1145/3550318

[19] Niu, K., Zhang, F., Wang, X., Lv, Q., Luo, H., & Zhang, D. (2022). Understanding WiFi signal frequency features for position-independent gesture sensing. *IEEE Transactions on Mobile Computing*, *21*(11), 4156–4171. https://doi.org/10.1109/TMC.2021.3063135

[20] Chen, J., Bi, S., Lin, X.-H., & Quan, Z. (2024). LAGER: Label-free domain-adaptive wireless gesture recognition via latent feature alignment and augmentation. *IEEE Internet of Things Journal*, *11*(23), 37928–37941. https://doi.org/10.1109/JIOT.2024.3445960

[21] Cárdenas, J. D., Gutiérrez, C. A., & Aguilar-Ponce, R. M. (2024). Fall detection using WiFi signals with Doppler frequency diversity. *IEEE Consumer Electronics Magazine*, *13*(5), 75–82. https://doi.org/10.1109/MCE.2023.3330302

[22] Wang, P., Huang, H., Zhao, L., Zhu, B., Huang, H., & Wu, H. (2025). ExtRe: Extended temporal-spatial network for consumer-electronic WiFi-based human activity recognition. *IEEE Transactions on Consumer Electronics*, *71*(1), 230–238. https://doi.org/10.1109/TCE.2024.3435881

[23] Yang, S., Yang, J., Yang, C., Yan, W., & Li, P. (2025). Human activity recognition using WiFi signal features and efficient residual packet attention network. *IEEE Sensors Letters*, *9*(4), 6004304. https://doi.org/10.1109/LSENS.2025.3551337

[24] Hu, J., Ge, F., Cao, X., & Yang, Z. (2025). RGANet: A human activity recognition model for extracting temporal and spatial features from WiFi channel state information. *Sensors*, *25*(3), 918. https://doi.org/10.3390/s25030918

[25] Xiao, Y., Zhu, J., Zhang, S., Liu, X., & Guo, S. (2024). Fall-attention: An attention-based fall detection method for adjoint activities. *IEEE Transactions on Mobile Computing*, *23*(7), 7895–7909. https://doi.org/10.1109/TMC.2023.3344125

[26] Chen, X., Li, C., Jiang, C., Meng, W., & Xiao, W. (2025). WiPhase: A human activity recognition approach by fusing of reconstructed WiFi CSI phase features. *IEEE Transactions on Mobile Computing*, *24*(1), 394–406. https://doi.org/10.1109/TMC.2024.3461672

[27] Li, R., Deng, T., Feng, S., Sun, M., & Jia, J. (2025). ConSense: Continually sensing human activity with WiFi via growing and picking. In *Proceedings of the AAAI Conference on Artificial Intelligence*, *39*(13), 14292–14300. https://doi.org/10.1609/aaai.v39i13.33565

[28] Zhou, Z., Wang, F., Yu, J., Ren, J., Wang, Z., & Gong, W. (2022). Target-oriented semi-supervised domain adaptation for WiFi-based HAR. In *IEEE Conference on Computer Communications*, 420–429. https://doi.org/10.1109/INFOCOM48880.2022.9796782

[29] Liu, L., Zhou, Z., Chen, M., & Gong, W. (2024). CAT: Cross-adversarial training for WiFi-based human activity recognition. In *2024 27th International Conference on Computer Supported Cooperative Work in Design*, 1023–1028. https://doi.org/10.1109/CSCWD61410.2024.10580434

[30] Cui, W., Wu, K., Wu, M., Li, X., & Chen, Z. (2024). WiCAU: Comprehensive partial adaptation with uncertainty-aware for WiFi-based cross-environment activity recognition. *IEEE Transactions on Instrumentation and Measurement*, *73*, 5503110. https://doi.org/10.1109/TIM.2024.3398094

[31] Sharma, L., Chao, C.-H., Wu, S.-L., & Li, M.-C. (2021). High accuracy WiFi-based human activity classification system with time-frequency diagram CNN method for different places. *Sensors*, *21*(11), 3797. https://doi.org/10.3390/s21113797

[32] Xu, K., Wang, J., Zhu, H., & Zheng, D. (2025). Evaluating self-supervised learning for WiFi CSI-based human activity recognition. *ACM Transactions on Sensor Networks*, *21*(2), 1–38. https://doi.org/10.1145/3715130

[33] Kang, H., Kim, D., & Toh, K.-A. (2025). Human activity recognition through augmented WiFi CSI signals by lightweight attention-GRU. *Sensors*, *25*(5), 1547. https://doi.org/10.3390/s25051547

[34] Liang, W., Tang, R., Jiang, S., Wang, R., Zhao, Y., Xu, C.-Z., & Li, X. (2024). LiWi-HAR: Lightweight WiFi-based human activity recognition using distributed AIoT. *IEEE Internet of Things Journal*, *11*(1), 597–611. https://doi.org/10.1109/JIOT.2023.3286455

[35] Hu, Y., Zhang, F., Wu, C., Wang, B., & Liu, K. R. (2020). A WiFi-based passive fall detection system. In *2020 IEEE International Conference on Acoustics, Speech and Signal Processing*, 1723–1727. https://doi.org/10.1109/ICASSP40776.2020.9054753

[36] Huang, J., Liu, B., Miao, C., Lu, Y., Zheng, Q., Wu, Y., & Chen, C. W. (2023). PhaseAnti: An anti-interference WiFi-based activity recognition system using interference-independent phase component. *IEEE Transactions on Mobile Computing*, *22*(5), 2938–2954. https://doi.org/10.1109/TMC.2021.3127721

[37] Xia, Z., & Chong, S. (2023). WiFi-based indoor passive fall detection for medical Internet of Things. *Computers and Electrical Engineering*, *109*, 108763. https://doi.org/10.1016/j.compJ.eleceng.2023.108763

[38] Zhao, G., Zhou, Z., Huang, Y., Nayak, A., Gong, W., & Zhou, H. (2023). ALSensing: Human activity recognition using WiFi based on active learning. In *2023 IEEE Wireless Communications and Networking Conference*, 1–6. https://doi.org/10.1109/WCNC55385.2023.10119036

[39] Yousefi, S., Narui, H., Dayal, S., Ermon, S., & Valaee, S. (2017). A survey on behavior recognition using WiFi channel state information. *IEEE Communications Magazine*, *55*(10), 98–104. https://doi.org/10.1109/MCOM.2017.1700082

[40] Ma, Y., Zhou, G., Wang, S., Zhao, H., & Jung, W. (2018). SignFi: Sign language recognition using WiFi. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *2*(1), 1–21. https://doi.org/10.1145/3191755

[41] Wang, F., Feng, J., Zhao, Y., Zhang, X., Zhang, S., & Han, J. (2019). Joint activity recognition and indoor localization with WiFi fingerprints. *IEEE Access*, *7*, 80058–80068. https://doi.org/10.1109/ACCESS.2019.2923743

[42] Guo, L., Wang, L., Lin, C., Liu, J., Lu, B., Fang, J., & Guo, S. (2019). WiAR: A public dataset for WiFi-based activity recognition. *IEEE Access*, *7*, 154935–154945. https://doi.org/10.1109/ACCESS.2019.2947024

[43] Zhang, Y., Zheng, Y., Qian, K., Zhang, G., Liu, Y., Wu, C., & Yang, Z. (2022). Widar3.0: Zero-effort cross-domain gesture recognition with Wi-Fi. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(11), 8671–8688. https://doi.org/10.1109/TPAMI.2021.3105387

[44] Nguyen, T.-D. H., & Nguyen, H.-N. H. (2020). Towards a robust WiFi-based fall detection with adversarial data augmentation. In *2020 54th Annual Conference on Information Sciences and Systems*, 1–6. https://doi.org/10.1109/CISS48834.2020.1570617398

[45] Yang, J., Chen, X., Zou, H., Wang, D., Xu, Q., & Xie, L. (2022). EfficientFi: Toward large-scale lightweight WiFi sensing via CSI compression. *IEEE Internet of Things Journal*, *9*(15), 13086–13095. https://doi.org/10.1109/JIOT.2021.3139958

[46] Yang, J., Huang, H., Zhou, Y., Chen, X., Xu, Y., Yuan, S., . . . , & Xie, L. (2023). Mm-Fi: Multi-modal non-intrusive 4D human dataset for versatile wireless sensing. In *Proceedings*

of the 37th International Conference on Neural Information Processing Systems, 18756–18768.

[47] Al-Rajab, M., Qassem, K., Seyam, S., Al Hamadi, Y., Al-Dubaisi, A. H., Mallat, J., & Alzraiqat, S. (2026). Artificial intelligence-enhanced CSI-based Wi-Fi sensing for non-contact vital sign monitoring: A systematic review. *PeerJ Computer Science*, *12*, e3375. https://doi.org/10.7717/peerj-cs.3375

[48] Wang, F., Zhang, T., Xi, W., Ding, H., Wang, G., Zhang, D., . . . , & Han, T. X. (2026). A survey on Wi-Fi sensing generalizability: Taxonomy, techniques, datasets, and future research prospects. *IEEE Communications Surveys & Tutorials*, *28*, 5227–5266. https://doi.org/10.1109/COMST.2026.3670854

[49] Ahmad, I., Ullah, A., & Choi, W. (2024). WiFi-based human sensing with deep learning: Recent advances, challenges, and opportunities. *IEEE Open Journal of the Communications Society*, *5*, 3595–3623. https://doi.org/10.1109/OJCOMS.2024.3411529

[50] Moghaddam, M. G., Shirehjini, A. A. N., & Shirmohammadi, S. (2023). A WiFi-based method for recognizing fine-grained multiple-subject human activities. *IEEE Transactions on Instrumentation and Measurement*, *72*, 2520313. https://doi.org/10.1109/TIM.2023.3289547

How to Cite: Chua, K. C., Lew, K. L., Toa, C. K., Paudianto, M. A., Nathanael, I., & Hutapea, A. N. (2026). WiFi-Based Human Activity Recognition and Fall Detection with Taxonomy, Benchmarks, and Future Directions: A Narrative Review. *Artificial Intelligence and Applications*. https://doi.org/10.47852/bonviewAIA62027517