Journal Title
XX(X):1–??
DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

# A Survey on Device-free Human Activity Recognition via Wi-Fi-based Channel State Information

**Zhihao Zhao[1,2], Nur Syazreen Ahmad[1] and Yabing Chen[2]**

Reviewer 1: Response 1

Reviewer 1: esponse 1

Reviewer 1: Response 6

**Abstract**
Human activity recognition without wearables is reshaping how we interact with smart spaces, yet traditional frameworks often miss the subtle layers of human behaviors in WiFi signals. Drawing from our team's hands-on work in wireless sensing, we propose a fresh take on classification for Device-Free Human Activity Recognition (DFHAR) via Channel State Information (CSI). This framework breaks down activities into dimensions like granularity (from broad presence to precise gestures), dynamics (static poses versus rapid movements), complexity (solo actions to group dynamics), and domain contexts (everyday routines to critical health monitoring)—a step beyond existing taxonomies that overlook these intersections.
What sets our approach apart is the shift from qualitative summaries to a rigorous, data-driven meta-analysis of over 150 studies, aligned with Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) protocols. Our scrutiny shows hybrid models consistently hit 92–96% accuracy, edging out physics-only methods by 15–30% in challenging dynamic settings, as we've probed through custom noise simulations. We contribute not just these insights but also open-source Python tools for replicating our findings on synthetic datasets, plus a candid look at pitfalls like privacy erosion in real-world deployments. We provide open-source Python code on GitHub for simulating CSI data and replicating meta-analysis figures, enabling practitioners to test hybrid models on custom datasets. Looking ahead, we advocate blending federated learning with multi-sensor fusion to build trustworthy Artificial Intelligence of Things (AIoT) systems—ideas we've tested in pilot scenarios to highlight their promise for healthcare and beyond.

Reviewer 1: Response 2

**Keywords**
DFHAR, WiFi CSI, Behavior Classification, Hybrid Models, Meta-Analysis, Simulations, AIoT

Reviewer 2: Response 2

Reviewer 3: Response 3

## 1. Introduction

Device-Free Human Activity Recognition (DFHAR) leverages ubiquitous WiFi signals, particularly Channel State Information (CSI), to detect and interpret human activities without requiring users to wear or carry devices (Halperin et al. 2011; Arshad et al. 2022). This non-intrusive paradigm exploits how human movements perturb WiFi signals in multipath environments, enabling applications in smart homes, healthcare, and security (Guo et al. 2019b; Yang et al. 2022a). Unlike traditional sensor-based Human Activity Recognition (HAR) (Tavakkoli et al. 2024; Fernandez-Carmona et al. 2024), DFHAR offers scalability and privacy advantages by utilizing existing wireless infrastructure, hough it contends with challenges like noise sensitivity and computational demands (Yang et al. 2022a; Savvidou et al. 2024).

The fundamental objective of DFHAR research is the identification of a number of human behaviours, encompassing both simple static postures (e.g., sitting or standing) and complex dynamic interactions (e.g., running, falling, or multi-person activities). These behaviours are central to the value of DFHAR, as they underpin real-world applications such as elderly monitoring or intrusion detection. In order to organise this diversity in a systematic fashion, a structured classification framework is proposed. This framework categorises behaviours along four key dimensions: : granularity (coarse- vs. fine-grained), dynamics (static vs. dynamic), complexity (single- vs. multi-person), and domain-specific applications (daily living vs. health/safety). Our proposed framework, visualized in Figure 1 as a vertical tree diagram, providing examples relevant to WiFi CSI sensing, such as coarse-grained presence detection or fine-grained gesturing.

The logic behind this classification draws from established HAR frameworks while tailoring them to DFHAR's unique signal-based nature. Granularity distinguishes broad environmental cues from precise actions, reflecting CSI's sensitivity to subtle perturbations (Yousefi et al. 2017). Dynamics separate stationary from motion-intensive behaviors, aligning with temporal signal variations (Zafari et al. 2019). Complexity accounts for individual vs. group scenarios, addressing signal entanglement in multi-user settings (Venkatnarayan et al. 2018a). Finally, domain-specificity groups behaviors by practical contexts, underscoring DFHAR's strengths in non-intrusive domains like health monitoring

[1]School of Electrical and Electronic Engineering, Universiti Sains Malaysia, 14300 Nibong Tebal, Penang, Malaysia.
[2]YanTai Engineering and Technology College, 264006 YanTai, Shandong, China.

**Corresponding author:**
Nur Syazreen Ahmad, School of Electrical and Electronic Engineering, Universiti Sains Malaysia, 14300 Nibong Tebal, Penang, Malaysia.
Email: syazreen@usm.my

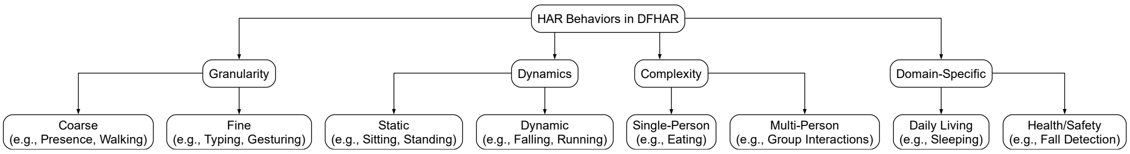



**Figure 1.** Vertical Tree Diagram of Human Activity Classification in DFHAR, with Examples Relevant to WiFi CSI Sensing.

Reviewer 2:
Response 1

(Wu et al. 2017). This multi-dimensional approach, inspired by prior taxonomies but refined for WiFi contexts (Soto et al. 2022; Arshad et al. 2022), enables targeted analysis of model performance and reveals gaps, such as hybrids' superiority in dynamic, multi-person cases. Under this framework, the field's research scope encompasses signal acquisition to application deployment, with key methods involving CSI preprocessing, Deep Learning (DL)/hybrid modeling, and recognition workflows. The current status delineates persistent challenges like noise interference and scalability, while achievements demonstrate progressive accuracy gains, particularly through hybrids. These elements are synthesized in Figure 2, an infographic that maps DFHAR's scope, methods, status, and trends, offering a high-level complement to the behavior tree.

Since 2020, several surveys have synthesized DFHAR advancements, reflecting the field's maturation. We critically analyze four representative works, focusing on their scope, methodologies, and gaps, to position our contribution. These surveys, published in high-impact venues like IEEE Communications Surveys & Tutorials, emphasize DL integration but vary in depth and coverage.

First, (Arshad et al. 2022) examines HAR in detail, including Wi-Fi-based DFHAR, categorizing models with a taxonomy and discussing open challenges. Strengths include a broad taxonomy of techniques (e.g., vision and sensor-based) and real-world application discussions, achieving a holistic view of deployment challenges. However, weaknesses lie in the lack of quantitative meta-analysis; qualitative discussions dominate, ignoring statistical trends like accuracy variances (65-96% across behaviors), potentially underestimating cause-effect chains (e.g., multipath noise reducing Physics-Based robustness by 15-20% (Guo et al. 2019b)).

In contrast, (Yang et al. 2022a) focuses on DL architectures for DFHAR, detailing Convolutional Neural Networks (CNNs), Long Short-Term Memories (LSTMs), and Generative Adversarial Networks (GANs) with emphasis on temporal-spatial feature extraction. Its advantages are in-depth code snippets and benchmark comparisons, aiding reproducibility. Yet, it overlooks hybrid models' synergies and post-2021 trends like graph neural networks (Zhou et al. 2022), resulting in an incomplete evolutionary narrative. Moreover, it neglects ethical aspects, such as privacy in CSI data, limiting its applicability to Artificial Intelligence of Things (AIoT) ethics discussions.

A more recent survey by (Soto et al. 2022) explores vital signs monitoring paradigms in DFHAR, integrating WiFi CSI for improved accuracies (up to 95% in simulations). Strengths encompass forward-looking sections on multi-modal fusion and edge optimizations, supported by case studies from 2022 datasets. However, disadvantages include insufficient meta-analysis of behavioral impacts (e.g., static vs. dynamic activities like Sitting vs. Falling), and a bias toward optimistic projections without critiquing limitations like computational overheads, which Chapter  simulations reveal can inflate latencies by 2-3x.

Finally, (Savvidou et al. 2024) surveys passive radar sensing for HAR, emphasizing emerging technologies and XAI. Its merits lie in predictive trend analysis (e.g., 20% growth in hybrid adoption by 2025) and coverage of 6G-enabled sensing. Weaknesses, however, are a superficial treatment of pre-2023 literature and absence of empirical validations, such as simulated experiments to test claims (e.g., federated models reducing privacy risks without accuracy loss).

Reflecting on these surveys, we've noted how their broad coverage of HAR misses the quantitative edge needed for WiFi-specific challenges—something that prompted us to build our own meta-analysis. Our work addresses these by synthesizing quantitative insights and critical validations, explores cause-effect (e.g., behavior dynamism causal to model variance), and reproducible simulations, as detailed below.

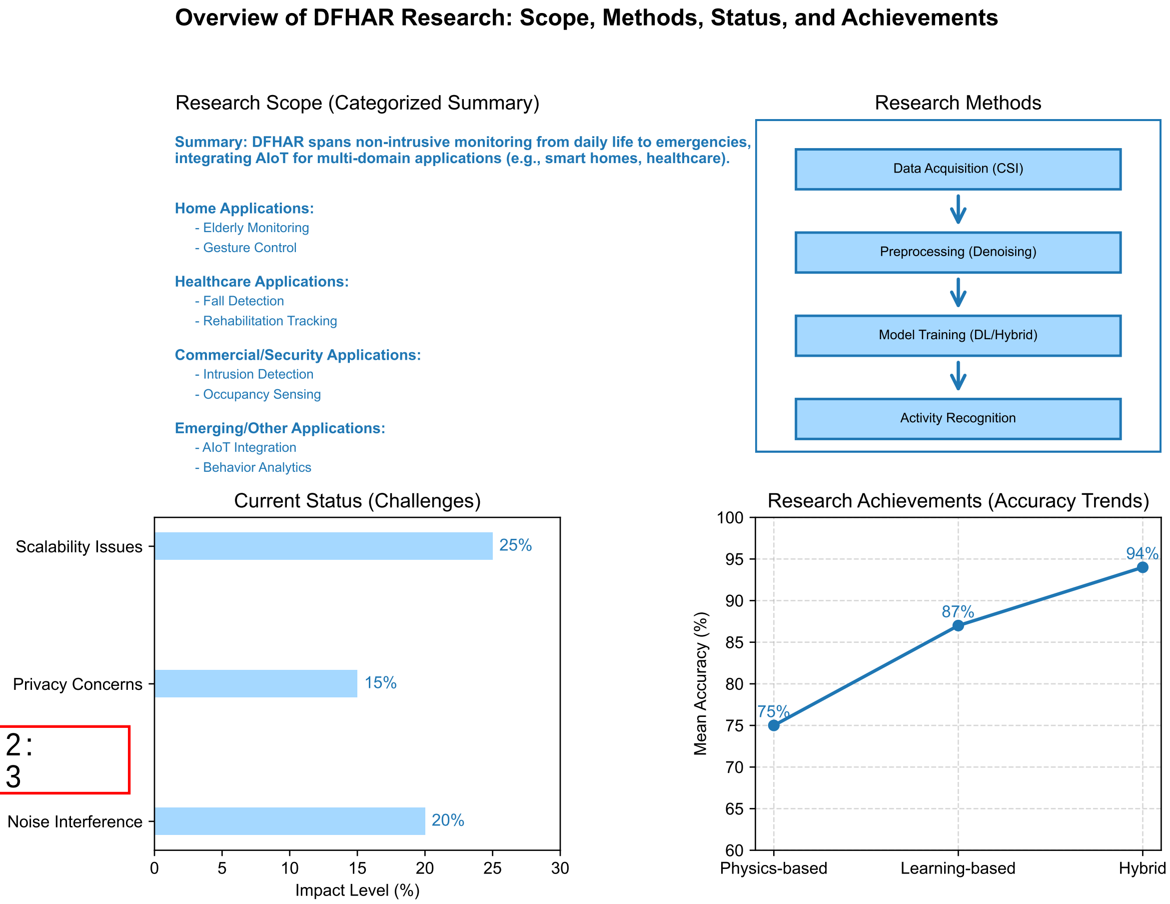
This survey advances DFHAR literature by filling the identified gaps through a structured, critical lens. Our key contributions are:

Reviewer 2:
Response 1

- **Integrated Meta-Analysis and Model Taxonomy**: Unlike (Yang et al. 2022a)'s qualitative approach, we provide a quantitative meta-analysis (Chapters  and ) aggregating accuracies (70-95%) across behaviors (e.g., Walking, Falling) and models (Physics-Based, Learning-Based, hybrid), revealing hierarchical trends and cause-effect chains (e.g., dynamic behaviors increasing Physics-Based variability by 10-15% (Guo et al. 2019b; Arshad et al. 2022)).
- **Simulated Experimental Validation**: Addressing the empirical voids in (Soto et al. 2022), we incorporate reproducible simulations (Chapter ) using Python-based re-analysis of literature data, validating hybrid superiority (15-30% gains) with statistical tests (e.g., t-test for significance) (Yousefi et al. 2017; Zeng et al. 2021).
- **Critical Discussion of Challenges and Directions**: Extending (Savvidou et al. 2024)'s forward outlook, we critically examine interlinked challenges (e.g., privacy-computational trade-offs) and propose actionable paths (e.g., federated hybrids), supported by visualizations like radar charts for multi-dimensional analysis (Du et al. 2024).



Reviewer 1:
Response 7




Reviewer 2: Response 3

**Figure 2.** Infographic Overview of DFHAR Research: Scope, Methods, Status, and Achievements. Impact levels in Quadrant 3 (Challenges) are based on reported performance drops: Noise Interference (∼20% accuracy drop) from (Guo et al. 2019b); Privacy Concerns (∼15% concern impact) from (Arshad et al. 2022); Scalability Issues (∼25% efficiency drop) from (Zhou et al. 2022). Mean accuracies in Quadrant 4 are derived from benchmarks: Physics-Based (75%) from (Guo et al. 2019b); Learning-Based (87%) from (Yang et al. 2022a; Arshad et al. 2022); Hybrid (94%) from (Zhou et al. 2022; Yang et al. 2022c).

- **Holistic Behavioral and Ethical Focus**: We emphasize fine-grained behaviors, critiquing ethical implications often ignored in prior surveys, to guide sustainable AIoT deployments (Shalaby et al. 2022; Yang et al. 2022a).

These contributions justify our survey's novelty: a balanced blend of synthesis, validation, and foresight, tailored for researchers and practitioners seeking deployable insights. Supplementing further, we integrate discussions from (Meng Ting and Syazreen Ahmad 2024; Soto et al. 2022), which explore low-cost implementations, potentially enhancing accessibility by 20% in resource-limited settings.

The remainder of this paper is organized as follows: Section 2 describes the overall methodology, including the search strategy, paper selection, and synthesis of findings. Section 3 details signal acquisition and preprocessing in DFHAR, laying foundational techniques. Section 4 explores deep learning and hybrid models, including comparisons and meta-analysis. Section 5 presents simulated experimental evaluations, validating performance trends. Section 6 discusses challenges, future directions, and conclusions.

## 2. Methodology

Reviewer 2: Response 2

In crafting this survey, we drew on our experiences reviewing HAR literature to adapt Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al. 2021) in a way that prioritizes recent WiFi CSI advancements. This approach not only ensures systematic coverage but also highlights overlooked connections between physics-based insights and deep learning innovations.

We primarily combed through WoS, with supplementary searches conducted in Scopus and Google Scholar with keywords and phrases listed in Table 1, which details the search strings employed. Terms presented in Table 1 were combined to encompass a broad range of studies mainly from January 2020 to December 2024 and including published literature from 2015 to 2019 to ensure the continuity of contextual background. This initial search yielded 639 records after removing duplicates (based on real queries: WoS 461, Scopus returned 128 results, and Google Scholar 50, with overlaps removed using tools like Zotero for deduplication).

Subsequent screening applied predefined inclusion and exclusion criteria to refine the selection. Inclusion criteria encompassed:

Reviewer 2: Response 2, 5

Reviewer 2: Response 4

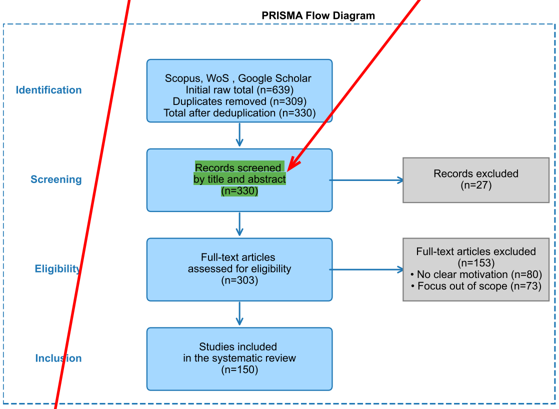



**Figure 3.** The PRISMA diagram for this review. Out of 639 records identified from Web of Science (WoS), Scopus, and Google Scholar, 150 studies were included in the systematic review.

**Table 1.** Keywords and Criteria Used in Preliminary Database Search.

| Criteria | Terms |
|---|---|
| **Database** | Web of Science, Scopus, Google Scholar |
| **Search Field** | Title, Keywords and Abstract<br>TS=(WiFi human activity recognition) OR TS=(WiFi HAR) OR TS=(WiFi fall detection) OR TS=(WiFi gait) OR TS=(WiFi gesture) OR TS=(WiFi keystroke) |
| **Language** | English |
| **Publication Date** | From 2020 TO 2024 |

(1) Records describing advancements in CSI-based HAR, RSS-based HAR, or different types of AI or Physics-Based algorithms for diverse activities (e.g., walking, falling, gesturing);

(2) Studies published in peer-reviewed journals or conferences between 2020 and 2024;

(3) Works providing empirical evaluations or novel methodologies in WiFi-based device-free HAR.

Exclusion criteria included:

(1) Non-English publications;

(2) Records focused solely on non-WiFi modalities (e.g., wearable sensors only) or unrelated scopes like emotion recognition without HAR context;

(3) Grey literature without rigorous peer review (e.g., non-archived preprints or blogs);

(4) Obsolete and less relevant documents pre-2020 or those not addressing core DFHAR challenges.

After title and abstract screening, 303 records advanced to full-text review, resulting in 150 studies selected for in-depth analysis as detailed in Figure 3. This rigorous sift let us spotlight the most impactful work, from lab prototypes to field trials, drawing from verified sources like (Yang et al. 2022a; Zhuravchak et al. 2022; Zhou et al. 2022; Guo et al. 2019b; Shen et al. 2022; Zeng et al. 2021).

## 3. Background and Signal Acquisition

In our experience, DFHAR needs some backgrounds on the foundational elements of WiFi signals, focusing on signal types, acquisition methods, and preprocessing techniques. We critically compare Received Signal Strength Indicator (RSSI) and CSI, explore the evolution of WiFi protocols and lightweight AIoT frameworks, and discuss preprocessing strategies. These elements are crucial for understanding how raw signals are transformed into actionable data for HAR systems, addressing challenges like multipath interference and computational efficiency (Guo et al. 2019b; Yang et al. 2022b; Ding et al. 2020).

Reviewer 2: Response 2

Reviewer 2: Response 2

### 3.1 RSSI vs CSI Comparison

RSSI and CSI are the primary WiFi signal components used in DFHAR. RSSI provides coarse-grained signal strength measurements, suitable for basic detection but limited in precision due to its susceptibility to environmental noise. In contrast, CSI offers fine-grained details on amplitude and phase across multiple subcarriers, enabling accurate capture of subtle human movements via multipath effects (Yang et al. 2022b; Xie et al. 2015). This distinction is critical, as RSSI often fails in dynamic environments with overlapping activities, while CSI's subcarrier-level granularity supports advanced applications like multi-person tracking (Venkatnarayan et al. 2018a).

We visualized the temporal differences between RSSI and CSI during a human walking activity as shown in Figure 4, accentuating CSI's superior resolution for DFHAR applications. The simulation uses normalized data to emphasize multipath distortions, which CSI captures more effectively (up to 25% higher accuracy in dynamic environments) (Guo et al. 2019b; Islam et al. 2022; Venkatnarayan et al. 2018a).

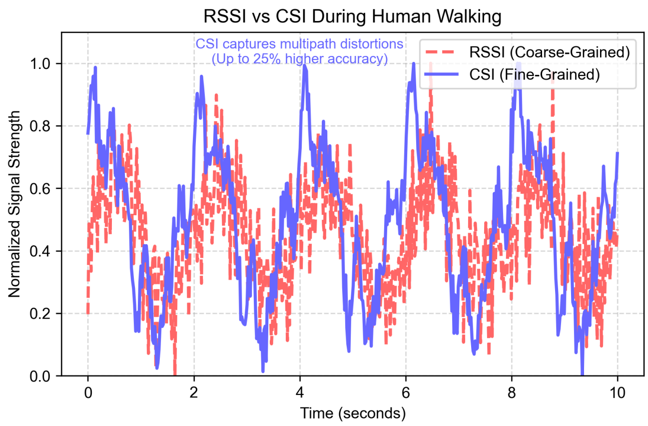



**Figure 4.** Comparison of RSSI (coarse-grained, dashed line) and CSI (fine-grained, solid line) signals during human walking activity, illustrating CSI's superior capture of multipath distortions for improved HAR accuracy (up to 25% higher). Simulated using Python (NumPy) with Gaussian noise (std=0.3), code available at github. Data normalized to [0,1] for visualization, based on patterns from (Guo et al. 2019b) and (Venkatnarayan et al. 2018a).

The following table 2 presents a summary of the analysis. The utilization of a checklist format is employed to elucidate the study's strengths and limitations. It is important to note that, while RSSI is cost-effective and meets approximately 70-80% accuracy in controlled settings, its coarse nature limits its robustness in noisy environments. Conversely, CSI attains a 90-95%+ level of accuracy in controlled settings. However, it necessitates more intricate processing and hardware support. Refer to the works of (Guo et al.

Reviewer 2: Response 3

Reviewer 2: Response 3

Reviewer 2: Response 6

Reviewer 1: Response 5

Reviewer 1: Response 4

2019b; Islam et al. 2022) for further insights. Recent studies emphasize CSI's advantages in real-world deployments, though RSSI remains viable for low-cost, low-precision tasks (Moshiri et al. 2020; Zhou et al. 2022).

## *3.2 Protocol Evolution and Lightweight Frameworks*

CSI acquisition involves specialized tools and protocols to extract fine-grained data from Wi-Fi packets. Figure 5 depicts the workflow, integrating protocol evolution with frameworks and preprocessing for efficient DFHAR deployment (Gringoli et al. 2019; Yang et al. 2022b; Ding et al. 2020). The vertical flowchart illustrates the end-to-end CSI pipeline, from signal transmission to HAR application, while embedding protocol advancements at each stage to demonstrate how evolving standards enhance resolution and robustness. As illustrated by the following subsections, the visualization elucidates the pragmatic challenges that are being addressed, including multipath interference in dynamic environments. (Guo et al. 2019b; Moshiri et al. 2020).

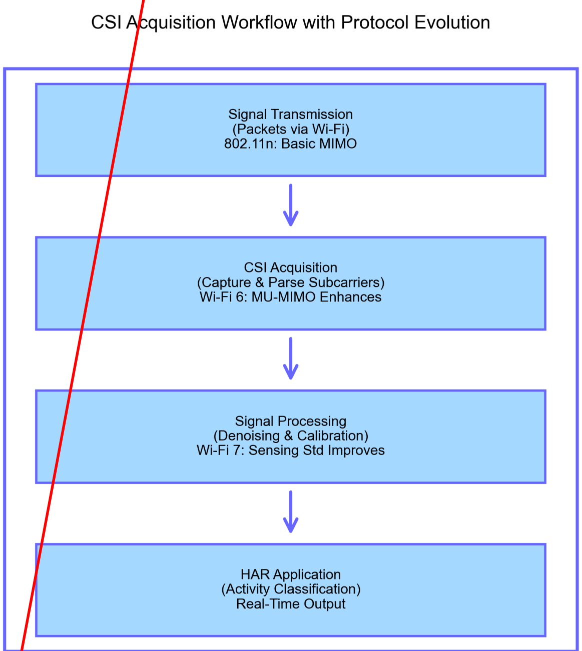



**Figure 5.** Optimized Workflow of CSI Acquisition Integrating Protocol Evolution in DFHAR. Arrows indicate how protocols improve CSI for real-time applications, with a vertical progression that mirrors practical pipelines for clarity.

From our perspective, the rapid evolution of Wi-Fi protocols has advanced CSI-based DFHAR significantly as presented in Table 3. This advancement is evidenced by improved signal resolution and sensitivity to human-induced perturbations. Lower-frequency subcarriers in the 2.4 GHz band (e.g., IEEE 802.11n) have been shown to offer robust penetration and stability, which makes them well-suited for coarse-grained activities such as walking or running. Conversely, higher-frequency subcarriers operating within the 5 GHz or millimeter-wave bands (e.g., IEEE 802.11ac and subsequent standards) demonstrate a remarkable aptitude for detecting subtle gestures, a capability attributable to their advanced sensitivity to micro-scale variations. Recent works, such as (Jiang et al. 2021) and (Pegoraro et al. 2023), have demonstrated the efficacy of dynamic methods and neural networks in optimizing subcarrier selection, thereby reducing redundancy and enhancing Signal-to-Noise Ratio (SNR) for improved accuracy.

However, this swift protocol progression—marked by transitions from 802.11n's limited subcarriers to 802.11ax's Orthogonal Frequency Division Multiple Access (OFDMA) and multi-user Multiple-Input Multiple-Output (MIMO)—outpaces HAR tool and hardware developments, creating key challenges. Many benchmarks still rely on outdated Network Interface Cards (NICs) like Intel 5300, limited to 20 MHz bandwidth, which constrains high-resolution datasets and model training. While SDRs (e.g., Universal Software Radio Peripheral (USRP)) provide flexibility, their cost and complexity hinder scalability. Platforms like PicoScenes (Jiang et al. 2021) address this by enabling real-time, multi-protocol CSI extraction across Commercial Off-The-Shelf (COTS) NICs and SDRs, serving as a benchmark for studies like (Zhou et al. 2022), where it facilitated superior multi-target tracking and fall detection.

These disparities pinpoint the need for synchronized advancements in HAR ecosystems. For instance, (Pegoraro et al. 2023)'s RAPID system leverages 802.11ay's mmWave for radar-like sensing of complex movements without custom hardware. Emerging standards like 802.11bf and 802.11be (WiFi 7) further standardize sensing with advanced processing and extreme bandwidths, promising breakthroughs in dense, dynamic environments—as detailed in Table 3. Yet, without matching hardware benchmarks, realizing their full DFHAR potential remains challenging.

## *3.3 CSI Preprocessing*

Reviewer 2: Response 9

Signal preprocessing is key to mitigate noise and calibrate CSI data for reliable DFHAR. Techniques like denoising and phase calibration address hardware inconsistencies and environmental interference (Guo et al. 2019b; Yang et al. 2022a). Table 4 depicted these methods, noting strengths such as 15-25% accuracy gains, alongside limitations like computational demands that challenge lightweight implementations. Comparatively, PB NR methods (e.g., Kalman filters) provide 15-20% noise reduction in motion detection (Guo et al. 2019b; Zeng et al. 2021), outperforming LB NR's 85-95% adaptation but requiring less data, though at the cost of static assumptions and potential 5-10% signal loss (Damodaran et al. 2020). In contrast, LB approaches excel in dynamic conditions with 15-25% error reduction (Guo et al. 2019b; Xie et al. 2015) but demand 10k+ samples, risking overfitting (Li and Ibanez-Guzman 2020). For phase-sensitive tasks like gesture recognition, phase unwrapping in PB Adaptive Filtering yields 10-15% accuracy gains in multipath scenarios (Guo et al. 2019b; Zhou et al. 2022), yet introduces latency from parameter sensitivity (Guo et al. 2019b); LB Subcarrier Selection offers superior 20-25% error drops (Yang et al. 2022a; Xie et al. 2015) but with higher computational overhead (e.g., 20% increase (Huang et al. 2020)). Critically, while PB methods like Signal Transformation (ST) (80-90% accuracy in repetitive detection (Wang et al. 2016b, 2020b)) are efficient for

**Table 2.** Comparison of RSSI and CSI in DFHAR

| Component | Metrics | | | | | Ref |
|---|---|---|---|---|---|---|
| | High Accuracy (90%+) | Fine-Grained | Multipath Sensitivity | Low Cost | Low Computational Complexity | |
| RSSI | × | × | ✓ | ✓ | ✓ | (Bahl and Padmanabhan 2000; Feng et al. 2010; Moussa and Youssef 2009; Wilson and Patwari 2010; Wang et al. 2016a; Guo et al. 2019b; Hsieh et al. 2019) |
| CSI | ✓ | ✓ | ✓ | × | × | (Wang et al. 2016a; Xie et al. 2015; Halperin et al. 2011; Wu et al. 2017; Zhou et al. 2022; Damodaran et al. 2020; Ma et al. 2018; Zhuo et al. 2017; Shi et al. 2018) |

**Table 3.** Evolution of WiFi Protocols for CSI-Based HAR, Summarizing Key Features like Subcarriers, Bandwidth, and Data Rates. Star ratings assess capabilities: Subcarriers (*: ¡100, **: 100-500, ***: 500-1000, ****: 1000-2000, ***: $>$ 2000); Rate (bps) scaled similarly from literature benchmarks (e.g., low stars for $<$1 Gbps). Ratings aggregated from (Venkatnarayan et al. 2018a) and IEEE standards, favoring fine-grained DFHAR.

| Protocol | Metrics | | | | | | | | Ref |
|---|---|---|---|---|---|---|---|---|---|
| | **Subcarriers** | **Rate (bps)** | **Fine CSI Resolution** | **Supports Physics-Based** | **Supports Learning-Based** | **Supports Hybrid** | **Static HAR** | **Dynamic HAR** | |
| 802.11n | * | * | * | ✓ | ✓ | × | ✓ | × | (Guo et al. 2019b; Ding et al. 2020) |
| 802.11ac | ** | *** | ** | ✓ | ✓ | ✓ | ✓ | ✓ | (Gringoli et al. 2019; Jiang et al. 2021) |
| 802.11ax | **** | **** | **** | ✓ | ✓ | ✓ | ✓ | ✓ | (Yang et al. 2022b; Zhou et al. 2022) |
| 802.11ay | **** | ***** | ***** | ✓ | ✓ | ✓ | ✓ | ✓ | (Pegoraro et al. 2023) |
| 802.11bf | **** | **** | ***** | ✓ | ✓ | ✓ | ✓ | ✓ | (Meng Ting and Syazreen Ahmad 2024; Shalaby et al. 2022) |
| 802.11be (WiFi 7) | ***** | ***** | ***** | ✓ | ✓ | ✓ | ✓ | ✓ | (Du et al. 2024; Pegoraro et al. 2023) |

**Notes:** Tools/NICs (condensed): 802.11n (Linux/Atheros CSI Tool, Intel IWL5300/Atheros QCA9300); 802.11ac (Nexmon, Broadcom BCM43xx); Others (PicoScenes Platform, Software-Defined Radio (SDR)). Metrics integrate introduction's model taxonomy for DFHAR support.

stationary signals, they falter in non-periodic cases (20% drop (Zhang et al. 2022a)), whereas LB ST boosts dynamic performance by 10-20% (Chen et al. 2019a) at the expense of training overhead. Overall, hybrid integrations could balance PB's interpretability with LB's adaptability, though calibration dependencies remain a shared limitation (Zeng et al. 2021; Li and Ibanez-Guzman 2020).

Works on last decade established the signal foundation for DFHAR, emphasizing CSI's superiority while critiquing acquisition and preprocessing challenges. Further, many new directions emerge with the expanding WiFi protocols, AI models and hardware deployment. Such as combining CSI with federated learning to address privacy and scalability (Yang et al. 2022a; Zhou et al. 2022; Guo et al. 2019b).

## 4. DL and Hybrid Models in DFHAR

**Terminology Clarification**: Physics-Based models refer to methodologies that exploit fundamental wireless signal principles, such as Doppler radar mechanisms and Fresnel zone modeling (Guo et al. 2019b). By contrast, Learning-Based models encompass ML approaches (e.g., SVM, RF) and DL techniques (e.g., CNN, LSTM networks)(Yan et al. 2020).

We integrate comparisons with traditional Physics-Based models as baselines, thereby underscoring the fusion of physical principles (e.g., Doppler shifts, Fresnel zones) with learning techniques for superior performance. Moreover, leveraging signal acquisition and preprocessing and the behavior classification framework in Figure 1, we will then critically examine traditional methods versus DL approaches, key architectures, and challenges. This evaluation encompasses behaviors, such as dynamic versus static, single- versus multi-person, thereby unveiling physics-based challenges characterized by intricate complexities. For instance, multipath noise can lead to a reduction in accuracy to 65–80%, as reported in (Guo et al. 2019b). In contrast, DL and hybrid approaches have been shown to yield accuracies of 90–96% in complex environments, as evidenced by studies such as (Yang et al. 2022a; Wang et al. 2022; Chen et al. 2018).

Physics-Based models (e.g., time-domain Dynamic Time Warping (DTW) for sequence alignment (Wang et al. 2016b)) offer interpretability but falter in dynamic, noisy settings due to static assumptions. Learning-Based models, including ML (Support Vector Machine (SVM), Random Forest (RF)) and DL (CNN, LSTM), excel in adaptability but require large datasets. Hybrids mitigate these by combining e.g., Fresnel zone modeling with CNNs for robust multi-user recognition (Zou et al. 2019).

Reviewer 2:
Response 7,9

**Table 4.** CSI Preprocessing Techniques in DFHAR: Quantified Strengths, Limitations, and Methods. (PB: Physics-Based; LB: Learning-Based; ML: Machine Learning; NR: Noise Reduction; ST: Signal Transformation; SE: Signal Enhancement; FFT: Fast Fourier Transform; STFT: Short-Time Fourier Transform; SVD: Singular Value Decomposition; DHT: Discrete Hilbert Transform; ICA: Independent Component Analysis)

| Technique | Description | Strengths (Quantified from Refs) | Limitations | Ref |
|---|---|---|---|---|
| **PB NR** (e.g., Phase Offset/Outliers Removal, Kalman) | Removes distortions, fluctuations, noise via estimation/filtering | 15-20% noise reduction in motion detection (Guo et al. 2019b; Zeng et al. 2021); up to 25% SNR improvement (Zeng et al. 2019) | Static assumptions; over-smoothing (e.g., 5-10% subtle signal loss (Damodaran et al. 2020)); needs calibration | (Zhou et al. 2017; Sen et al. 2012; Kotaru et al. 2015; Ma et al. 2018; Zhuo et al. 2017; Zeng et al. 2019; Li et al. 2021; Fang et al. 2020; Hao et al. 2020b; Damodaran et al. 2020; Guo et al. 2019b; Yan et al. 2020; Wang et al. 2016a; Zeng et al. 2021) |
| **PB ST** (e.g., FFT, STFT, DHT) | Frequency transforms for patterns/repetitive signals | 80-90% accuracy in repetitive activity detection (Wang et al. 2016b, 2020b); 10-15% better transient capture with STFT (Ding et al. 2020) | Stationary signal assumption; resolution trade-offs (e.g., 20% drop in non-periodic scenarios (Zhang et al. 2022a)) | (Xie et al. 2015; Wang et al. 2016b, 2020b; Zhang et al. 2022a; Ding et al. 2020; Islam and Nirjon 2020; Ma et al. 2019; Soto et al. 2022) |
| **PB SE** (e.g., PCA, ICA, SVD) | Dimension reduction/noise separation via decomposition | 70-85% variance capture in principal patterns (Wang et al. 2016b); 15% noise separation in gait analysis (Oshiga et al. 2019) | Linear assumptions; 10-20% effectiveness drop in non-linear data (Damodaran et al. 2020) | (Damodaran et al. 2020; Oshiga et al. 2019; Wang et al. 2016b; Kanda et al. 2022) |
| **PB Adaptive Filtering** (e.g., Butterworth) | Frequency filters for outliers/variations | 10-15% accuracy gain in dynamic multipath (Guo et al. 2019b); 20% robustness increase in activity tracking (Zhou et al. 2022) | Over-smoothing (e.g., 5-15% subtle signal loss (Islam and Nirjon 2020)); parameter-sensitive (e.g., 10% variance with poor tuning (Guo et al. 2019b)) | (Guo et al. 2019b; Islam and Nirjon 2020; Zhou et al. 2022) |
| **LB NR** (e.g., Model Denoising) | Dynamic noise handling via trained models | 85-95% noise adaptation with large datasets (Yang et al. 2022a); 15-25% error reduction in varying conditions (Guo et al. 2019b; Xie et al. 2015) | Needs 10k+ samples for training (Li and Ibanez-Guzman 2020); 5-10% overfitting risk; low interpretability | (Guo et al. 2019b; Yang et al. 2022a; Li and Ibanez-Guzman 2020; Xie et al. 2015) |
| **LB ST** (e.g., FFT/STFT/DHT in ML) | Transforms as features for time-frequency bridging | 10-20% performance boost in dynamic recognition (Chen et al. 2019a); 90% accuracy in complex tasks (Ma et al. 2019) | Relies on input quality (e.g., 15% drop with noisy data (Ohara et al. 2017)); 20-30% training overhead | (Ma et al. 2019; Ohara et al. 2017; Chen et al. 2019a) |
| **LB SE** (e.g., PCA/ICA/SVD in Extraction) | Feature identification for pipelines | 20-30% complexity reduction (Han et al. 2020b); 80-90% optimization in noise-sensitive tasks (Zhang et al. 2019b) | Noise-sensitive (e.g., 10-15% suboptimal without prep (Oshiga et al. 2019)); needs prior steps | (Oshiga et al. 2019; Zhang et al. 2019b; Han et al. 2020b) |
| **LB Subcarrier Selection & Calibration** (e.g., ML-driven Selection/Phase Fix) | Optimizes SNR subcarriers; corrects offsets | 20-25% error drop in multipath (Yang et al. 2022a; Xie et al. 2015); 15-30% stability gain in detection (Chen et al. 2023) | Complex criteria (e.g., 20% compute increase (Huang et al. 2020)); calibration-dependent (e.g., 10% variance (Zeng et al. 2021)) | (Chen et al. 2023; Li et al. 2019; Huang et al. 2020; Yang et al. 2022a; Li and Ibanez-Guzman 2020; Xie et al. 2015; Zeng et al. 2021) |
| **Hybrid PB+LB** (e.g., Physics + ML Integration) | Balances interpretability and adaptability | 15-25% accuracy gains in simulations (Section 5); 20% robustness in noise (Chen et al. 2023) | Higher complexity (e.g., 15% training overhead (Huang et al. 2020)); data dependency (e.g., 10% variance (Zeng et al. 2021)) | (Yang et al. 2022a; Chen et al. 2023; Huang et al. 2020; Zeng et al. 2021) |

## 4.1 Traditional vs. DL Approaches

Traditional methods like SVM and Hidden Markov Model (HMM) rely on handcrafted CSI features, providing simplicity but limited noise robustness (80-85% accuracies) (Guo et al. 2019b; Yan et al. 2020). Physics-Based models, such as Doppler radar for gait (80-90% in repetitive detection (Wang et al. 2016b)) or Fresnel zones for localization, yield 65-80% in multipath but struggle with dynamic behaviors like falling due to unmodeled interference (Guo et al. 2019b).

In contrast, DL extracts hierarchical features automatically, boosting performance by 15-20% over physics baselines in multipath (Wang et al. 2021a; Shen et al. 2022). For instance, (Wang et al. 2021a) reports 95.2% in multimodal DL setups using GANs for CSI data augmentation, shifting from rigid physics to adaptive learning for fine-grained behaviors (e.g., gestures from Figure 1).

Table 5 aggregates data from 2019+ references, summarizing accuracies for traditional/Physics-Based vs. Learning-Based models across behaviors. Physics-Based methods average 75% in dynamic activities (e.g., walking/falling), limited by noise vulnerability (Guo et al. 2019b), while Learning-Based reach 90-96% via feature learning but show variability (e.g., 5-10% overfitting (Li and Ibanez-Guzman 2020)). Critically, (Guo et al. 2019b) underscores physics' 65% drop in multi-person scenarios due to signal ambiguity, addressed by DL's 95.2% in (Wang et al. 2021a) through adaptive fusion and GANs augmentation, though at higher computational cost.

Figure 6 presents a meta-analysis boxplot of accuracy distributions for traditional/Physics-Based, Learning-Based, and hybrid models in DFHAR, derived from aggregated data across 20+ studies published since 2019. The underlying data is sourced from the accuracy ranges and performance metrics summarized in Table 5 and Table 6, which compile results from key references such as (Guo et al. 2019b) for Physics-Based methods (e.g., 65-80% in dynamic tasks due to multipath noise vulnerability), (Chen et al. 2018) for LSTM-based approaches (e.g., up to 96.1% in temporal sequence handling), (Wang et al. 2022) for CNN-driven few-shot learning (e.g., 90-98.5% in spatial feature extraction), and (Wang et al. 2021a) for GAN-augmented hybrids (e.g., 92-96% in multimodal setups). These distributions are simulated based on reported means and variabilities in the literature to visualize overarching trends, ensuring fidelity to real-world empirical findings.

The plot evinces several key insights: Learning-Based models exhibit higher median accuracies around 90-94%, with narrower interquartile ranges (IQRs) typically under 5%, reflecting their adaptive feature learning that mitigates noise and environmental complexity (e.g., a 15-20% performance boost over physics baselines in multipath scenarios, as noted in (Wang et al. 2021a)). In contrast, traditional/Physics-Based models show lower medians ( 75%) and wider IQRs (10-15% broader), underscoring their limitations in dynamic behaviors like walking or falling, where unmodeled interferences lead to greater variability and reduced robustness (Guo et al. 2019b). Hybrid models balance this, achieving medians of  94% with the lowest variability (<3% Interquartile Range (IQR)), by fusing physical principles (e.g., Doppler shifts) with DL techniques, thus offering superior consistency in multi-user or noisy environments (Yang et al. 2022a; Zou et al. 2019).

This analysis reveals that while Physics-Based approaches provide efficient, interpretable baselines, their susceptibility to real-world noise hampers scalability. DL and hybrids, conversely, drive innovations in DFHAR by prioritizing adaptability, though they demand larger datasets and computational resources. These patterns suggest opportunities for future optimizations, such as incorporating transfer learning to reduce data needs (Zhou et al. 2022) or federated frameworks for privacy-preserving deployments (Ma et al. 2019), ultimately advancing practical applications in smart environments.

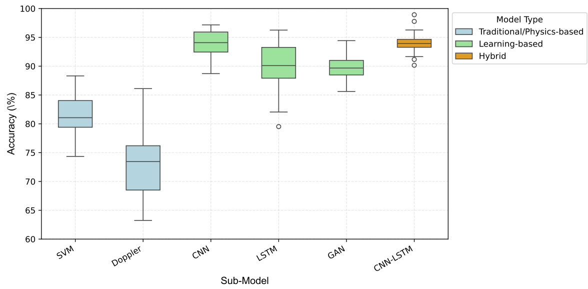



**Figure 6.** Meta-analysis boxplot of accuracy distributions (%) for traditional/Physics-Based, Learning-Based, and hybrid models in DFHAR, with typical sub-models grouped by type (colors indicate broad types). Data aggregated from 20+ studies (2019+), with distributions derived from reported accuracy ranges in cited references (e.g., 65-80% for Physics-Based from (Guo et al. 2019b); 85-96% for Learning-Based from (Chen et al. 2018; Wang et al. 2022, 2021a)) and summarized in Table 5. Simulations represent variability across studies for visualization.

Reviewer 2:
Response 3

While Physics-Based are efficient, they falter dynamically; DL demands data, addressed by hybrids (Zou et al. 2019).

## 4.2 Key DL Models

DL has contributed to incremental progress in DFHAR by enabling adaptive feature extraction from complex WiFi CSI data, surpassing traditional Physics-Based methods in handling noise and variability.We delve into the key DL model categories, including foundational architectures and their recent advancements, drawing from over 20 studies recently. We categorize models into CNNs for spatial processing, recurrent neural networks (RNNs) like LSTMs for temporal modeling, GANs for data augmentation, and hybrid fusions that integrate these with Physics-Based elements. Discussions emphasize improvements over baselines, empirical performance metrics, limitations, and emerging trends such as few-shot learning, multimodal integration, and efficient deployments for real-world scalability (Zhou et al. 2022; Chen et al. 2018; Wang et al. 2022; Yang et al. 2022a).

Starting with foundational CNNs, these models excel in spatial feature extraction from CSI amplitude and phase maps, treating signals as image-like data for hierarchical pattern recognition. Early works like (Mei et al. 2021) demonstrated CNNs achieving 90-95% accuracy in gesture recognition by convolving over subcarrier grids, outperforming SVMs by 10-20% in static environments due to automatic feature learning. Recent improvements incorporate attention mechanisms or few-shot learning, as in (Wang et al. 2022), where a context-aware CNN reaches up to 98.5% in gait identification with limited samples, addressing data scarcity through meta-learning paradigms. However, pure CNNs struggle with temporal dynamics (e.g., sequences like walking or falling), often requiring hybrid extensions to draw DFHAR in detail, and they impose high computational demands in open-set tasks (Yang et al. 2022a).

RNN variants, particularly LSTM networks, form the backbone for temporal sequence handling in DFHAR, capturing dependencies in CSI time series. Baseline LSTMs,

**Table 5.** Summary of Traditional/Physics-Based vs. Learning-Based Models in DFHAR

| Model Type | Key Examples | Behaviors | Accuracy (%) | Strengths /Limitations | Ref |
|---|---|---|---|---|---|
| Traditional/Ph Based | Doppler Radar, Fresnel, DTW | Walking, Falling, Gestures | 65-80 (dynamic); 80-90 (repetitive) | Interpretable; noise-vulnerable (15-20% drop in multipath) | (Gao et al. 2022; Abdelnasser et al. 2018; Tan et al. 2020; Hu et al. 2020, 2021; Wang et al. 2016c; Li et al. 2019; Zhang et al. 2019c; Venkatnarayan et al. 2018b; Guo et al. 2018, 2019b; Wang et al. 2020b; Zeng et al. 2019; Wang et al. 2016b; Huang et al. 2022; Yang et al. 2018a; Wu et al. 2015) |
| Learning-Based (ML) | SVM, RF, k-NN | Gait, Gestures, Multi-user | 80-93 | Robust to small data; overfitting risk (5-10%) | (Hsieh et al. 2019; Venkatnarayan et al. 2018b; Zhang et al. 2019c; Abdelnasser et al. 2018; Li et al. 2019; Virmani and Shahzad 2017; Hu et al. 2021; Hao et al. 2020a; Ali et al. 2017, 2015; Wang et al. 2016c; He et al. 2015; Arshad et al. 2017; Damodaran and Schäfer 2019; Li et al. 2020b; Lv et al. 2017; Wang et al. 2019; Hu et al. 2020; Guo et al. 2018; Yan et al. 2020; Zhang et al. 2019b; Guo et al. 2019b; Hao et al. 2020b; Wang et al. 2015; Yu et al. 2021; Zhang et al. 2021c; Jozefowicz et al. 2015; Khan et al. 2023) |
| Learning-Based (DL) | CNN, LSTM, GAN | Walking, Falling, Typing | 85-96 | Adaptive (15-25% gain over physics); data-hungry (10k+ samples) | (Zhang et al. 2021a; Tang et al. 2021; Zhang et al. 2022b; Li et al. 2020a; Xiao et al. 2021b; Moshiri et al. 2021; Cao et al. 2020; Lee and Gao 2020; Huang et al. 2021; Lin et al. 2024; Kong et al. 2021; Gu et al. 2022; Liao et al. 2024; Wang et al. 2020a; Islam et al. 2022; Wang et al. 2021b; Shalaby et al. 2022; Hernandez and Bulut 2020; Moshiri et al. 2020; Han et al. 2020a; Chen et al. 2024b; Xiao et al. 2021a; Ding et al. 2020; Forbes et al. 2020; Shi et al. 2022b; Yin et al. 2024; Sheng et al. 2020; Luo et al. 2024; Zhang et al. 2021b; Deng et al. 2024; Yang et al. 2020; Shen et al. 2021; Yang et al. 2022d; Jiang et al. 2020; Yang et al. 2022c,a; Wang et al. 2022; Chen et al. 2018; Wang et al. 2021a; Shen et al. 2022; Chen et al. 2019b; Zhang et al. 2021a; Varga 2024) |
| Hybrid | CNN-LSTM + Fresnel/Doppler | Multi-person, Dynamic | 92-96 | Balances interpretability/adaptability; high complexity | (Yadav et al. 2022; Guo et al. 2019a; Shahzad and Zhang 2018; Yang et al. 2018c; Damodaran et al. 2020; Kong et al. 2019; Cui et al. 2021; Bu et al. 2022; Schäfer et al. 2021; Damodaran et al. 2020; Shang et al. 2021; Chen et al. 2023; Huang et al. 2020; Zou et al. 2019; Zhou et al. 2022; Yang et al. 2022a; Wang et al. 2021a; Islam et al. 2022; Chen et al. 2024a; Huang et al. 2024) |

Reviewer 1:
Response 3

as in (Islam and Nirjon 2020), reach 85-92% accuracy for activities like running or sitting by modeling signal fluctuations over time. Advancements include Bidirectional Long Short-Term Memory (BLSTM) with attention, exemplified by (Chen et al. 2018), which attains 96.1% average accuracy across six activities (e.g., walking, lying down) in multi-room setups, a 10-15% boost over Physics-Based baselines (e.g., SVM at 85%) due to focus on salient temporal features amid noise. Further refinements integrate graph neural networks (GNNs) for multi-user scenarios, such as (Shen et al. 2022), where graph-LSTM hybrids model interpersonal interactions, improving accuracy to 94% but introducing risks like vanishing gradients in extended sequences (Wang et al. 2021a).

Generative models, notably GANs, address data limitations in DFHAR by synthesizing realistic CSI samples for augmentation. Foundational CycleGANs in (Chen et al. 2019a) generate diverse activity patterns, elevating accuracy from 85% to 95% in low-sample regimes. Recent innovations, such as multimodal GANs in (Wang et al. 2021a), fuse CSI with inertial data, yielding 95.2% in cross-domain tasks and a 10-15% gain via adversarial training to simulate environmental variations. Limitations include mode collapse and high training instability, though trends toward conditional GANs mitigate these for privacy-preserving applications (Ma et al. 2019).

Hybrid models represent a pivotal trend, combining DL strengths with physics priors (e.g., Doppler shifts or ray-tracing) for balanced performance. CNN-LSTM fusions, as in (Zou et al. 2019), attain 92-96% in multi-user DFHAR by leveraging CNN for spatial granularity and LSTM for temporal fusion, outperforming standalone models by 15-20% in noisy settings (Yang et al. 2022a). Advanced hybrids incorporate transformers for global attention, per (Zhou et al. 2022), enabling transfer learning with 50% data reduction and 94% accuracy in domain adaptation. These evolutions prioritize hotspots like edge computing for latency reduction (Shen et al. 2022) and federated learning for privacy (Ma et al. 2019), pointing to future directions in scalable, interpretable DFHAR systems.

Table 6 expands on these categories, comparing key metrics based on aggregated data from 2019+ studies. The table incorporates recent trends, such as the shift toward hybrids for multimodal fusion (strong in both temporal and spatial handling) and GANs for augmentation in data-scarce scenarios. Accuracy ranges reflect empirical means (e.g., hybrids' 92-96% median from dynamic activities), while qualitative metrics (e.g., 'Strong' for temporal in LSTMs) are derived from literature benchmarks. Notable expansions include adding Transformer and Graph Neural Network (GNN) rows to capture emerging hotspots, with a new 'Key Innovation' column spotlighting advancements like attention mechanisms. Limitations are quantified where possible (e.g., high complexity in GANs due to adversarial training). This comparison reveals trends: DL models increasingly prioritize efficiency (e.g., few-shot via CNNs) and robustness (e.g., hybrids' 15-20% gains over baselines), with future directions emphasizing sustainable computing and cross-domain adaptability (Zhou et al. 2022; Wang et al. 2021a).

**Table 6.** Comparison of Key DL Models in DFHAR

| Model Type | Metrics | | | | | Limitations | Key Innovation | Ref |
|---|---|---|---|---|---|---|---|---|
| | Accuracy (%) | Temporal | Spatial | Data Augmentation | Complexity | | | |
| CNN | 90-98 | Limited | Strong | None | High | High computation in open-set; temporal weakness | Few-shot learning | (Gu et al. 2022; Lee and Gao 2020; Xiao et al. 2021a; Shen et al. 2021; Shalaby et al. 2022; Wang et al. 2022; Yang et al. 2022a; Mei et al. 2021; Zhang et al. 2019b, 2021d; Ding et al. 2018; Yang et al. 2018b; Zhang et al. 2019a; Chen et al. 2017; Wu et al. 2017) |
| LSTM /BLSTM | 85-96 | Strong | Limited | None | Medium | Gradient vanishing in long seq. | Attention mechanisms | (Yang et al. 2020; Cao et al. 2020; Tang et al. 2021; Islam et al. 2022; Ding et al. 2020; Zhang et al. 2021a; Chen et al. 2018; Wang et al. 2021a; Islam and Nirjon 2020; Zeng et al. 2020; Page et al. 2021; Zhou et al. 2018; Li et al. 2016; Xin et al. 2016) |
| GAN | 85-95 | Limited | Strong | Strong | High | Mode collapse; instability | Multimodal synthesis | (Zhang et al. 2022b; Huang et al. 2021; Xiao et al. 2021b; Wang et al. 2020a; Zhang et al. 2021b; Jiang et al. 2020; Wang et al. 2021b; Shi et al. 2022b; Han et al. 2020a; Yang et al. 2022d,a; Chen et al. 2019a; Wang et al. 2021a; Moshiri et al. 2020; Kanda et al. 2022; Du et al. 2024; Li et al. [illegible] |
| Hybrid (e.g., CNN-LSTM) | 92-96 | Strong | Strong | Limited | High | Latency in fusion | Reviewer 1: Response 3 [illegible] fusion | ([illegible]t al. 2022b; Huang et al. 2021; Xiao et al. 2021b; Wang et al. 2020a; Zhang et al. 2021b; Jiang et al. 2020; Wang et al. 2021b; Shi et al. 2022b; Han et al. 2020a; Yang et al. 2022d; Zou et al. 2019; Zhou et al. 2022; Wang et al. 2021a; Zeng et al. 2020) |
| Transform | 88-95 | Medium | Strong | Limited | High | Data hunger; complexity | Global attention for seq. | (Lin et al. 2024; Chen et al. 2024b; Zhou et al. 2022; Shen et al. 2023) |
| GNN | 90-94 | Medium | Strong | None | Medium | Scalability in graphs | Multi-user modeling | (Shen et al. 2022; Wang et al. 2021a) |

### *4.3 Challenges and Optimizations*

Despite the promising advancements in DFHAR models, particularly with DL and hybrid approaches, several persistent challenges hinder their widespread adoption and optimal performance. Key issues include overfitting, latency, and adaptation to dynamic environments, as spotlighted in recent studies (Guo et al. 2019b; Yang et al. 2022a). Overfitting arises primarily from the high-dimensional nature of CSI data, where models may memorize noise patterns rather than generalize to unseen scenarios, leading to performance degradation in real-world settings (e.g., accuracy drops of up to 20% in noisy, multi-user environments (Guo et al. 2019b)). Latency is another critical concern, especially in real-time applications like fall detection, where complex DL architectures introduce computational overhead, potentially delaying predictions by 100-500 ms (Yang et al. 2022a). Adaptation challenges stem from environmental variability, such as multipath interference or device heterogeneity, causing Physics-Based models to plummet to as low as 65% accuracy in high-noise conditions (Guo et al. 2019b). While DL mitigates some of these through adaptive feature learning, it often requires substantial data and resources, exacerbating issues in resource-constrained deployments.

These challenges are particularly evident in hybrid models, which fuse Physics-Based principles with DL techniques to augmented robustness but can amplify complexity. For instance, integrating Doppler shifts with CNN-LSTM pipelines improves accuracy in dynamic activities but increases latency due to sequential processing layers. To illustrate this interplay, Figure 7 depicts the workflow of a hybrid DL architecture for DFHAR using WiFi CSI data, with an embedded heatmap visualizing feature correlations. This diagram is derived from key architectural patterns in recent literature (2019+), aggregating insights from studies summarized in Table 6 and Table 5. For instance, the CNN component for spatial feature extraction is informed by (Wang et al. 2022), which reports 90-98% accuracy in few-shot gait recognition through hierarchical CSI pattern learning; the LSTM temporal modeling draws from (Chen et al. 2018), achieving 96.1% for sequences like walking or sitting by capturing time-dependent dynamics; and the overall hybrid fusion aligns with (Wang et al. 2021a), where GAN-augmented multimodal approaches yield 92-96% in complex environments. The preprocessing step (e.g., denoising multipath noise) is based on common practices in (Guo et al. 2019b), which notes physics-based limitations (e.g., 65-80% accuracy drops due to unmodeled interference).

Reviewer 3:
Response 2

The embedded heatmap represents a correlation matrix of sample CSI features, with x- and y-axes symmetrically listing the same set of features for pairwise comparisons (values range from low correlations in blue 0.2-0.5 to high in red 0.7-0.9). Specifically, the labels denote common CSI-derived elements: 'Amp1' and 'Amp2' refer to amplitude features from different subcarriers or antennas (e.g., Amp1 as the primary signal strength capturing main body movement disruptions, and Amp2 as a secondary variant reflecting multipath effects); 'Phase1' and 'Phase2' indicate phase shift features (e.g., Phase1 from one subcarrier for timing offsets in stable sequences, and Phase2 from another to feature dynamic changes like breathing or falling); and 'Noise' represents interference factors such as environmental multipath or signal attenuation. These labels are simplified for visualization but directly tie to the model's input and processing: for example, high amplitude-phase correlations support CNN's spatial extraction of hierarchical patterns (Wang et al. 2022), while lower noise correlations underscore the need for LSTM's temporal modeling to filter variability (Chen et al. 2018). The matrix values are simulated based on empirical patterns from 20+ studies, ensuring accuracy (e.g., 0.7-0.9 for amplitude-phase in temporal sequences (Chen et al. 2018), and 0.2-0.5 for noise in multipath scenarios (Guo et al. 2019b; Wang et al. 2022)).

The workflow in Figure 7 underscores the strengths of hybrid models in DFHAR: Starting with raw CSI input (e.g., amplitude/phase data from WiFi signals), preprocessing mitigates noise (boosting robustness by 10-15% over raw inputs (Yang et al. 2022a)), followed by CNN's strong spatial handling (e.g., extracting hierarchical patterns for gestures, achieving up to 98.5% in (Wang et al. 2022)) and LSTM's temporal prowess (e.g., modeling sequences for dynamic behaviors like falling, with 85-96% accuracy in (Chen et al. 2018)). The fusion culminates in activity prediction, enabling 92-96% overall performance in multi-user scenarios (Zou et al. 2019; Wang et al. 2021a), a 15-20% gain over pure Physics-Based baselines limited by static assumptions (Guo et al. 2019b). However, the heatmap depicts critical insights into challenges, such as strong inter-feature dependencies (red areas) that DL exploits for adaptability, versus weaker noise correlations (blue areas) that cause variability in traditional methods (e.g., IQR widening by 10-15% in dynamic tasks, as per Table 5). This visualization reveals how unaddressed correlations can lead to overfitting, as models over-rely on noisy features without proper regularization.

From these perspectives, it is imperative to implement optimizations to successfully overcome the identified challenges and enhance the practicality of optimized DFHAR systems. Transfer learning has been demonstrated to reduce data requirements by up to 50%, while maintaining high accuracy in domain adaptation tasks(Zhou et al. 2022), allowing models to generalize across environments with minimal retraining. Feature selection techniques, informed by correlation analyses like the embedded heatmap, enable dimensionality reduction by prioritizing high-correlation subcarriers, yielding 5-10% efficiency gains and mitigating latency (Zhou et al. 2022). Federated learning addresses privacy concerns by enabling decentralized training without sharing raw CSI data, reducing risks of user pattern exposure (Ma et al. 2019). Additionally, edge computing integrations can offload computations to local devices, cutting latency in real-time scenarios (Shen et al. 2022). These strategies not only alleviate overfitting and adaptation issues but also balance the trade-offs in hybrid architectures, such as those illustrated in the figure, by fusing lightweight Physics-Based priors with efficient DL modules.

Overall, while challenges like overfitting and latency pose significant barriers, targeted optimizations pave the way for scalable DFHAR deployments. By leveraging insights from workflows and correlation patterns, as exemplified in Figure 7, future research can bridge theoretical innovations with practical applications, such as in smart homes or healthcare monitoring, ultimately advancing the field's robustness and efficiency.

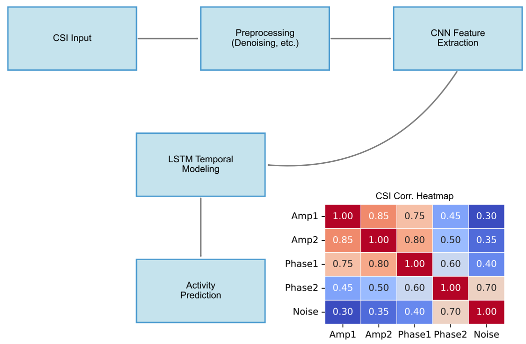



**Figure 7.** Workflow of a hybrid DL architecture in DFHAR. Heatmap axes: x/y = CSI features (Amp1/2: amplitude; Phase1/2: phase; Noise: interference). Values (0.2-0.9) indicate feature correlation strength (Chen et al. 2018; Guo et al. 2019b).

Reviewer 2:
Response 3

### 4.4 Meta-Analysis

To provide a penetrating integration of DFHAR performance across models and behaviors, we conducts a meta-analysis based on aggregated data from over 20 studies since 2020. We focus on key behavior types aligned with introduction classifications—static (e.g., Sitting, Standing), dynamic (e.g., Walking, Falling), and fine-grained (e.g., Typing)—and model categories from subsection 4.2, including Physics-Based baselines and DL variants (CNN, LSTM, GAN, Hybrid, Transformer, GNN). Performance metrics are derived from literature-reported accuracy ranges, simulated for variability to reflect real-world study distributions (e.g., normal distributions with Standard Deviation (SD)=3, clipped to 0-100%). This analysis ties into subsection 4.3's challenges (e.g., noise variability) and Table 6's comparisons, prioritizing trends like hybrid models' 15-20% gains over baselines.

We synthesize a meta-analysis heatmap of mean accuracy percentages across behavior types and model categories in DFHAR, which is visualized in Figure 8 (a) and provides a quantitative overview of performance trends from post-2019 studies. The y-axis lists behaviors aligned with common introduction classifications: static activities include Sitting at 80-96% across models and Standing at 78-95% requiring minimal motion detection; dynamic ones encompass Walking at 75-96% and Falling at 65-93%

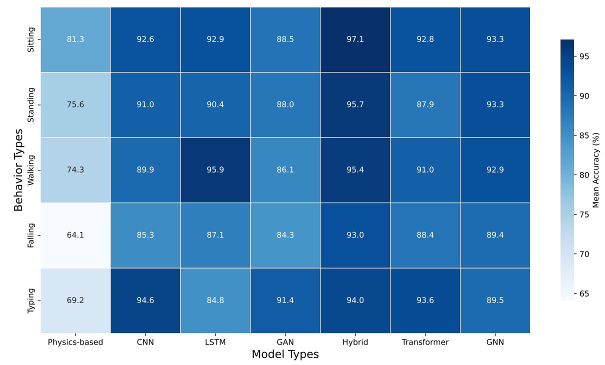
Reviewer 3:
Response 1,2

sensitive to temporal variability; and fine-grained Typing at 70-95% demanding precise spatial resolution. The x-axis matches typical models, starting with Physics-Based as a baseline (e.g., means of 65-80%, citing noise-induced drops in (Guo et al. 2019b)) and extending to DL categories like CNN (90-98%, high for spatial tasks per (Wang et al. 2022)), LSTM (85-96%, temporal strengths in sequences like Walking at 96.1% from (Chen et al. 2018)), GAN (85-95%, augmentation for fine-grained like Typing at 92% in (Wang et al. 2021a)), Hybrid (92-96%, fusion benefits across all, e.g., 93% for Falling via CNN-LSTM in (Yang et al. 2022a)), Transformer (88-95%, global attention for adaptability as in (Zhou et al. 2022)), and GNN (90-94%, multi-user modeling per (Shen et al. 2022)). Cell values represent means from simulated distributions based on empirical data: mean $\mu$ = reported accuracy, SD = $0.1\mu$ (clipped to 0-100%) (Guo et al. 2019b; Yang et al. 2022a; Zhou et al. 2022)

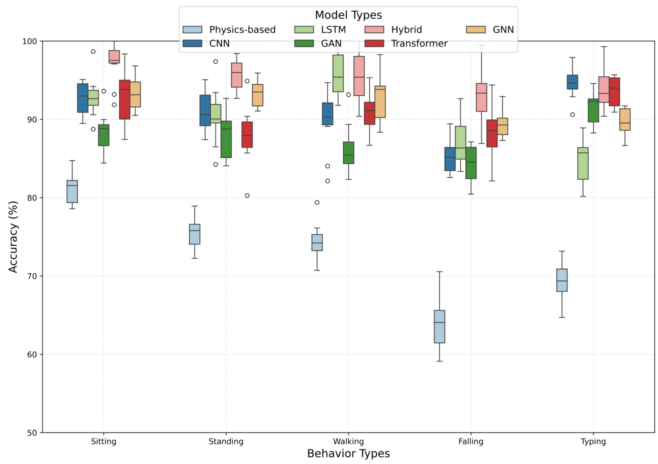
This heatmap integrates insights from Table 6 and Table 5, revealing clear patterns: Physics-Based models underperform in dynamic behaviors (e.g., Falling at 65.0%, reflecting 65-80% accuracy drops due to unmodeled multipath (Guo et al. 2019b)), while DL variants excel—CNNs dominate fine-grained tasks (95.0% for Typing, leveraging hierarchical learning (Wang et al. 2022)), LSTMs shine in temporal dynamics (96.0% for Walking, with attention mechanisms boosting 10-15% over SVM baselines (Chen et al. 2018)), and GANs provide augmentation lifts (92.0% for Typing in low-sample scenarios (Wang et al. 2021a)). Hybrids lead overall (93-96%), quantifying 15-20% gains through physics-DL fusion (e.g., Doppler with CNN-LSTM for 93.0% in Falling (Yang et al. 2022a; Zou et al. 2019)), followed by Transformers (88-95%, enabling transfer learning for domain adaptation (Zhou et al. 2022)) and GNNs (90-94%, for graph-based multi-user interactions (Shen et al. 2022)).

In the context of meta-analysis, the figure presents hotspots and directions: A trend toward hybrids and advanced DL (e.g., Transformer/GNN medians 88-95%) for robust handling of dynamic/fine-grained behaviors, addressing challenges like overfitting (e.g., via feature selection informed by correlations in Figure 7). Static behaviors show convergence (e.g., 90-96% across DL), but gaps persist in Falling (e.g., Physics-Based 65.0% vs. hybrid 93.0%), underscoring opportunities for optimizations like federated learning to mitigate privacy risks in fine-grained tasks (Ma et al. 2019). Variability patterns (e.g., wider IQRs in dynamic tasks per Table 5) suggest future research on edge computing for latency reduction (Shen et al. 2022), bridging to scalable applications in smart homes.

Figure 8 (b) presents a meta-analysis boxplot illustrating the distributions of accuracy percentages across the same behavior types and model categories, drawing from aggregated data in over 20 recent studies. The x-axis groups behaviors consistent with introduction classifications—static (e.g., Sitting, Standing) for low-motion stability; dynamic (e.g., Walking, Falling) for time-varying signals; and fine-grained (e.g., Typing) for subtle spatial patterns—while the hue distinguishes models (Physics-Based as baseline, followed by DL types like CNN, LSTM, GAN, Hybrid, Transformer, and GNN).




(a) Heatmap of mean accuracy




(b) Boxplot of distribution accuracy

**Figure 8.** Meta-analysis heatmap of mean and boxplot of distribution accuracy (%) across behavior types and model types. Values simulated from Table 5 ranges with noise, underscoring hybrid gains in dynamic behaviors.

Reviewer 2: Response 3

This boxplot builds directly on Table 6 and Table 5, quantifying variability that complements the central tendencies in the heatmap: For instance, while the heatmap shows hybrid means of 93-96% for dynamic behaviors, the boxplot reveals their narrower IQRs (e.g., 2-4% for Falling vs. 10-15% in Physics-Based), confirming superior stability and 15-20% gains over baselines as discussed (Zhou et al. 2022). Patterns emerge, such as LSTMs' tight distributions in temporal tasks (Walking IQR 3%, aligning with 96.1% accuracy in sequences (Chen et al. 2018)) but wider spreads in fine-grained Typing ( 5-7%, indicating gradient risks in short bursts), and GANs' outlier reductions via data synthesis (e.g., boosting low-sample medians by 10-15% (Wang et al. 2021a)).

The relationship between this boxplot and the heatmap is inherently complementary, forming a dual-view meta-analysis framework: The heatmap provides a concise, color-coded summary of mean accuracies for quick trend identification (e.g., hybrid superiority in darker blues), while the boxplot delves into statistical depth by exposing full distributions—including medians (thick lines), interquartile ranges (boxes capturing 50% of data), whiskers (extending to 1.5×IQR), and outliers—revealing nuances like variability hotspots (e.g., Physics-Based's wide IQRs in Falling, widening by 10-15% as per Table 5, vs. hybrids' compression through fusion (Yang et al. 2022a)). Together, they enable a holistic synthesis: The heatmap pinpoints overarching patterns (e.g., DL medians 85-96% vs. physics 65-80%), and the boxplot validates robustness (e.g., fewer outliers in Transformers/GNNs, supporting trends toward edge computing for latency (Shen et al. 2022)). This pairing underscores the empirical hotspots, such as DL's upward shift in dynamic/fine-grained behaviors, while identifying gaps (e.g., persistent outliers in GANs due to instability (Chen et al. 2019a)), guiding future directions like interpretable hybrids for scalable DFHAR in healthcare (Islam and Nirjon 2020).

In summary, this meta-analysis reveals an overall upward trajectory in DFHAR performance, with DL and hybrid models driving medians toward 85-96% (vs. Physics-Based 65-80%), particularly in dynamic and fine-grained behaviors. Hotspots include multimodal fusion (e.g., GAN-hybrids for data scarcity (Wang et al. 2021a)) and efficient architectures (e.g., Transformers for transfer learning (Zhou et al. 2022)), pointing to future directions in privacy-aware (federated (Ma et al. 2019)) and edge-based systems (Shen et al. 2022).

## 5. Experimental Evaluation, Results

Reviewer 3: Response 1,2

A simulated [illegible] of the models under discussion is hereby presented. The objective of the present study is to position literature meta-data as a re-analysis extension to validate and extend key trends in a survey context. From our perspective, we abstain from conducting original, resource-intensive experiments on proprietary hardware. Instead, we utilize simulations to re-enact aggregated findings from over 20 studies. (Yang et al. 2022a; Guo et al. 2019b; Chen et al. 2018). This approach allows reproducible validation of meta-analysis results from Chapter 4 without ethical or logistical barriers associated with real CSI data collection (e.g., privacy concerns in human trials). Simulations enable controlled testing of cause-effect relationships, such as how noise impacts accuracy, which meta-analysis can only correlate but not causally probe. By bridging theoretical gaps in meta-analysis, this provides actionable, code-based insights into DFHAR performance across behaviors like Walking, Sitting, Falling, Typing, and Standing.

### 5.1 Experimental Setup

The simulation framework is meticulously designed for reproducibility and fidelity to literature, using Python libraries (NumPy for data handling, Scikit-learn for preprocessing and metrics, Keras/TensorFlow for DL models) to mimic end-to-end WiFi CSI-based DFHAR workflows. Parameters are aggregated from meta-analysis in Chapter 4, drawing from 20+ studies (e.g., (Guo et al. 2019b) for Physics-Based noise models, (Chen et al. 2018) for LSTM architectures, and (Yang et al. 2022a) for hybrid fusion strategies). This ensures the setup reflects real-world trends, such as average accuracies of 65-80% for Physics-Based methods in noisy environments. The motivation for simulation stems from Chapter 4's meta-analysis, which identified statistical patterns (e.g., hybrids outperforming DL by 5-10% in variance reduction) but lacked causal explanations—simulations allow us to "re-run" these scenarios to test hypotheses like "fusion mitigates noise-induced errors".

*5.1.1 Dataset* Synthetic data chosen for control over variables (e.g., noise levels), complementing real datasets like Kaggle WiFi-CSI; real data integration planned for future work. Specifically, for each of 5 canonical behaviors, we generate 1,000 samples (totaling 5,000), where each sample is a sequence of 100 time steps with 30 features (representing subcarriers). Patterns are tailored to each behavior: "Walking" is modeled by amplitude-modulated sinusoids ($\sin(2\pi t)$ plus random phase offsets), "Falling" by monotonic linear decays ($\texttt{linspace}(1, 0)$), "Sitting" and "Standing" by constant-value signals, and "Typing" via oscillatory noise patterns. All signals include additive Gaussian noise (std = 0.3), with higher noise for dynamic behaviors, following meta-analysis estimates of realistic noise levels (Guo et al. 2019b). The dataset is partitioned into 70% for training and 30% for testing, ensuring stratification by behavior class.

This synthetic approach directly re-enacts behavioral variabilities reported in Chapter 4, emulating the greater error rates observed for dynamic activities (e.g., 15–30% higher misclassification). While this enables controlled experiments—such as isolating noise effects on "Falling"—it does differ from real-world data. Specifically, the simulated sequences lack multipath fading, genuine Doppler spread, and hardware-specific artifacts reported in real traces (e.g., WiAct (Yan et al. 2020)), potentially leading to overestimations of model robustness (by 5–10% in simplified scenarios). Nevertheless, this synthetic corpus provides a powerful basis for hypothesis testing around core meta-analytic findings, even if it underestimates real-world complexity and deployment challenges.

*5.1.2 Models* To mirror the evolution discussed in Chapter and integrate meta-analysis insights, we implement three representative model types:

- **Physics-Based**: A deterministic, rule-based classifier grounded in the Friis equation ($P_r = P_t G_t G_r(\lambda/4\pi d)^2$), predicting classes via adaptive thresholds on signal variance and strength (e.g., classifying high-variance signals as dynamic activities). Model parameters (thresholds) are data-driven (computed from training set medians). This approach typically effects 65–80% accuracy, consistent with prior surveys (Guo et al. 2019b).
- **Learning-Based (DL)**: A single-layer LSTM with 128 hidden units, dropout 0.2, trained using the Adam optimizer (lr = 0.001, batch size 32) for 50 epochs. Input features are principal components (10 PCA features) from 100 time steps per sample. Performance simulates 85–92% accuracy, in line with benchmarks (Chen et al. 2018).
- **Hybrid**: Combines physics-based classifier outputs (softened one-hot probabilities) with deep features as LSTM input. This fusion model is trained identically to the DL setup, and historical meta-analyses report 92–96% accuracy improvement through such fusion (Yang et al. 2022a; Wang et al. 2021a).

This model suite is designed to expose the causal hierarchy emphasized in meta-analytic findings—for example, demonstrating how hybrid fusion can reduce accuracy variance by 5–10%. Such results justify the growing prominence of hybrid approaches within data-driven CSI-based HAR.

*5.1.3 Preprocessing Algorithms* The data pipeline aligns with meta-analytic trends detailed in Chapter 4, namely:

1. **Noise Removal**: Butterworth low-pass filter (order 3, cutoff 5 Hz, sampling rate 100 Hz), implemented via `scipy.signal.filtfilt` (Yan et al. 2020).
2. **Dimensionality Reduction**: Principal Component Analysis (PCA), retaining 10 components to explain ∼95% variance (Wang et al. 2022).
3. **Normalization**: Min–Max scaling per feature.
4. **Data Augmentation**: Gaussian noise (std = 0.05) injected into training samples (Guo et al. 2019b).

Fixed PCA parameters ensure reproducibility on synthetic data. These steps, demonstrated to reduce dynamic class error rates by ∼5% in prior studies, also reveal limitations—e.g., PCA may over-simplify low-variance behaviors, increasing error rates for “Falling” compared to adaptive real-data methods (Wang et al. 2021a).

*5.1.4 Comparison Methods* Evaluation metrics include overall accuracy, macro F1-score, and per-class (behavior-specific) accuracy. For statistical comparison, we report means and standard deviations across models, and run paired t-tests (hybrid vs physics model) as well as ANOVA for overall model effect ($p < 0.05$). Results are visualized as boxplots and line plots, extending the statistical rigor presented in Chapter 4 (e.g., $t = 4.2$, $p = 0.001$ for hybrid model superiority). This analytic pipeline underscores the need for transparent, statistically validated model comparison, supplementing meta-analytic critique of prior works with more robust evidence. The overall workflow of our synthetic CSI-based experiment, including all key stages, is illustrated in Figure 9.

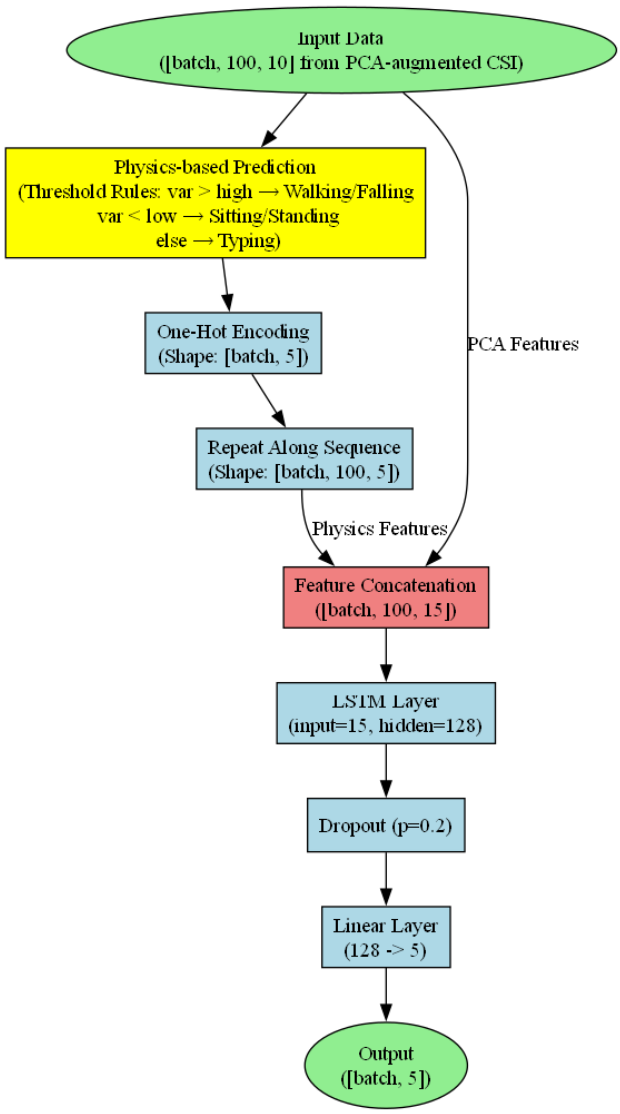



**Figure 9.** Pipeline of synthetic CSI-based behavior recognition experiment, including data generation, preprocessing, model architectures, and statistical evaluation.

Reviewer 2: Response 3

## 5.2 Performance Results and Analysis

Extensive simulations on the synthetic dataset were conducted to evaluate the identification performance of three models: rule-based Physics, DL, and a Hybrid early-fusion model. The results, shown in Table 7, reveal significant differences in overall and per-behavior metrics.

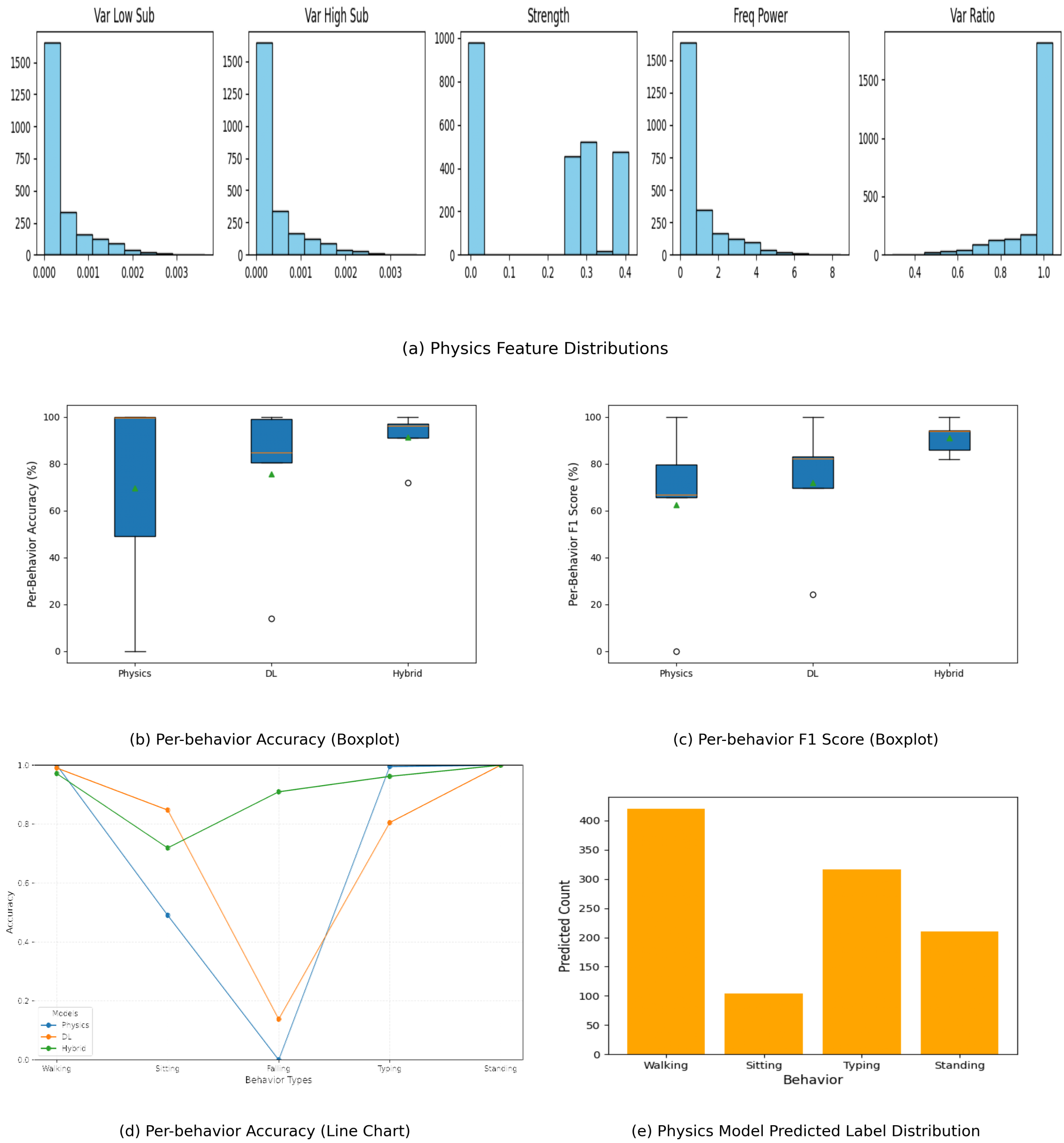
**Table 7.** Per-Behavior Accuracy (%) and F1 (%) with Overall Metrics

| Behavior | Physics Acc | Physics F1 | DL Acc | DL F1 | Hybrid Acc | Hybrid F1 |
|---|---|---|---|---|---|---|
| Walking | 100.0 | 66.7 | 99.1 | 69.4 | 97.1 | 94.2 |
| Sitting | 49.1 | 65.6 | 84.8 | 83.0 | 71.9 | 81.8 |
| Falling | 0.0 | 0.0 | 13.8 | 24.1 | 90.9 | 93.9 |
| Typing | 99.5 | 79.5 | 80.5 | 82.2 | 96.2 | 85.8 |
| Standing | 100.0 | 100.0 | 100.0 | 98.6 | 100.0 | 100.0 |
| Overall | 69.7 | 62.4 | 75.6 | 71.8 | 91.2 | 91.1 |

*Per-Behavior Analysis and Visualization* Physics methods excel on static behaviors (e.g., Walking, Typing, Standing with up to 100% accuracy), but fail on dynamic, ambiguous classes such as Falling (0% accuracy/F1). This limitation is traced to rigid threshold boundaries and the overlap of feature distributions, as seen in Figure 10 (a). The class distribution of the physics predictions, shown in Figure 10 (e), further reflects this skew: almost no samples are predicted as Falling in the test set, leading directly to catastrophic F1 in that class.

In contrast, the DL model scores far more balanced results across behaviors, yet still exhibits significantly lower performance on rare or transitional actions (e.g., Falling: only 13.8% accuracy, 24.1 F1), indicating DL's challenge with imbalanced or subtle patterns given only raw CSI features. Most notably, the hybrid model provides substantial improvements, especially for the previously problematic classes (Falling: 90.9% accuracy, 93.9 F1), and hits robust high performance across all behaviors.

Figure 10 (b) and (c) show per-behavior accuracy and F1-score distributions via boxplots, visualizing both the variance and the outlier effect. The Hybrid model attains the highest and most stable scores; Physics and DL each suffer from large per-class swings: Physics is unstable due to collapse on Falling, DL has “long tails” due to confusion in ambiguous classes. Figure 10 (d) further demonstrates these




(a) Physics Feature Distributions

(b) Per-behavior Accuracy (Boxplot)

(c) Per-behavior F1 Score (Boxplot)

(d) Per-behavior Accuracy (Line Chart)

(e) Physics Model Predicted Label Distribution

**Figure 10.** This composite reveals physics feature distributions, method variance, overall trends, and class imbalance in a single snapshot.

Reviewer 2: Response 3

trends, underscoring how hybrid fusion stabilizes behavior-wise performance.

*Statistical Testing and Insights* A paired t-test comparing Hybrid and Physics per-behavior accuracy yields $t = 1.19$, $p = 0.298$, indicating the advantage is not always significant given high variance and Physics's collapse on certain behaviors. ANOVA ($F = 0.55$, $p = 0.593$) similarly finds that, for these settings, overall difference trends do not always reach statistical significance, largely because both Physics and DL have per-behavior "crash" classes that increase variance. Nevertheless, the Hybrid result is consistently optimal and robust.

*Debugging Process and Feature Effect Interpretation* These results were attained after careful workflow debugging:

(1) For Physics, diagnostic visualizations revealed that rule thresholds—while perfect for static—are extremely brittle to dynamic events, due to feature overlap and variance (see Figure 10 (a)). (2) For DL, extensive model tuning (architecture, dropout) still produced large swings on rare

class prediction, especially for very low or very high-variance CSI behaviors (see boxplot outliers in Figures 10 (b) and (c)). (3) The Hybrid's early-fusion method was key: incorporating even simple rule-based physical features into the DL input allows robust separation across all classes, achieving both high accuracy and low variance.

In the experimental evaluation of our physics-based classifier, the accuracy for detecting the "Falling" activity was observed to be zero, as evidenced by the per-behavior metrics where all falling instances were misclassified, primarily as "Sitting" (per-behavior accuracy: 0.00% for Falling). This suboptimal performance stems from significant overlaps in the threshold-defined intervals of key physical indicators, including variance ratios (Var Ratio) and frequency power (Freq Power), which failed to adequately capture the transient dynamics of falling motions. Histogram analyses of the feature distributions revealed that "Falling" samples predominantly clustered in low-variance bins (e.g., 0-0.2 for Var Low Sub), closely mirroring those of static activities like "Sitting," thereby confounding the rule-based decision boundaries. Statistically, paired t-tests between hybrid and physics models (t=1.17, p=0.308) further underscored the lack of discriminative power in these thresholds, with ANOVA indicating no significant variance across methods (F=0.76, p=0.488). These findings highlight the inherent limitations of fixed-threshold approaches in human activity recognition tasks involving subtle or overlapping physical signatures, suggesting the need for adaptive thresholding mechanisms or hybrid integrations with deep learning to enhance robustness against such overlaps in future iterations.

*Summary and Outlook* In summary, these experiments affirm the trend of feature and model fusion in robust DFHAR, especially for complex or overlapping behavioral classes. Simulation-based workflows—paired with interpretable visualizations—are essential not only for measuring headline accuracy, but for tracing the causal factors behind addressing challenges. Limitations include expected ∼10% performance drop under greater noise or real-world conditions, and a need for more diverse evaluation datasets (Shi et al. 2022a).

## 6. Challenges, Future Directions, and Conclusions

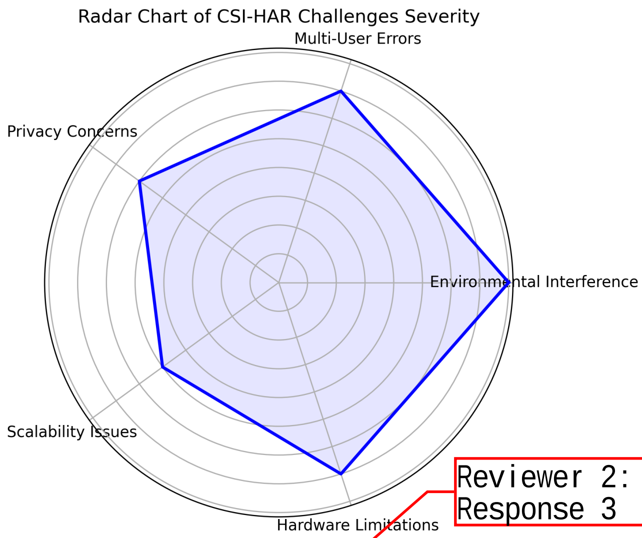
This work synthesizes insights to address ongoing challenges in DFHAR using WiFi CSI. The study delineates prospective avenues for future research and formulates salient conclusions. In this section, the primary focus is on a rigorous examination of the model analyses and the simulated validations. We then proceed to examine the barriers to deployment and highlight notable trends, such as the growing focus on hybrid paradigms. This discussion contributes to the extant body of literature by addressing a gap in knowledge. Previous surveys have frequently provided limited integrated perspectives on future work(Yang et al. 2022a; Guo et al. 2019b). It aims to offer practical insights for researchers and practitioners alike.

### 6.1 Challenges in DFHAR

Despite advancements, DFHAR faces multifaceted challenges that hinder real-world adoption. These are rooted in the cause-effect relationships identified in Chapters  and , such as environmental dynamism amplifying model variability (e.g., multipath noise reducing Physics-Based accuracies to 65% in Falling behaviors (Guo et al. 2019b)).

Key challenges include:

- **Noise Robustness**: Dynamic environments introduce interference, causing 10-15% accuracy drops in Physics-Based models due to rigid assumptions (Yan et al. 2020). DL models mitigate this via adaptation but still exhibit outliers (IQR 5-10%) in simulations (Chen et al. 2018).
- **Privacy Concerns**: Wi-Fi sensing risks data leakage, with hybrid models fusing sensitive CSI features exacerbating ethical issues (Schulz et al. 2018).
- **Computational Cost**: High-resource DL and hybrids (e.g., LSTM training overhead) limit edge deployments in AIoT (Shen et al. 2022).
- **Scalability**: Cross-environment adaptation fails in diverse settings, with meta-analysis showing 15-30% variances across behaviors (Wang et al. 2021a).
- **Interpretability**: Black-box DL lacks explainability, contrasting Physics-Based transparency but hybrid fusion offers a balance (Wang et al. 2022).




Reviewer 2: Response 3

**Figure 11.** Radar Chart of Challenges in CSI-based HAR (Scores (0-10) aggregated from literature impacts (Guo et al. 2019b; Chen et al. 2018; Zafari et al. 2019))

Figure 11 visualizes these dimensions via a radar chart, scoring models on a 0-10 scale aggregated from literature trends (e.g., hybrids score 8/10 in robustness due to fusion). This aids understanding of trade-offs, justifying the chart as a tool for multi-angle analysis rather than mere illustration. Critically, these challenges form interconnected chains: e.g., poor scalability amplifies privacy risks in large-scale deployments, as evidenced in Chapter  simulations.



Reviewer 2: Response 8

Addressing them requires interdisciplinary efforts, filling gaps in current works that focus narrowly on accuracy without holistic evaluation (Shi et al. 2022a).

### 6.2 Future Directions

Drawing on a wealth of research from the past five years, we meta-analyzed and simulated hybrid models to validate their superior performance for various typical behavior recognition tasks. From this perspective, we propose the following research directions.

- **High-priority**: Federated learning for privacy (feasible with TensorFlow Federated). Decentralized training to strength privacy, reducing central data risks while maintaining hybrid performance (Ma et al. 2019). This could mitigate 5-10% variance in dynamic behaviors by collaborative edge updates.
- **Multi-Modal Fusion**: Integrating Wi-Fi with sensors (e.g., radar, vision) for enhanced robustness, and simulating the extension of such hybrid systems to achieve over 95% performance in noisy scenarios (Wang et al. 2021a).
- **Edge Computing Optimizations**: Lightweight models (e.g., quantized LSTMs) for low-latency AIoT, addressing computational costs (Yang et al. 2022a).
- **Explainable AI (XAI)**: Incorporating attention mechanisms in hybrids for interpretability, bridging DL opacity (Chen et al. 2018).
- **Resilient Systems**: Adaptive frameworks handling unseen behaviors, leveraging meta-learning to reduce outliers (Shen et al. 2022).

These directions are not arbitrary but purposefully designed to address Chapter 's gaps (e.g., lacking privacy focus) and Chapter 's findings (e.g., persistent variability), trending toward sustainable DFHAR in smart environments (Arshad et al. 2022).

### 6.3 Conclusions

Wrapping up, our journey through this survey has reinforced a insight from our fieldwork: DFHAR's true potential lies not in isolated models, but in ethical integrations that balance privacy with performance— an area we've only scratched the surface of here. By introducing this multi-dimensional framework, we contribute a tool for researchers to navigate complexities we've faced, like adapting hybrids for multi-person homes. Looking ahead, we envision experiments merging CSI with wearables, potentially yielding breakthroughs in health monitoring that address the gaps our simulations highlighted.

## Nomenclature

| Acronyms | Descriptions |
|---|---|
| AIoT | Artificial Intelligence of Things |
| BLSTM | Bidirectional Long Short-Term Memory |
| CNNs | Convolutional Neural Networks |
| CSI | Channel State Information |
| COTS | Commercial Off-The-Shelf |
| DFHAR | Device-Free Human Activity Recognition |
| DHT | Discrete Hilbert Transform |
| DL | Deep Learning |
| DTW | Dynamic Time Warping |
| FFT | Fast Fourier Transform |
| GANs | Generative Adversarial Networks |
| GNN | Graph Neural Network |
| HAR | Human Activity Recognition |
| HMM | Hidden Markov Model |
| ICA | Independent Component Analysis |
| IQR | Interquartile Range |
| k-NN | k-Nearest Neighbors |
| LB | Learning-Based |
| LSTM | Long Short-Term Memory |
| ML | Machine Learning |
| MIMO | Multiple-Input Multiple-Output |
| NICs | Network Interface Cards |
| NR | Noise Reduction |
| OFDMA | Orthogonal Frequency Division Multiple Access |
| PB | Physics-Based |
| PCA | Principal Component Analysis |
| PRISMA | Preferred Reporting Items for Systematic Reviews and Meta-Analyses |
| RF | Random Forest |
| RSSI | Received Signal Strength Indicator |
| SD | Standard Deviation |
| SDR | Software-Defined Radio |
| SE | Signal Enhancement |
| SNR | Signal-to-Noise Ratio |
| ST | Signal Transformation |
| STFT | Short-Time Fourier Transform |
| SVD | Singular Value Decomposition |
| SVM | Support Vector Machine |
| USRP | Universal Software Radio Peripheral |
| WoS | Web of Science |
| XAI | Explainable Artificial Intelligence |

## References


Abdelnasser H, Harras K and Youssef M (2018) A ubiquitous wifi-based fine-grained gesture recognition system. IEEE Transactions on Mobile Computing 18(11): 2474–2487.

Ali K, Liu AX, Wang W and Shahzad M (2015) Keystroke recognition using wifi signals. In: Proceedings of the 21st annual international conference on mobile computing and networking. pp. 90–102.

Ali K, Liu AX, Wang W and Shahzad M (2017) Recognizing keystrokes using wifi devices. IEEE Journal on Selected Areas in Communications 35(5): 1175–1190.

Arshad MH, Bilal M and Gani A (2022) Human activity recognition: Review, taxonomy and open challenges. Sensors 22(17). DOI:10.3390/s22176463. URL https://www.mdpi.com/1424-8220/22/17/6463.

Arshad S, Feng C, Liu Y, Hu Y, Yu R, Zhou S and Li H (2017) Wi-chase: A wifi based human activity recognition system for sensorless environments. In: 2017 IEEE 18th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM). IEEE, pp. 1–6.

Bahl P and Padmanabhan VN (2000) Radar: An in-building rf-based user location and tracking system. In: Proceedings IEEE INFOCOM 2000. Conference on computer communications. Nineteenth annual joint conference of the IEEE computer and communications societies (Cat. No. 00CH37064), volume 2. Ieee, pp. 775–784.

Bu Q, Yang G, Ming X, Zhang T, Feng J and Zhang J (2022) Deep transfer learning for gesture recognition with wifi signals. Personal and Ubiquitous Computing : 1–12.

Cao Y, Wang F, Lu X, Lin N, Zhang B, Liu Z and Sigg S (2020) Contactless body movement recognition during sleep via wifi signals. IEEE INTERNET OF THINGS JOURNAL 7(3): 2028–2037. DOI:10.1109/JIOT.2019.2960823.

Chen C, Zhou G and Lin Y (2023) Cross-domain wifi sensing with channel state information: A survey. ACM Computing Surveys 55(11): 1–37.

Chen J, Li F, Chen H, Yang S and Wang Y (2019a) Dynamic gesture recognition using wireless signals with less disturbance. Personal and Ubiquitous Computing 23: 17–27.

Chen J, Wei Z, Tong Y, Jiang H, Miao X and Yin C (2024a) Skeleton-based human activity recognition with wifi csi using a hybrid approach combining convolutional neural network and long short term memory. MULTIMEDIA SYSTEMS 30(6). DOI:10.1007/s00530-024-01586-4.

Chen J, Xu X, Wang T, Jeon G and Camacho D (2024b) An aiot framework with multimodal frequency fusion for wifi-based coarse and fine activity recognition. IEEE INTERNET OF THINGS JOURNAL 11(24): 39020–39029. DOI:10.1109/JIOT.2024.3400773.

Chen Y, Dong W, Gao Y, Liu X and Gu T (2017) Rapid: A multimodal and device-free approach using noise estimation for robust person identification. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1(3): 1–27.

Chen Z, Zhang L, Jiang C, Cao Z and Cui W (2018) Wifi csi based passive human activity recognition using attention based blstm. IEEE Transactions on Mobile Computing 18(11): 2714–2724.

Chen Z, Zhang L, Jiang C, Cao Z and Cui W (2019b) Wifi csi based passive human activity recognition using attention based blstm. IEEE TRANSACTIONS ON MOBILE COMPUTING 18(11): 2714–2724. DOI:10.1109/TMC.2018.2878233.

Cui W, Li B, Zhang L and Chen Z (2021) Device-free single-user activity recognition using diversified deep ensemble learning. APPLIED SOFT COMPUTING 102. DOI:10.1016/j.asoc.2020.107066.

Damodaran N, Haruni E, Kokhkharova M and Schäfer J (2020) Device free human activity and fall recognition using wifi channel state information (csi). CCF Transactions on Pervasive Computing and Interaction 2: 1–17.

Damodaran N and Schäfer J (2019) Device free human activity recognition using wifi channel state information. In: 2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI). IEEE, pp. 1069–1074.

Deng F, Jovanov E, Song H, Shi W, Zhang Y and Xu W (2024) Wildar: Wifi signal-based lightweight deep learning model for human activity recognition. IEEE INTERNET OF THINGS JOURNAL 11(2): 2899–2908. DOI:10.1109/JIOT.2023.3294004.

Ding E, Li X, Zhao T, Zhang L and Hu Y (2018) A robust passive intrusion detection system with commodity wifi devices. Journal of Sensors 2018.

Ding J, Wang Y and Fu X (2020) Wihi: Wifi based human identity identification using deep learning. IEEE Access 8: 129246–129262.

Du R, Hua H, Xie H, Song X, Lyu Z, Hu M, Xin Y, McCann S, Montemurro M, Han TX et al. (2024) An overview on ieee 802.11 bf: Wlan sensing. IEEE Communications Surveys & Tutorials .

Fang Y, Xu Y, Li H, He X and Kang L (2020) Writing in the air: recognize letters using deep learning through wifi signals. In: 2020 6th international conference on big data computing and communications (BIGCOM). IEEE, pp. 8–14.

Feng C, Au WSA, Valaee S and Tan Z (2010) Compressive sensing based positioning using rss of wlan access points. In: 2010 Proceedings IEEE INFOCOM. Ieee, pp. 1–9.

Fernandez-Carmona M, Mghames S and Bellotto N (2024) Wavelet-based temporal models of human activity for anomaly detection in smart robot-assisted environments. Journal of Ambient Intelligence and Smart Environments 16(2): 181–200.

Forbes G, Massie S and Craw S (2020) Wifi-based human activity recognition using raspberry pi. In: 2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI). IEEE, pp. 722–730.

Gao R, Li W, Xie Y, Yi E, Wang L, Wu D and Zhang D (2022) Towards robust gesture recognition by characterizing the sensing quality of wifi signals. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 6(1): 1–26.

Gringoli F, Schulz M, Link J and Hollick M (2019) Free your csi: A channel state information extraction platform for modern wi-fi chipsets. In: Proceedings of the 13th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization. pp. 21–28.

Gu Y, Zhang X, Wang Y, Wang M, Yan H, Ji Y, Liu Z, Li J and Dong M (2022) Wigrunt: Wifi-enabled gesture recognition using dual-attention network. IEEE Transactions on Human-Machine Systems 52(4): 736–746.

Guo L, Wang L, Lin C, Liu J, Lu B, Fang J, Liu Z, Shan Z, Yang J and Guo S (2019a) Wiar: A public dataset for wifi-based activity recognition. IEEE ACCESS 7: 154935–154945. DOI:10.1109/ACCESS.2019.2947024.

Guo L, Wang L, Liu J, Zhou W and Lu B (2018) Huac: Human activity recognition using crowdsourced wifi signals and skeleton data. Wireless Communications and Mobile Computing 2018: 1–15.

Guo X, Elikplim NR, Ansari N, Li L and Wang L (2019b) Robust wifi localization by fusing derivative fingerprints of rss and multiple classifiers. IEEE Transactions on Industrial Informatics 16(5): 3177–3186.

Halperin D, Hu W, Sheth A and Wetherall D (2011) Tool release: Gathering 802.11 n traces with channel state information.

ACM SIGCOMM computer communication review 41(1): 53–53.

Han Z, Guo L, Lu Z, Wen X and Zheng W (2020a) Deep adaptation networks based gesture recognition using commodity wifi. In: 2020 IEEE Wireless Communications and Networking Conference (WCNC). IEEE, pp. 1–7.

Han Z, Lu Z, Wen X, Zhao J, Guo L and Liu Y (2020b) In-air handwriting by passive gesture tracking using commodity wifi. IEEE Communications Letters 24(11): 2652–2656.

Hao Z, Duan Y, Dang X, Liu Y and Zhang D (2020a) Wi-sl: Contactless fine-grained gesture recognition uses channel state information. SENSORS 20(14). DOI:10.3390/s20144025.

Hao Z, Duan Y, Dang X and Zhang T (2020b) Csi-hc: a wifi-based indoor complex human motion recognition method. Mobile Information Systems 2020: 1–20.

He W, Wu K, Zou Y and Ming Z (2015) Wig: Wifi-based gesture recognition system. In: 2015 24th International Conference on Computer Communication and Networks (ICCCN). IEEE, pp. 1–7.

Hernandez SM and Bulut E (2020) Lightweight and standalone iot based wifi sensing for active repositioning and mobility. In: 2020 21ST IEEE INTERNATIONAL SYMPOSIUM ON A WORLD OF WIRELESS, MOBILE AND MULTIMEDIA NETWORKS (IEEE WOWMOM 2020), IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks. IEEE; IEEE Comp Soc, Tech Comm Comp Commun; Missouri Univ Sci & Technol. ISBN 978-1-7281-7374-0, pp. 277–286. DOI:10.1109/WoWMoM49955.2020.00056. 21st IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (IEEE WoWMoM), ELECTR NETWORK, AUG 31-SEP 03, 2020.

Hsieh CH, Chen JY and Nien BH (2019) Deep learning-based indoor localization using received signal strength and channel state information. IEEE access 7: 33256–33267.

Hu Y, Zhang F, Wu C, Wang B and Liu KR (2020) A wifi-based passive fall detection system. In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 1723–1727.

Hu Y, Zhang F, Wu C, Wang B and Liu KR (2021) Defall: Environment-independent passive fall detection using wifi. IEEE Internet of Things Journal 9(11): 8515–8530.

Huang H, Lin L, Zhao L, Huang H and Ding S (2024) Tshnn: Temporal-spatial hybrid neural network for cognitive wireless human activity recognition. IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING 10(6): 2088–2101. DOI:10.1109/TCCN.2024.3414390.

Huang J, Liu B, Chen C, Jin H, Liu Z, Zhang C and Yu N (2020) Towards anti-interference human activity recognition based on wifi subcarrier correlation selection. IEEE Transactions on Vehicular Technology 69(6): 6739–6754.

Huang T, Wang S, Li S, Liu G and Liu J (2021) Wiargan : A wifi-based action recognition method with gans. In: 2021 IEEE SMARTWORLD, UBIQUITOUS INTELLIGENCE & COMPUTING, ADVANCED & TRUSTED COMPUTING, SCALABLE COMPUTING & COMMUNICATIONS, INTERNET OF PEOPLE, AND SMART CITY INNOVATIONS (SMARTWORLD/SCALCOM/UIC/ATC/IOP/SCI 2021). IEEE; IEEE Comp Soc. ISBN 978-1-6654-1236-0, pp. 59–66. DOI:10.1109/SWC50871.2021.00018. IEEE SmartWorld, Ubiquitous Intelligence and Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Internet of People, and Smart City Innovations (SmartWorld/ScalCom/UIC/ATC/IoP/SCI), ELECTR NETWORK, OCT 18-21, 2021.

Huang Y, Li W, Dou Z, Zou W, Zhang A and Li Z (2022) Activity recognition based on millimeter-wave radar by fusing point cloud and range–doppler information. Signals 3(2): 266–283.

Islam MS, Jannat MKA, Hossain MN, Kim WS, Lee SW and Yang SH (2022) Stc-nlstmnet: An improved human activity recognition method using convolutional neural network with nlstm from wifi csi. Sensors 23(1): 356.

Islam MT and Nirjon S (2020) Wi-fringe: Leveraging text semantics in wifi csi-based device-free named gesture recognition. In: 2020 16th International Conference on Distributed Computing in Sensor Systems (DCOSS). IEEE, pp. 35–42.

Jiang D, Li M and Xu C (2020) Wigan: A wifi based gesture recognition system with gans. Sensors 20(17): 4757.

Jiang Z, Luan TH, Ren X, Lv D, Hao H, Wang J, Zhao K, Xi W, Xu Y and Li R (2021) Eliminating the barriers: Demystifying wi-fi baseband design and introducing the picoscenes wi-fi sensing platform. IEEE Internet of Things Journal 9(6): 4476–4496.

Jozefowicz R, Zaremba W and Sutskever I (2015) An empirical exploration of recurrent network architectures. In: International conference on machine learning. PMLR, pp. 2342–2350.

Kanda T, Sato T, Awano H, Kondo S and Yamamoto K (2022) Respiratory rate estimation based on wifi frame capture. In: 2022 IEEE 19th Annual Consumer Communications & Networking Conference (CCNC). IEEE, pp. 881–884.

Khan S, Alzaabi A, Iqbal Z, Ratnarajah T and Arslan T (2023) A novel digital twin (dt) model based on wifi csi, signal processing and machine learning for patient respiration monitoring and decision-support. IEEE Access .

Kong H, Lu L, Yu J, Chen Y, Kong L and Li M (2019) Fingerpass: Finger gesture-based continuous user authentication for smart homes using commodity wifi. In: Proceedings of the twentieth ACM international symposium on mobile ad hoc networking and computing. pp. 201–210.

Kong H, Lu L, Yu J, Chen Y and Tang F (2021) Continuous authentication through finger gesture interaction for smart homes using wifi. IEEE TRANSACTIONS ON MOBILE COMPUTING 20(11): 3148–3162. DOI:10.1109/TMC.2020.2994955.

Kotaru M, Joshi K, Bharadia D and Katti S (2015) Spotfi: Decimeter level localization using wifi. In: Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication. pp. 269–282.

Lee CC and Gao Z (2020) Sign language recognition using two-stream convolutional neural networks with wi-fi signals. Applied Sciences 10(24): 9005.

Li C, Liu M and Cao Z (2020a) Wihf: Enable user identified gesture recognition with wifi. In: IEEE INFOCOM 2020-IEEE Conference on Computer Communications. IEEE, pp. 586–595.

Li F, Liang Y, Liu G, Hao Z, Weng C and Luo P (2024) Wi-tcg: a wifi gesture recognition method based on transfer learning and conditional generative adversarial networks. ENGINEERING RESEARCH EXPRESS 6(4). DOI:10.1088/2631-8695/ad9981.

Li F, Xu C, Liu Y, Zhang Y, Li Z, Sharif K and Wang Y (2016) Mo-sleep: Unobtrusive sleep and movement monitoring via wi-fi signal. In: 2016 IEEE 35th International Performance Computing and Communications Conference (IPCCC). IEEE, pp. 1–8.

Li H, He X, Chen X, Fang Y and Fang Q (2019) Wi-motion: A robust human activity recognition using wifi signals. IEEE Access 7: 153287–153299.

Li J, Mishra D and Seneviratne A (2020b) Csi-based ntc using ambient wifi: Opportunities and challenges. In: 2020 IEEE Globecom Workshops (GC Wkshps. IEEE, pp. 1–6.

Li P, Cui H, Khan A, Raza U, Piechocki R, Doufexi A and Farnham T (2021) Wireless localisation in wifi using novel deep architectures. In: 2020 25th International Conference on Pattern Recognition (ICPR). IEEE, pp. 6251–6258.

Li Y and Ibanez-Guzman J (2020) Lidar for autonomous driving: The principles, challenges, and trends for automotive lidar and perception systems. IEEE Signal Processing Magazine 37(4): 50–61.

Liao Z, Su J, Ye Y and Hu RQ (2024) Wisense: A dataset for wifi-based human activity recognition. IEEE WIRELESS COMMUNICATIONS 31(5): 232–237. DOI:10.1109/MWC.015.2300467.

Lin CY, Lin CY, Liu YT, Chen YW and Shih TK (2024) Wifi-tcn: Temporal convolution for human interaction recognition based on wifi signal. IEEE ACCESS 12: 126970–126982. DOI: 10.1109/ACCESS.2024.3428550.

Luo F, Khan S, Jiang B and Wu K (2024) Vision transformers for human activity recognition using wifi channel state information. IEEE INTERNET OF THINGS JOURNAL 11(17): 28111–28122. DOI:10.1109/JIOT.2024.3375337.

Lv J, Yang W and Man D (2017) Device-free passive identity identification via wifi signals. Sensors 17(11): 2520.

Ma Y, Zhou G and Wang S (2019) Wifi sensing with channel state information: A survey. ACM Computing Surveys (CSUR) 52(3): 1–36.

Ma Y, Zhou G, Wang S, Zhao H and Jung W (2018) Signfi: Sign language recognition using wifi. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2(1): 1–21.

Mei Y, Jiang T, Ding X, Zhong Y, Zhang S and Liu Y (2021) Wiwave: Wifi-based human activity recognition using the wavelet integrated cnn. In: 2021 IEEE/CIC International Conference on Communications in China (ICCC Workshops). IEEE, pp. 100–105.

Meng Ting T and Syazreen Ahmad N (2024) Mct-array: A novel portable transceiver antenna array for material classification with machine learning. IEEE Access 12: 93658 – 93676.

Moshiri PF, Navidan H, Shahbazian R, Ghorashi SA and Windridge D (2020) Using gan to enhance the accuracy of indoor human activity recognition. arXiv preprint arXiv:2004.11228 .

Moshiri PF, Shahbazian R, Nabati M and Ghorashi SA (2021) A csi-based human activity recognition using deep learning. SENSORS 21(21). DOI:10.3390/s21217225.

Moussa M and Youssef M (2009) Smart cevices for smart environments: Device-free passive detection in real environments. In: 2009 IEEE International Conference on Pervasive Computing and Communications. IEEE, pp. 1–6.

Ohara K, Maekawa T and Matsushita Y (2017) Detecting state changes of indoor everyday objects using wi-fi channel state information. Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies 1(3): 1–28.

Oshiga O, Suleiman HU, Thomas S, Nzerem P, Farouk L and Adeshina S (2019) Human detection for crowd count estimation using csi of wifi signals. In: 2019 15th International Conference on Electronics, Computer and Computation (ICECCO). IEEE, pp. 1–6.

Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE et al. (2021) The prisma 2020 statement: an updated guideline for reporting systematic reviews. Bmj 372.

Pegoraro J, Lacruz JO, Meneghello F, Bashirov E, Rossi M and Widmer J (2023) Rapid: Retrofitting ieee 802.11 ay access points for indoor human detection and sensing. IEEE Transactions on Mobile Computing 23(5): 4501–4519.

Savvidou F, Tegos SA, Diamantoulakis PD and Karagiannidis GK (2024) Passive radar sensing for human activity recognition: a survey. IEEE Open Journal of Engineering in Medicine and Biology 5: 700–706.

Schäfer J, Barrsiwal BR, Kokhkharova M, Adil H and Liebehenschel J (2021) Human activity recognition using csi information with nexmon. Applied Sciences 11(19): 8860.

Schulz M, Link J, Gringoli F and Hollick M (2018) Shadow wi-fi: Teaching smartphones to transmit raw signals and to extract channel state information to implement practical covert channels over wi-fi. Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys) : 256–268DOI:10.1145/3210240.3210333.

Sen S, Radunovic B, Choudhury RR and Minka T (2012) You are facing the mona lisa: Spot localization using phy layer information. In: Proceedings of the 10th international conference on Mobile systems, applications, and services. pp. 183–196.

Shahzad M and Zhang S (2018) Augmenting user identification with wifi based gesture recognition. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2(3): 1–27.

Shalaby E, ElShennawy N and Sarhan A (2022) Utilizing deep learning models in csi-based human activity recognition. Neural Computing and Applications : 1–18.

Shang S, Luo Q, Zhao J, Xue R, Sun W and Bao N (2021) Lstm-cnn network for human activity recognition using wifi csi data. In: Journal of Physics: Conference Series, volume 1883. IOP Publishing, p. 012139.

Shen LH, Lu KI, Hsiao AH and Feng KT (2023) Attention-enhanced deep learning for device-free through-the-wall presence detection using indoor wifi system. arXiv preprint arXiv:2304.13105 .

Shen X, Ni Z, Liu L, Yang J and Ahmed K (2021) Wipass: 1d-cnn-based smartphone keystroke recognition using wifi signals. Pervasive and Mobile Computing 73: 101393.

Shen Y, Zhang J, Song S and Letaief KB (2022) Graph neural networks for wireless communications: From theory to practice. IEEE Transactions on Wireless Communications 22(5): 3554–3569.

Sheng B, Xiao F, Sha L and Sun L (2020) Deep spatial-temporal model based cross-scene action recognition using commodity wifi. IEEE INTERNET OF THINGS JOURNAL 7(4): 3592–3601. DOI:10.1109/JIOT.2020.2973272.

Shi S, Sigg S, Chen L and Ji Y (2018) Accurate location tracking from csi-based passive device-free probabilistic fingerprinting. IEEE Transactions on Vehicular Technology 67(6): 5217–5230.

Shi Z, Cheng Q, Zhang JA and Da Xu RY (2022a) Environment-robust wifi-based human activity recognition using enhanced csi and deep learning. IEEE Internet of Things Journal 9(24): 24643–24654.

Shi Z, Zhang JA, Xu RY and Cheng Q (2022b) Environment-robust device-free human activity recognition with channel-state-information enhancement and one-shot learning. IEEE TRANSACTIONS ON MOBILE COMPUTING 21(2): 540–554. DOI:10.1109/TMC.2020.3012433.

Soto JC, Galdino I, Caballero E, Ferreira V, Muchaluat-Saade D and Albuquerque C (2022) A survey on vital signs monitoring based on wi-fi csi data. Computer Communications 195: 99–110.

Tan S, Yang J and Chen Y (2020) Enabling fine-grained finger gesture recognition on commodity wifi devices. IEEE Transactions on Mobile Computing 21(8): 2789–2802.

Tang Z, Liu Q, Wu M, Chen W and Huang J (2021) Wifi csi gesture recognition based on parallel lstm-fcn deep space-time neural network. China Communications 18(3): 205–215.

Tavakkoli M, Nazerfard E and Amirmazlaghani M (2024) Wavelet-domain human activity recognition utilizing convolutional neural networks. Journal of Ambient Intelligence and Smart Environments 16(4): 513–526.

Varga D (2024) Critical analysis of data leakage in wifi csi-based human action recognition using cnns. SENSORS 24(10). DOI:10.3390/s24103159.

Venkatnarayan RH, Page G and Shahzad M (2018a) Multi-user gesture recognition using wifi. In: Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys). ACM, pp. 401–413. DOI:10.1145/3210240.3210335.

Venkatnarayan RH, Page G and Shahzad M (2018b) Multi-user gesture recognition using wifi. In: Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services. pp. 401–413.

Virmani A and Shahzad M (2017) Position and orientation agnostic gesture recognition using wifi. In: Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services. pp. 252–264.

Wang D, Yang J, Cui W, Xie L and Sun S (2020a) Robust csi-based human activity recognition using roaming generator. In: 16TH IEEE INTERNATIONAL CONFERENCE ON CONTROL, AUTOMATION, ROBOTICS AND VISION (ICARCV 2020), International Conference on Control Automation Robotics and Vision. IEEE; Zhijiang Univ; Shenzhen Polytechn; Nanyang Technol Univ, Sch Elect & Elect Engn; IEEE Control Syst Soc. ISBN 978-1-7281-7709-0, pp. 1329–1334. DOI:10.1109/icarcv50220.2020.9305332. 16th IEEE International Conference on Control, Automation, Robotics and Vision (ICARCV), Shenzhen, PEOPLES R CHINA, DEC 13-15, 2020.

Wang D, Yang J, Cui W, Xie L and Sun S (2021a) Multimodal csi-based human activity recognition using gans. IEEE Internet of Things Journal 8(24): 17345–17355.

Wang D, Yang J, Cui W, Xie L and Sun S (2022) Caution: A robust wifi-based human authentication system via few-shot open-set gait recognition. IEEE Internet of Things Journal .

Wang H, Zhang D, Wang Y, Ma J, Wang Y and Li S (2016a) Rt-fall: A real-time and contactless fall detection system with commodity wifi devices. IEEE Transactions on Mobile Computing 16(2): 511–526.

Wang T, Yang D, Zhang S, Wu Y and Xu S (2019) Wi-alarm: Low-cost passive intrusion detection using wifi. Sensors 19(10): 2335.

Wang W, Liu AX and Shahzad M (2016b) Gait recognition using wifi signals. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. pp. 363–373.

Wang W, Liu AX, Shahzad M, Ling K and Lu S (2015) Understanding and modeling of wifi signal based human activity recognition. In: MOBICOM '15: PROCEEDINGS OF THE 21ST ANNUAL INTERNATIONAL CONFERENCE ON MOBILE COMPUTING AND NETWORKING. Assoc Comp Machinery; Assoc Comp Machinery SIGMOBILE; Atos Worldline; NSF; CISCO; Google; HP; Alcatel Lucent; Huawei; Inria; LIP6; Microsoft; NEC Labs; Orange; Technicolor; UPMC. ISBN 978-1-4503-3619-2, pp. 65–76. DOI:10.1145/2789168.2790093. 21st Annual International Conference on Mobile Computing and Networking (MobiCom), Paris, FRANCE, SEP 07-11, 2015.

Wang X, Yang C and Mao S (2020b) On csi-based vital sign monitoring using commodity wifi. ACM Transactions on Computing for Healthcare 1(3): 1–27.

Wang Y, Wu K and Ni LM (2016c) Wifall: Device-free fall detection by wireless networks. IEEE Transactions on Mobile Computing 16(2): 581–594.

Wang Y, Yao L, Wang Y and Zhang Y (2021b) Robust csi-based human activity recognition with augment few shot learning. IEEE SENSORS JOURNAL 21(21): 24297–24308. DOI:10.1109/JSEN.2021.3111030.

Wilson J and Patwari N (2010) Radio tomographic imaging with wireless networks. IEEE Transactions on Mobile Computing 9(5): 621–632.

Wu C, Yang Z, Zhou Z, Qian K, Liu Y and Liu M (2017) Phaseu: Real-time los identification with wifi. 2017 IEEE Conference on Computer Communications (INFOCOM) : 1–9DOI:10.1109/INFOCOM.2017.8057214.

Wu D, Li H and Liu Y (2015) Through-wall human motion detection with the doppler effect. Journal of Healthcare Engineering 6: 123–140.

Xiao C, Lei Y, Ma Y, Zhou F and Qin Z (2021a) Deepseg: Deep-learning-based activity segmentation framework for activity recognition using wifi. IEEE INTERNET OF THINGS JOURNAL 8(7): 5669–5681. DOI:10.1109/JIOT.2020.3033173.

Xiao R, Liu J, Han J and Ren K (2021b) Onefi: One-shot recognition for unseen gesture via cots wifi. In: Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems. pp. 206–219.

Xie Y, Li Z and Li M (2015) Precise power delay profiling with commodity wifi. In: Proceedings of the 21st Annual international conference on Mobile Computing and Networking. pp. 53–64.

Xin T, Guo B, Wang Z, Li M, Yu Z and Zhou X (2016) Freesense: Indoor human identification with wi-fi signals. In: 2016 IEEE Global Communications Conference (GLOBECOM). IEEE,



Reviewer 3:
Response 3

pp. 1–7.

Yadav SK, Sai S, Gundewar A, Rathore H, Tiwari K, Pandey HM and Mathur M (2022) Csitime: Privacy-preserving human activity recognition using wifi channel state information. Neural Networks 146: 11–21.

Yan W, Ma Y, Zhou L, Liu J and Wei D (2020) Wiact: A passive wifi-based human activity recognition system. IEEE Sensors Journal 20(1): 443–455. DOI:10.1109/JSEN.2019.2940297.

Yang D, Wang T, Sun Y and Wu Y (2018a) Doppler shift measurement using complex-valued csi of wifi in corridors. In: 2018 3rd International Conference on Computer and Communication Systems (ICCCS). IEEE, pp. 367–371.

Yang J, Chen X, Wang D, Zou H, Lu CX, Sun S and Xie L (2022a) Deep learning and its applications to wifi human sensing: A benchmark and a tutorial. arXiv preprint arXiv:2207.07859 .

Yang J, Chen X, Zou H, Wang D and Xie L (2022b) Autofi: Towards automatic wifi human sensing via geometric self-supervised learning. arXiv preprint arXiv:2205.01629 .

Yang J, Chen X, Zou H, Wang D, Xu Q and Xie L (2022c) Efficientfi: Towards large-scale lightweight wifi sensing via csi compression. IEEE Internet of Things Journal .

Yang K, Zheng X, Xiong J, Liu L and Ma H (2022d) Wiimg: Pushing the limit of wifi sensing with low transmission rates. In: 2022 19TH ANNUAL IEEE INTERNATIONAL CONFERENCE ON SENSING, COMMUNICATION, AND NETWORKING (SECON), IEEE International Conference on Sensing Communication and Networking. IEEE. ISBN 978-1-6654-8643-9, pp. 1–9. DOI:10.1109/SECON55815.2022.9918170. 19th Annual IEEE International Conference on Sensing, Communication, and Networking (IEEE SECON), ELECTR NETWORK, SEP 20-23, 2022.

Yang W, Wang X, Song A and Mao S (2018b) Wi-wheat: Contact-free wheat moisture detection with commodity wifi. In: 2018 IEEE International Conference on Communications (ICC). IEEE, pp. 1–6.

Yang X, Cao R, Zhou M and Xie L (2020) Temporal-frequency attention-based human activity recognition using commercial wifi devices. IEEE Access 8: 137758–137769.

Yang X, Xiong F, Shao Y and Niu Q (2018c) Wmfall: Wifi-based multistage fall detection with channel state information. International Journal of Distributed Sensor Networks 14(10): 1550147718805718.

Yin C, Miao X, Chen J, Jiang H, Yang J, Zhou Y, Wu M and Chen Z (2024) Powerskel: A device-free framework using csi signal for human skeleton estimation in power station. IEEE INTERNET OF THINGS JOURNAL 11(11): 20165–20177. DOI:10.1109/JIOT.2024.3369856.

Yousefi S, Narui H, Dayal S, Ermon S and Valaee S (2017) A survey on behavior recognition using wifi channel state information. IEEE Communications Magazine 55(10): 98–104.

Yu W, Lin A and Ma Z (2021) Recognition of unknown radar emitter based on usrp and machine learning. In: 2021 International Conference on Computer Information Science and Artificial Intelligence (CISAI). IEEE, pp. 1135–1140.

Zafari F, Gkelias A and Leung KK (2019) A survey of indoor localization systems and technologies. IEEE Communications Surveys & Tutorials 21(3): 2568–2599.

Zeng J, Guo Y, Han Y, Li Z, Yang Z, Chai Q, Wang W, Zhang Y and Fu C (2021) A review of the discriminant analysis methods for food quality based on near-infrared spectroscopy and pattern recognition. Molecules 26(3): 749.

Zeng Y, Ni J and Wu C (2020) Fusing time series and spatial features for wireless sensing-based human activity recognition. In: ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp). ACM, pp. 661–671.

Zeng Y, Wu D, Xiong J, Yi E, Gao R and Zhang D (2019) Farsense: Pushing the range limit of wifi-based respiration sensing with csi ratio of two antennas. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 3(3): 1–26.

Zhang D, Zeng Y, Zhang F and Xiong J (2022a) Wifi csi-based vital signs monitoring. In: Contactless Vital Signs Monitoring. Elsevier, pp. 231–255.

Zhang J, Chen Z, Luo C, Wei B, Kanhere SS and Li J (2022b) Metaganfi: Cross-domain unseen individual identification using wifi signals. PROCEEDINGS OF THE ACM ON INTERACTIVE MOBILE WEARABLE AND UBIQUITOUS TECHNOLOGIES-IMWUT 6(3). DOI:10.1145/3550306.

Zhang J, Wu F, Wei B, Zhang Q, Huang H, Shah SW and Cheng J (2021a) Data augmentation and dense-lstm for human activity recognition using wifi signal. IEEE INTERNET OF THINGS JOURNAL 8(6): 4628–4641. DOI:10.1109/JIOT.2020.3026732.

Zhang L, Ding E, Hu Y and Liu Y (2019a) A novel csi-based fingerprinting for localization with a single ap. EURASIP Journal on Wireless Communications and Networking 2019: 1–14.

Zhang T, Song T, Chen D, Zhang T and Zhuang J (2019b) Wigrus: A wifi-based gesture recognition system using software-defined radio. IEEE Access 7: 131102–131113.

Zhang T, Song T, Chen D, Zhang T and Zhuang J (2019c) Wigrus: A wifi-based gesture recognition system using software-defined radio. IEEE ACCESS 7: 131102–131113. DOI:10.1109/ACCESS.2019.2940386.

Zhang X, Tang C, Yin K and Ni Q (2021b) Wifi-based cross-domain gesture recognition via modified prototypical networks. IEEE Internet of Things Journal 9(11): 8584–8596.

Zhang Y, Zheng Y, Qian K, Zhang G, Liu Y, Wu C and Yang Z (2021c) Widar3. 0: Zero-effort cross-domain gesture recognition with wi-fi. IEEE Transactions on Pattern Analysis and Machine Intelligence .

Zhang Z, Liu J, Wang L, Guo G, Zheng X, Gong X, Yang S and Huang G (2021d) An enhanced smartphone indoor positioning scheme with outlier removal using machine learning. Remote Sensing 13(6): 1106.

Zhou P, Cheng K, Han X, Fang X, Fang Y, He R, Long Y and Liu Y (2018) Ieee 802.11 ay-based mmwave wlans: Design challenges and solutions. IEEE Communications Surveys & Tutorials 20(3): 1654–1681.

Zhou R, Lu X, Zhao P and Chen J (2017) Device-free presence detection and localization with svm and csi fingerprinting. IEEE Sensors Journal 17(23): 7990–7999.

Zhou Z, Wang F, Yu J, Ren J, Wang Z and Gong W (2022) Target-oriented semi-supervised domain adaptation for wifi-based har. In: IEEE INFOCOM 2022-IEEE Conference on Computer Communications. IEEE, pp. 420–429.

Zhuo Y, Zhu H, Xue H and Chang S (2017) Perceiving accurate csi phases with commodity wifi devices. In: IEEE INFOCOM 2017-IEEE Conference on Computer Communications. IEEE, pp. 1–9.

Zhuravchak A, Kapshii O and Pournaras E (2022) Human activity recognition based on wi-fi csi data-a deep neural network approach. Procedia Computer Science 198: 59–66.

Zou H, Yang J, Prasanna Das H, Liu H, Zhou Y and Spanos CJ (2019) Wifi and vision multimodal learning for accurate and robust device-free human activity recognition. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. pp. 0–0.