Device-free Human Activity Recognition for the Kitchen

Majid Ghosian Moghaddam

Thesis submitted to the University of Ottawa
in partial Fulfillment of the requirements for the
Doctorate in Philosophy (Ph.D.) degree
in Electrical and Computer Engineering

School of Electrical Engineering and Computer Science
Faculty of Engineering
University of Ottawa

Abstract

Device-free human activity recognition (HAR) is a promising approach for smart environments, but existing methods often experience performance drift with context-specific tasks, particularly in complex settings like kitchens. While prior work explored coarse activities (e.g., walking, sitting), fine-grained, context-specific tasks, such as kitchen activities involving object interactions, remain widely underexplored and are challenging due to signal noise, environmental variability, and limited datasets. In this thesis, we begin by evaluating three HAR approaches for kitchen tasks: deep learning (43% accuracy, limited by small datasets), transfer learning (39%, hindered by domain mismatch), and classical machine learning (68%, which proved best suited for the context-specific data). Given the superior performance of classical learning, its reliance on effective feature selection becomes a critical challenge. Current methods often rely on heuristics to select features, which can limit performance and require domain knowledge. To address this, we propose an optimized method that combines multivariate analysis of variance (MANOVA) with genetic algorithms for discriminative feature selection. This approach improves accuracy to 80% for coarse-grained tasks such as storing food and washing pots. However, fine-grained motions (chopping and slicing potatoes) remain challenging (with average recognition accuracy of 37%), highlighting the limitations of WiFi sensing for subtle activities. The key contributions of this thesis include: (1) an evaluation of HAR methods for kitchen tasks, (2) a novel approach to feature selection based on *MANOVA-genetic* algorithm to generalize the feature selection process, and (3) a public WiFi-based dataset to advance research in context-aware HAR. These findings highlight the potential of classical learning in context-specific tasks while revealing open challenges in fine-grained activity recognition.

Acknowledgement

The completion of this dissertation would not have been possible without the unwavering support, guidance, and encouragement of many individuals to whom I owe a profound debt of gratitude.

First and foremost, I extend my deepest appreciation to my supervisor, Dr. Shervin Shirmohammadi. His expert guidance, immense patience, and invaluable insights were instrumental in shaping this research from its inception to its final form. His rigorous academic standards and constant encouragement challenged me to strive for excellence and persevere through challenges. I am truly fortunate to have had the opportunity to learn from his expertise.

I am also profoundly grateful to Dr. Ali Asghar Nazari Shirehjini, for his invaluable advice and steadfast support throughout my academic journey. His thoughtful feedback and constructive critiques were essential in refining the direction and quality of this work.

On a personal note, my most heartfelt thanks go to my beloved wife, Zahra. Her unconditional love, endless patience, and immense sacrifices provided the emotional foundation that sustained me through the most demanding periods of this research. She was my greatest source of strength and comfort, and this achievement is as much hers as it is mine.

I owe a special debt of gratitude to my parents and my entire family. Their boundless love, unwavering belief in me, and constant encouragement have been a guiding light throughout my life and academic pursuits. Their support has always been my greatest blessing.

Finally, I would like to acknowledge the use of DeepSeek, an AI language model, which was employed solely for the purpose of proofreading and polishing the English language of this manuscript. The core research, ideas, analysis, and conclusions presented herein are entirely my own.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

5G	Fifth Generation (mobile networks)
6G	Sixth Generation (mobile networks)
ADC	Analog to Digital Converter
AI	Artificial Intelligence
ANN	Artificial Neural Network
AP	Access Point
BPSK	Binary Phase Shift Keying
BSSID	Basic Service Set Identifier
CFR	Channel Frequency Response
CNN	Convolutional Neural Network
CSI	Channel State Information
DC	Direct Current
DL	Deep Learning
DNN	Deep Neural Network
DT	Decision Tree
DTW	Dynamic Time Warping
DfP	Device-free Passive (localization)
DWT	Discrete Wavelet Transform
ELM	Extreme Learning Machine
ESP32	Espressif Systems' microcontroller
FMCW	Frequency Modulated Continuous Wave
FT	Fourier Transform
FWHM	Full Width at Half Maximum
GA	Genetic Algorithm
HAR	Human Activity Recognition
HMM	Hidden Markov Model
H-to-H	Human-to-Human (interaction)
H-to-O	Human-to-Object (interaction)
HT	High Throughput (as in 802.11n signal mode)

IDF	IoT Development Framework (Espressif)
IFFT	Inverse Fast Fourier Transform
IEEE	Institute of Electrical and Electronics Engineers
IFT	Inverse Fourier Transform
IoT	Internet of Things
KNN	K Nearest Neighbors
LDPC	Low Density Parity Check
LED	Light Emitting Diode
LoS	Line of Sight
LSTM	Long Short Term Memory
MAC	Media Access Control
MANOVA	Multivariate Analysis of Variance
MIMO	Multiple Input Multiple Output
ML	Machine Learning
MLP	Multi-Layer Perceptron
mmWave	Millimeter Wave
MPDU	MAC Protocol Data Unit
NLoS	Non-Line of Sight
NP-CSI	Normalized Power - Channel State Information
OFDM	Orthogonal Frequency Division Multiplexing
PCA	Principal Component Analysis
PCB	Printed Circuit Board
PDR	Packet Delivery Ratio
PSD	Power Spectral Density
RF	Radio Frequency
RFID	Radio Frequency Identification
RIS	Reconfigurable Intelligent Surfaces
RNN	Recurrent Neural Network
RSSI	Received Signal Strength Indicator
SLR	Systematic Literature Review
SMO	Sequential Minimal Optimization
SNR	Signal to Noise Ratio

SSID	Service Set Identifier
SSNR	Sensing-Signal-to-Noise Ratio
STA	Station
STBC	Space Time Block Coding
SVM	Support Vector Machine
UART	Universal Asynchronous Receiver/Transmitter
Wi-Fi	Wireless Fidelity
Wi-GRAK	WiFi Granular Activities in Kitchen Dataset
XGBoost	Extreme Gradient Boosting

Chapter 1 Introduction

1.1 Motivation

The increasing integration of technology into living spaces is fostering a new paradigm: privacy-aware insights-driven intelligent assistants for shopping or assisted living in smart homes. Smart homes are interconnected environments equipped with various sensors that help users manage and control daily tasks [1]. These sensors generate a continuous stream of data, when combined with AI assisted analysis, providing insights on user behavior, preferences, and routines. Such data serve as the primary source for recommender systems, which analyze patterns to provide personalized suggestions, from the next show to watch to the optimal temperature setting or products and services to purchase.

Recently, the concept of data monetization has entered the smart home domain [2]. In its simplest form, user-generated data from smart homes, with the owner's consent, can be shared with service providers with appropriate compensation. Service providers, in turn, use this data to deliver highly personalized recommendations, tailored to the user's needs and delivered at the right time. This approach enhances user engagement, drives revenue, and effectively turns the smart home into a platform for automated, intelligent, and commercially valuable services.

Recognizing human activities is important because it provides one of the most valuable forms of data in smart homes, enabling personalized services and intelligent decision-making. Various approaches have been explored to capture such activities, including camera-based monitoring and wearable sensors. While cameras can provide detailed information, they raise serious privacy concerns and are often unsuitable for continuous monitoring. Wearable devices, on the other hand, are intrusive and inconvenient for daily use. This is where device-free human activity recognition becomes essential. It offers the capability to monitor activities without visual recording or required user instrumentation. This makes it uniquely suited for privacy-critical and sensitive environments where other modalities are ethically or practically infeasible, such as monitoring for falls or distress in a hospital washroom or tracking sleep patterns and nighttime routines in a senior living facility's bedroom.

1.2 Problem Statement

Recommend systems and intelligent assistants like personalized shopping bots rely heavily on high-quality input data to generate accurate and relevant suggestions. Within smart homes, human activities provide a rich and context-aware source of such data. By understanding a user's actions, a recommender system can move beyond generic suggestions and instead deliver personalized and timely recommendations that align with the user's daily routines.

A variety of sensor modalities, such as cameras, wearables, and ambient sensors, has been proposed to recognize human activities in smart environments [3]. Each, however, carries specific strengths and limitations. Cameras raise privacy concerns, wearables require active user compliance, and ambient sensors often lack precision. This highlights the need for alternative approaches that are less intrusive, more practical, and still capable of accurately capturing human activity patterns.

An additional complexity arises from the nature of the activities themselves. Human activities can be recognized at different levels of granularity, ranging from coarse-grained actions (e.g., cooking, cleaning) to fine-grained movements (e.g., chopping vegetables, stirring a pot).

Device-free HAR, which leverages wireless sensing, has recently emerged as a promising solution. Numerous methods have been proposed in literature, often tailored to specific activities or environments. However, the efficacy of these methods in capturing and recognizing kitchen-related activities remains unclear. The kitchen presents a uniquely challenging and informative domain for HAR, especially for device-free sensing, for several reasons that distinguish it from less complex spaces like bedrooms or living rooms. First, it is a high-fidelity activity hub: it hosts a wide variety of structured, goal-oriented tasks (food preparation, cleaning, appliance use) with rich, sequential semantics. Second, these activities involve a high degree of object interaction and fine-grained hand-arm motility within a confined area, creating complex, overlapping signal patterns. Third, the environment itself is signal-challenging: it contains metallic appliances, reflective surfaces, and water, all of which significantly impact radio frequency (RF) propagation and multi-path effects. In contrast, activities in other environments like a bedroom (e.g., sleeping, reading, dressing) or living room (e.g., watching TV, relaxing) are often more static, involve fewer complex object manipulations, and occur in less RF-hostile environments. Therefore, the kitchen serves as a rigorous proving ground for HAR systems.

This dissertation addresses this gap by investigating whether a device-free HAR method existing for other domains can be utilized to recognize kitchen activities, and by developing and evaluating a novel device-free HAR method to recognize kitchen activities in various levels of granularity.

It is important to frame this investigation as foundational research. The current performance limitations of WiFi-based sensing for fine-grained activities, as documented in this dissertation, are not presented as a final solution ready for immediate commercial deployment. Rather, this work serves as a critical stepping stone in a nascent field. It systematically identifies the boundaries and challenges, such as the impact of granularity and sensor placement, and establishes empirical baselines and methodologies. The goal is to build a rigorous understanding of the problem space, which is a necessary precursor to developing practical, high-accuracy systems. Future practical implementations will likely integrate WiFi sensing with other complementary modalities or more advanced signal processing; this dissertation provides the essential groundwork for those advancements.

1.3 Research Objectives

The following research objectives are established to guide this study:

i. To investigate the feasibility of device-free human activity recognition (HAR) in recognition of kitchen activities.

ii. To evaluate and compare the performance of existing HAR methods for recognizing kitchen activities.

   - To classify the types of kitchen activities that can be recognized using device-free methods.

iii. To assess the impact of environmental and contextual factors on the performance of device-free HAR.

iv. To develop and evaluate a novel device-free HAR method to recognize kitchen activities.

v. To propose a structured framework and best practices for conducting device-free HAR research.

1.4 Contributions

This dissertation makes contributions to both the general field of device-free human activity recognition and, more specifically, to its application in kitchen environments. The general contributions are as follows:

- Conduct a systematic literature review that synthesizes existing methods and research trends in device-free HAR. [32]
- Propose a comprehensive framework outlining the key stages and requirements for deploying device-free HAR systems in smart home environments.

In addition to these general contributions, this work provides the following specific advancements for device-free HAR in kitchen environments:

- The design and implementation of a novel device-free HAR system tailored for recognizing both coarse- and fine-grained kitchen activities [47][106][107][177][178].
- The creation and public release of a rich WiFi Channel State Information (CSI) and Received Signal Strength Indicator (RSSI) dataset capturing a diverse set of kitchen activities [190].
- A novel evaluation method that progressively assesses recognition accuracy, demonstrating the performance drift as the system transitions from recognizing coarse-grained to fine-grained activities.
- An empirical demonstration of the efficacy of a device-free HAR in recognizing complex, sequential kitchen activities.
- An analysis of the impact of sensor placement on system performance, providing practical guidelines for deployment in real kitchens [73].
- A critical discussion about the applicability and limitations of state-of-the-art device-free HAR approaches within the challenging kitchen environment.

1.5 Research Publication

1.5.1. Journals

- M. G. Moghaddam, A. R. Nowdeh, N. Warnakulasuriya, G. S. S. P. Pala, A. N. Shirehjini, and S. Shirmohammadi, "Descriptor: WiFi-Granular Activities in Kitchen Dataset (Wi-GRAK)," *IEEE Data Descriptor*, Submitted.
- M. G. Moghaddam, A. A. N. Shirehjini and S. Shirmohammadi, "Device-free Kitchen Activity Recognition via AI-assisted Wireless Sensing," in *IEEE Transactions on Instrumentation and Measurement*, Submitted.
- M. G. Moghaddam, A. A. N. Shirehjini and S. Shirmohammadi, "Device-Free Human Activity Recognition: A Systematic Literature Review," in *IEEE Open Journal of Instrumentation and Measurement*, vol. 4, pp. 1-34, **2025**, Art no. 9500134, doi: 10.1109/OJIM.2024.3502885.
- M. G. Moghaddam, A. A. N. Shirehjini and S. Shirmohammadi, "A WiFi-based Method for

Recognizing Fine-grained Multiple-Subject Human Activities," in *IEEE Transactions on Instrumentation and Measurement*, **2022**, doi: 10.1109/TIM.2023.3289547.

1.5.2. Conferences

- M. G. Moghaddam, A. A. N. Shirehjini and S. Shirmohammadi, "The Effect of Activity Granularity on A Kitchen Activity Recognition System," *2024 IEEE Sensors Applications Symposium (SAS)*, Naples, Italy, **2024**, pp. 1-5, doi: 10.1109/SAS60918.2024.10636508.
- M. G. Moghaddam, A. A. Nazari Shirehjini and S. Shirmohammadi, "The Effect of Sensor Placement in a Cooking Activity Recognition System," *2024 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, Glasgow, United Kingdom, **2024**, pp. 1-6, doi: 10.1109/I2MTC60896.2024.10560619.
- M. G. Moghaddam, A. A. Nazari Shirehjini and S. Shirmohammadi, "Device-Free Fine-Grained Dining Activity Sensing," *2023 IEEE Sensors Applications Symposium (SAS)*, Ottawa, ON, Canada, **2023**, pp. 1-6, doi: 10.1109/SAS58821.2023.10254010.
- M. G. Moghaddam, A. A. N. Shirehjini and S. Shirmohammadi, "A WiFi-based System for Recognizing Fine-grained Multiple-Subject Human Activities", *2022 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, Ottawa, ON, Canada, **2022**, pp. 1-6, doi: 10.1109/I2MTC48687.2022.9806622.

1.5.3. Datasets

- M. G. Moghaddam, A. R. Nowdeh, N. Warnakulasuriya, G. S. S. P. Pala, A. N. Shirehjini, and S. Shirmohammadi, “Wi-GRAK,” Kaggle, **2025**, doi: 10.34740/KAGGLE/DSV/12123800.
- M. G. Moghaddam, A. A. Nazari Shirehjini, and S. Shirmohammadi, “WiFi (CSI and RSSI) Data of three Fine-grained Knife Activities (Chop, Slice, French Cut) in an Authentic Kitchen with Different Sensor Placements ”, Dec. **2023**, figshare. Dataset. https://doi.org/l0.6084/m9.figshare.24750738.vl.
- M. G. Moghaddam, A. A. N. Shirehjini and S. Shirmohammadi, "WiFi (CSI and RSSI) Data of Six Basic Knife Activities for Cooking (chopping, cubing, French cutting, julienning, mincing, and slicing)", IEEE Dataport, February 26, **2023**, doi:10.21227/gq04-b598

1.6 Thesis Organization

This dissertation is organized into seven chapters. Chapter 1 has introduced the motivation, research objectives, key contributions, and a list of publications originating from this work.

Chapter 2 provides the necessary background for the study, beginning with a comprehensive survey of human activity recognition research from the past decade. It discusses various sensor modalities used in HAR and justifies the selection of a device-free, WiFi-based approach for this research. The chapter also explains fundamental concepts critical to understanding WiFi sensing, including line-of-sight (LOS) propagation, Orthogonal Frequency-Division Multiplexing (OFDM), multipath effects, and the specifics of Channel State Information (CSI) and Received Signal Strength Indicator (RSSI). Finally, it reviews the machine learning, statistical, and optimization methods utilized in subsequent chapters.

Chapter 3 reviews the related literature, positioning this dissertation within the current state of the art. It analyzes key studies in HAR, discussing their advantages and drawbacks. A specific focus is placed on the importance of context specific HAR systems and on prior work that has targeted kitchen activities.

In Chapter 4, the research problem is formally formulated. The chapter begins by elaborating on the smart home ecosystem, personalized advertising, and data monetization, which collectively form the core motivation for this work. It then discusses the value of activity data and the role of HAR in unlocking it. The chapter concludes by defining key terminology and presenting a formal problem statement.

Chapter 5 details four preliminary feasibility studies conducted to validate the core assumptions, ensure the feasibility of the main research, and inform the final system design.

Chapter 6 presents the proposed device-free HAR system. It describes the sensor hardware, the data collection process, and the constructed dataset of kitchen activities. The chapter then outlines the experimental method, including the design of pre-experiments to select models, the evaluation of various machine learning methods, and the optimization of the best-performing approach.

Chapter 7 presents a comprehensive discussion of the research outcomes. It provides a detailed interpretation of the experimental results, a critical comparison of the proposed system's performance against existing methods, and an examination of the study's limitations. Finally, Chapter 8 serves as the concluding chapter, synthesizing the key findings and contributions of this research. It provides a final summary of how the work has addressed the initial research problems and outlines specific, promising directions for future work to build upon the foundations laid here.

Chapter 2 Background

2.1 Device-free Human Activity Recognition

Human Activity Recognition (HAR) is defined as the ability to identify and classify human activities based on data collected from sensors. HAR has significant potential for various applications [4]. In healthcare [5], for instance, detecting patients' activities helps caregivers monitor them more easily [6], [7], [8], [9]. Detecting anomalies in daily routines for security purposes [4], [10], [11] represents another key use case. Furthermore, within the domain of human-computer interaction, researchers have utilized HAR to capture user movements, postures , gestures, and actions; examples include detecting teachers' activities in classrooms [12] or children's activities in various environments [13].

Human activity recognition can be categorized into device-based, camera-based, and device-free methods [14]. In device-free methods, sensors (e.g., radars, Wi-Fi transmitters, acoustic sensors) propagate signals (e.g., sound, light, or radio waves) into the environment. These systems recognize activities by analyzing the reflections of these signals from various elements, such as walls, furniture, and users. In contrast to device-based approaches that require wearable sensors, the sensors in device-free systems are embedded in the environment, meaning the user is not required to carry any device. Among these modalities, device-free methods have gained significant popularity. This is primarily because, compared to camera-based approaches, they better preserve user privacy and, as stated, do not require the user to carry any additional devices.

Several categorizations for device-free HAR methods exist in the literature. This work adopts the taxonomy provided by [15] , which categorizes methods based on the underlying sensing technology:

Radio Frequency (RF): Radio Frequency (RF)-based sensing is the most prevalent device-free HAR method. These systems typically comprise a transmitter and one or more receivers. The transmitted RF signals propagate through the environment, reflecting off objects and humans. WiFi is a widely used source of RF signals for this purpose. Researchers leverage fine-grained Channel State Information (CSI) or coarse-grained Received Signal Strength Indicator (RSSI) to analyze alterations in the propagated WiFi signals caused by human activity [16].

Acoustic Signal: Acoustic-based methods utilize sound waves for activity detection. A transmitter emits sound waves, and one or more receivers (microphones) analyze the received waves to detect environmental changes indicative of human activity. While similar in principle to RF systems, the use of microphones as receivers can raise privacy concerns for users, as they may potentially capture private conversations [17].

Visible Light: Visible light methods employ embedded light sources (e.g., LEDs) and photodetectors. The HAR system recognizes activities by inferring changes in the environment based on the amount of light received by each sensor [18].

2.1.1. Review of Existing Literature Reviews

To establish a comprehensive background for device-free HAR, particularly in kitchen environments, an analysis of existing literature reviews in the field was conducted. These reviews each illuminate a specific subset of primary HAR studies, offering valuable categorizations and conclusions. Their scopes vary; some categorize HAR methods broadly based on sensor modality, while others focus on a single modality, providing a more detailed taxonomy (see Table 1). Certain reviews concentrate on specific modalities such as radio frequency (RF) [19], [20], [21], [22], WiFi [23], [24], or WiFi Channel State Information (CSI) [25], [26], [27], [28], while others focus on particular human activities or unique research conditions, as summarized in Table 1.

A comprehensive categorization of HAR sensing technologies is provided by [29]. The authors categorized studies based on sensor modality into six types: (1) Acoustic sensors, which measure waves traveling through materials (subcategories: acoustic, surface acoustic, ultrasonic); (2) Electric sensors, which measure charge from an electrified object (subcategories: capacitive, electrostatic); (3) Mechanical sensors, which indicate applied force (subcategories: resistive pressure, integrated pressure sensing); (4) Optical sensors, which quantify light intensity (subcategories: visible spectrum, depth channel, infrared); (5) Radiation sensors, which use high-frequency electromagnetic waves; and (6) Hybrid methods, which combine multiple sensor types.

Lentzas and Vrakas [30] propose a dual taxonomy for HAR methods. The first categorization is based on sensor types, including smartphones/wearables, radiofrequency identification, ambient sensors, and hybrid sensing. The second categorization is based on the artificial intelligence techniques employed. These techniques include decision trees (DT), random forests (RF), rule-based methods, hidden Markov models (HMM), hybrid HMM-DT approaches, support vector machines (SVM), deep learning (DL) methods, dictionary learning, the Dempster–Shafer method, threshold-based approaches, and probabilistic behavior models. The authors conclude that hybrid models are particularly effective for detecting activities with similar characteristics.

Other literature reviews proposed similar categorizations for device-free HAR, with a particular emphasis on RFID. Hussain et al. [14] introduce ten metrics that they choose for the comparison of different device-free RFID-based HAR methods. The metrics are approach, technology, information type, utilized machine learning (ML) algorithm, supervised/unsupervised classification method, application, cost, accuracy, latency, and real-time. They divided the device-free methods into three categories: action-based, interaction-based, and motion-based methods. In a narrative review [31], Gupta et al. conducted a comparison among different HAR methods. In contrast to other reviews, in their work, Gupta et al. considered device-free and RFID-based approaches into two separate categories. They categorized the rest of the papers as vision-based and sensor-based. However, in their paper, the focus of the categorization is the AI module as the core component of the HAR research method. They provided a critical analysis of existing HAR methods and compared the number of features, feature extraction method, ML/DL model, architecture, evaluation metrics, model validation, and optimization techniques of each AI module.

Table 1 Literature reviews focused on a specific sensor modality of device-free HAR

Ref.	Year	Modality	Activity	Other Condition
Our SLR [32]	2025	WiFi, radar, RFID, acoustic, ultrasonic, visible spectrum, infrared, hybrid	-	Systematic review
[33]	2025	WiFi	-	
[34]	2024	WiFi	-	
[35]	2023	-	Limb motion, hand gesture, lip tracking, movement tracking	
[36]	2022	Mechanical, RF, Acoustic, Optic, Physiological, Field Sensing	-	
[23]	2020	WiFi	-	Deep learning
[28]	2020	WiFi-CSI	Gesture	
[19]	2020	RF	-	
[26]	2019	WiFi-CSI	Behavior	Through-the-wall
[27]	2019	WiFi-CSI	Behavior	
[25]	2019	WiFi-CSI	Motion	
[24]	2018	WiFi	-	
[37]	2018	WiFi	-	
[21]	2017	RF	-	
[20]	2016	RF	Emotion	
[22]	2015	RF	Any	

Other reviews have focused on wireless sensing rather than the device-free paradigm specifically. For instance, Liu et al. [19] surveyed studies using WiFi data (e.g., RSSI, CSI, FMCW, Doppler effect). They categorized these articles based on recognition objective: (1) room-level activity sensing, (2) activity recognition, (3) gesture recognition, (4) vital sign monitoring, and (5) user identification and localization, further classifying them by the type of WiFi data used. Wang and Zhou [22] reviewed radio-based HAR methods, dividing them into four categories: (1) ZigBee, (2) WiFi, (3) RFID, and (4) other (e.g., FM radio, microwave). It is critical to distinguish between radio-based and device-free methods. Radio-based methods encompass any approach using radio waves, including those requiring wearable or object-tagged sensors, and are therefore not exclusively device-free. Conversely, device-free methods, while often using radio frequencies (e.g., WiFi, RFID), are not limited to them and also include acoustic or optical sensors, with the defining characteristic being that the user carries no device.

Recent literature reviews have built upon earlier efforts by adopting more systematic [32] and comprehensive [3], [34] approaches. Examples include the qualitative analysis of device-free HAR techniques based on sensing modality and processing technique in [38] ; the in-depth review of sensing technologies and their limitations in [36] ; the focused survey of Wi-Fi sensing challenges in [33] ; and the exploration of wireless, device-free sensing solutions across diverse applications in [35].

While these reviews provide valuable insights, a gap was identified for a systematic literature review (SLR) on device-free HAR that is not restricted to a specific sensor type or activity domain [32]. To address this, a systematic literature review was conducted as part of this dissertation. This review explored major digital libraries—including ACM Digital Library, IEEE Xplore, ScienceDirect, Scopus, and Web of Science—to identify relevant studies. The resulting SLR employs a mixed-methods approach to assess the quality of the reviewed papers and analyze the proposed HAR methods from both scientometric and technical perspectives. The scientometric analysis investigates publication trends from 2017 to 2023, examining distributions by country, institution, and venue. The technical analysis categorizes methods based on device-free sensing modalities, activity type and granularity, and discusses common challenges. It further compares methods based on their support for non-line-of-sight, multi-subject, user-independent, and environment-independent recognition. This review provides foundational knowledge on every stage of the device-free HAR pipeline (data acquisition, preprocessing, classification, and evaluation) and, most importantly, identifies critical research gaps and open questions that inform the work presented in this dissertation.

2.2 Wi-Fi-Based Device-Free HAR

Among all device-free HAR modalities, those based on radio frequency (RF) signals, and specifically WiFi-based systems, have garnered significant attention [32]. The primary reason for this is the ease of deployment and availability of existing WiFi infrastructure in most indoor environments, which eliminates the need for dedicated sensor networks. Furthermore, the ability of WiFi signals to penetrate walls and obstacles makes them suitable for cross-room monitoring and non-line-of-sight (NLOS) scenarios. The pervasive deployment of WiFi means that a WiFi-based HAR system can be readily implemented in environments such as smart homes, offices, and hospitals without requiring specialized hardware installation. This ubiquity makes WiFi-based HAR suitable for a wide range of applications, including elder care, health monitoring [39], energy management [40], and security [41].

2.2.1. Multipath Effects and Human Influence

As previously established, WiFi signals are electromagnetic waves that propagate through the environment between a transmitter and a receiver. These signals do not follow a single path; instead, they undergo multipath propagation, reflecting off various surfaces, objects, and obstacles. This phenomenon causes multiple delayed copies of the transmitted signal to arrive at the receiver. The composite effect of these signal copies creates a unique fingerprint of the propagation environment. It is the alteration of this fingerprint due to human movement that provides the fundamental information utilized for device-free HAR with WiFi.

2.2.2. Line-of-Sight (LoS) vs. Non-Line-of-Sight (NLoS)

Line-of-sight (LoS) propagation occurs when WiFi signals travel on a direct, unimpeded path from the transmitter to the receiver. In contrast, non-line-of-sight (NLoS) propagation occurs when signals reach the receiver indirectly after being reflected, diffracted, or scattered by objects, humans, walls, or furniture. The difference between these propagation conditions is illustrated in Figure 1 [73].

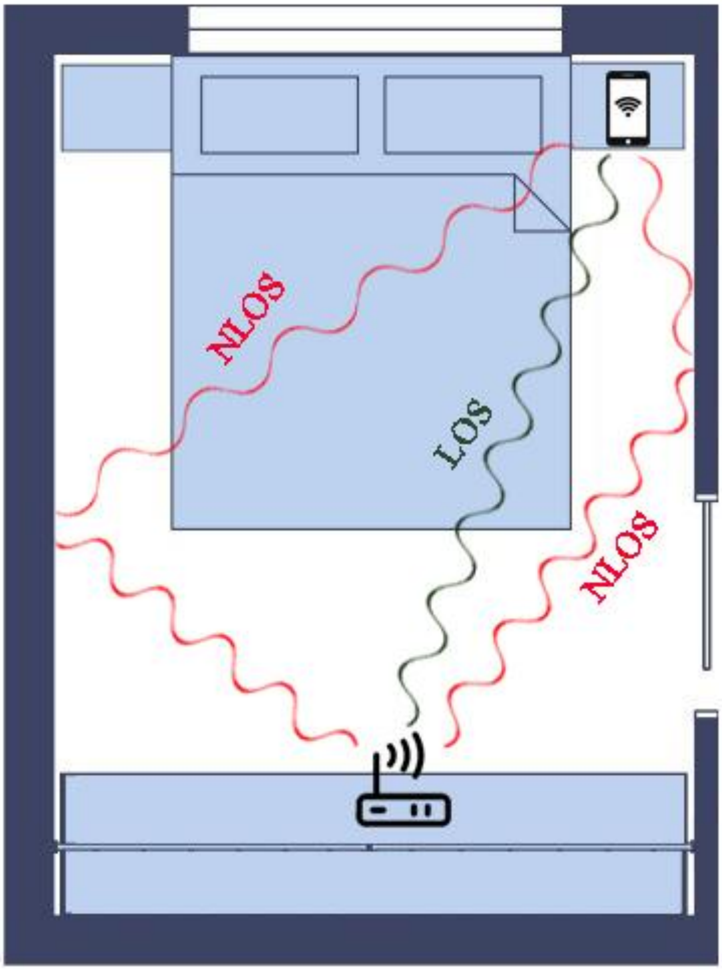

Figure 1 LoS and NLoS in WiFi signal propagation

In device-free HAR systems, Line-of-Sight (LoS) is typically crucial. Vision-based methods cannot recognize human activities behind a wall, whereas device-free methods leverage signal waves (light, sound, electromagnetic) that penetrate objects, allowing for the recognition of human activities in NLoS scenarios. A recent literature review [42] highlights the applications of LoS and NLoS in WiFi-based systems. While NLoS scenarios offer useful applications, [26] reviewed several studies and reported that the accuracy of methods in NLoS scenarios is always lower compared to LoS scenarios. In SLR, 19% of the papers we reviewed considered NLoS for HAR (46 papers) while 81% did not deal with NLoS(196 papers).

2.2.3. Orthogonal Frequency Division Multiplexing (OFDM)

WiFi is among the most widely used wireless communication technologies. It is defined by the IEEE 802.11 family of standards, which have evolved from 802.11b to include versions such as 802.11a/g/n/ac/ax. A foundational feature introduced in 802.11a/g is the Orthogonal Frequency Division

Multiplexing (OFDM) modulation scheme. OFDM divides the available bandwidth into multiple orthogonal, narrowband subcarriers, transmitting data in parallel across them to improve robustness against interference and multipath fading.

The OFDM modulation scheme can be conceptually understood by examining the transmission of a single data bit. Each bit can be represented by a rectangular function in the time domain (see Figure 2.b). This representation is multiplied by a sinusoidal carrier signal (Figure 2.a) to produce the final transmitted signal (Figure 2.c). The frequency-domain characteristics of these signals, obtained via Fourier transform, are shown in Figure 2.d, Figure 2.e, and Figure 2.f respectively.

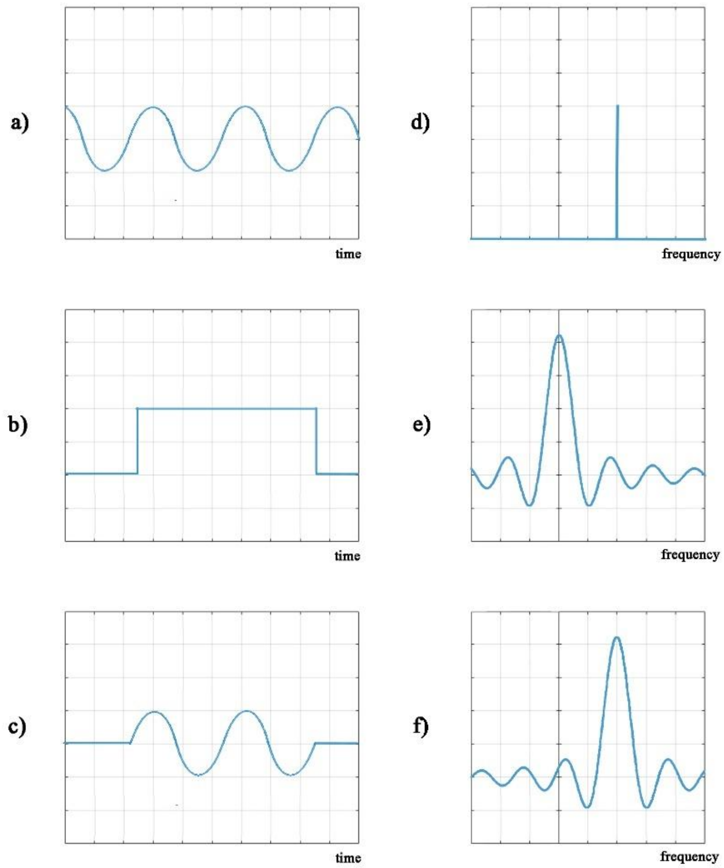

Figure 2 OFDM modulation fundamentals in 802.11 IEEE standard for WiFi communications. a) carrier signal in the time domain, b) one bit of data represented by a rectangular function in the time domain, c) multiplication of two signals represented in a and b, d) Fourier transform of signal a, e) Fourier transform of signal b, and f) convolution of two signals represented in d and e

The theoretical use of a rectangular time-domain window in OFDM creates sharp symbol transitions, leading to significant out-of-band spectral leakage (evident in the sinc function sidelobes in Figure 2.e). To mitigate this, the 802.11 standard mandates a cyclic prefix (CP), which copies the symbol's end to its beginning. This guard interval absorbs multipath delay, preventing inter-symbol interference (ISI) and enabling simpler frequency-domain equalization. For enhanced spectral containment, additional pulse shaping (e.g., with a raised cosine window) can be applied to taper the symbol's edges, further reducing leakage at the cost of a marginal increase in symbol duration. Consequently, OFDM implementations robustly manage these inherent discontinuities to achieve spectral efficiency.

OFDM enables high data rates by dividing the carrier frequency into multiple orthogonal subcarriers, each transmitting a portion of the data simultaneously. The data on each subcarrier is modulated independently. The signals from all subcarriers are combined in the time domain for transmission. At the receiver, a Fourier transform is applied to demultiplex the composite signal and recover the data transmitted on each individual subcarrier. Critically, the subcarriers are orthogonal, meaning their peak frequencies are chosen so that they do not interfere with each other despite spectral overlap, as illustrated in Figure 3. It is this division of the channel into fine-grained subcarriers that enables the detailed channel measurements known as Channel State Information (CSI).

The 802.11n standard in the 2.4 GHz band specifies 52 orthogonal subcarriers with a spacing of 312.5 KHz. These subcarriers are indexed from -26 to 26. According to the standard, four of these subcarriers (-21, -7, 7, and 21) are designated as pilot subcarriers. Unlike data subcarriers, pilots are modulated using Binary Phase Shift Keying (BPSK) with a known magnitude and phase, which aids in tracking and compensating for channel variations.

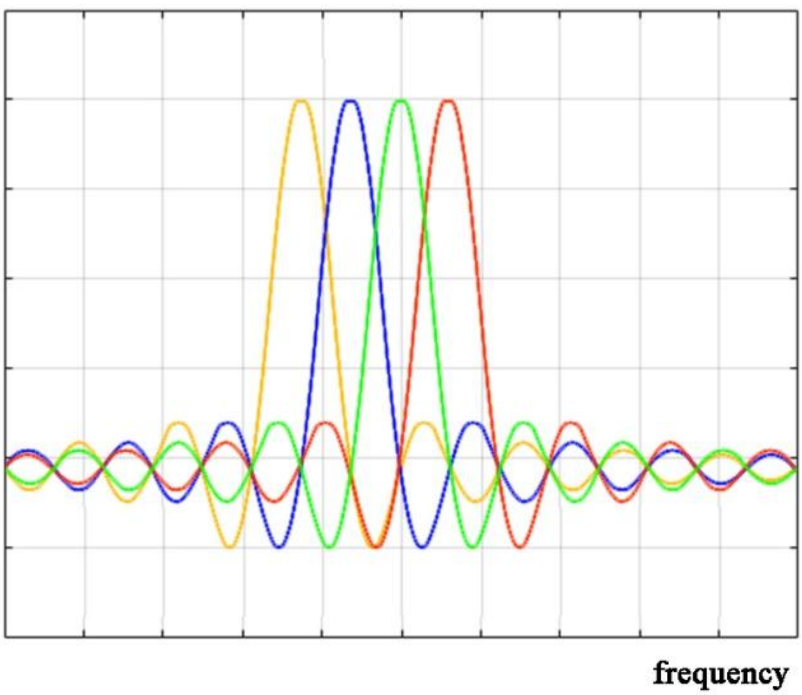

Figure 3 Fourier transform of the received signal in WiFi receiver supports OFDM modulation

Channel State Information (CSI) is derived from the frequency-domain representations of the transmitted and received signals across these subcarriers. It characterizes the propagation environment for each subcarrier. Mathematically, the channel frequency response H(f, t) at frequency f and time t is modeled by the following formula [26]:

$$H(f,t) = \sum_{n=1}^{N} a_n(t) e^{-j2\pi f \tau_n(t)}$$

Where the variables are defined in Table 2. Crucially, all variables in this model except the carrier frequency f are sensitive to the environment. Movements of objects or humans alter the attenuation $a_{n(t)}$ and delay $\tau_{n(t)}$ on various paths N. It is this sensitivity that makes CSI a powerful tool for device-free HAR. Consequently, any environmental layout characteristic that affects these variables must be considered during WiFi sensor placement [43].

Table 2 Definition of variables in the CSI formula

Variable	Definition
$a_i(t)$	Amplitude Attenuation Factor
$\tau_i(t)$	Propagation Delay
f	Carrier Frequency
N	Is the number of Paths

2.3 Signal Measurements in Wi-Fi

Several types of signal measurements are commonly used in WiFi technology to characterize wireless communication links. These include the Received Signal Strength Indicator (RSSI), which measures the total power of the received signal; Channel State Information (CSI), which captures detailed amplitude and phase information for each OFDM subcarrier; the Signal-to-Noise Ratio (SNR), which reflects signal quality relative to background noise; and the Packet Delivery Ratio (PDR), which indicates the ratio of successfully received packets to the total number sent. For device-free human activity recognition (HAR), CSI and RSSI are the most directly relevant metrics, as they are most sensitive to environmental changes caused by human movement.

2.3.1. Channel State Information (CSI)

Channel State Information (CSI) is a measurement that characterizes the frequency response of a wireless channel across its subcarriers, representing how signals propagate through multiple paths in the environment. Fundamentally, CSI is derived from the Channel Frequency Response (CFR), which is calculated for each subcarrier using the formula:

$$H(f,t) = \frac{Y(f,t)}{X(f,t)}$$

where H(f, t) is the complex-valued CFR at frequency f and time t, and X(f, t) and Y(f, t) are the frequency-domain representations of the transmitted and received orthogonal frequency-division multiplexing (OFDM) symbols, respectively [44]. CSI can thus be considered a sampled version of the CFR.

The number of subcarriers contained in CSI data is hardware-dependent and varies with channel bandwidth. For example, a 20 MHz channel may have 30 or 52 subcarriers (depending on the specific tool and standard), while an 80 MHz channel can have up to 256 [45]. Most commercial Wi-Fi devices utilizing the 802.11n standard report CSI for 30 subcarriers. Consequently, the size of the gathered CSI data is given by $N_T \times N_R \times 30 \times P$ where N_T and N_R are the number of transmitter and receiver antennas, respectively, and P is the number of packets.

The Multiple-Input Multiple-Output (MIMO) system model, which underpins the use of CSI from devices with multiple antennas, is defined by the following equation [46]:

$$y = Hx + N$$

where x is the transmitted signal vector, y is the received signal vector, H is the channel matrix containing the CSI, and N is additive white Gaussian noise. For a system with N_T transmit antennas and N_R receive antennas, the CSI matrix H has dimensions $N_R \times N_T \times N_{SC}$, where N_{SC} is the number of subcarriers. Each element in this matrix is a complex value representing the CFR for a specific transmit-receive antenna pair and subcarrier. The structure of this CSI data matrix is illustrated in Figure 4 [47].

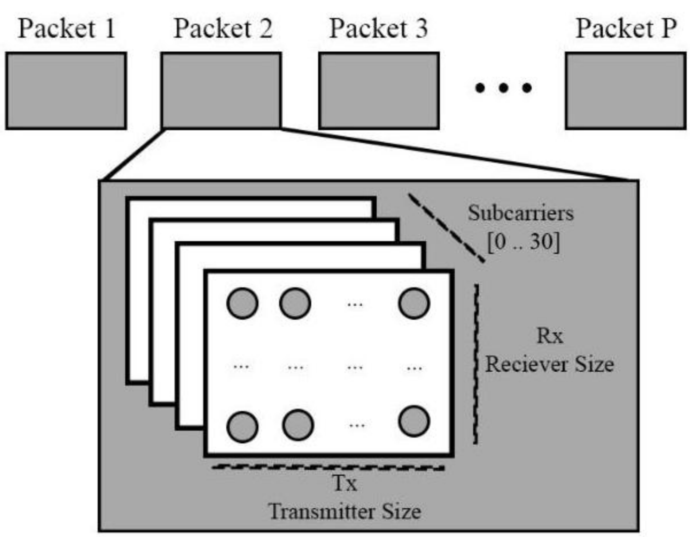

Figure 4 Structure of CSI data

2.3.2. Received Signal Strength Indicator (RSSI)

The Received Signal Strength Indicator (RSSI) is a measurement that quantifies the power level of a received radio signal from a transmitter. It is typically represented in decibels (dB) and provides an aggregate measure of the signal strength at the receiver. Changes in the propagation path between the transmitter and receiver—such as those caused by obstruction, reflection, or absorption—result in fluctuations in the RSSI value. While less fine-grained than CSI, analyzing these temporal changes in RSSI enables coarse-grained localization and activity recognition.

2.4 Machine Learning Methods

Machine learning (ML) forms the core computational engine of device-free human activity recognition (HAR) systems. The fundamental purpose of these systems is to learn the complex mapping between patterns in the extracted signal features (e.g., from CSI or RSSI) and the specific human activities that generated them. Therefore, all processed signal characteristics are presented to a machine learning model, which learns to identify which features and patterns are discriminative indicators of each activity.

2.4.1. Long Short-Term Memory (LSTM)

In recent years, deep learning methods—machine learning approaches based on deep neural networks—have outperformed classical shallow models in numerous applications [48]. A primary reason for this superiority is their ability to not only learn from engineered features but also to automatically extract relevant hierarchical features directly from raw or preprocessed data. These discriminative patterns, often imperceptible to human analysis, enable deep learning models to identify highly distinctive characteristics for complex tasks.

Long Short-Term Memory (LSTM) networks and their variants are among the most popular deep learning models for sequential data. As an extension of recurrent neural networks (RNNs), LSTM was specifically designed to overcome the vanishing gradient problem [49]. This architectural innovation enables LSTMs to effectively capture both short-term and long-term temporal dependencies, making them highly suitable for time-series analysis like HAR. An LSTM network is composed of interconnected memory cells, each of which learns to update its internal state based on both current inputs and information from previous time steps.

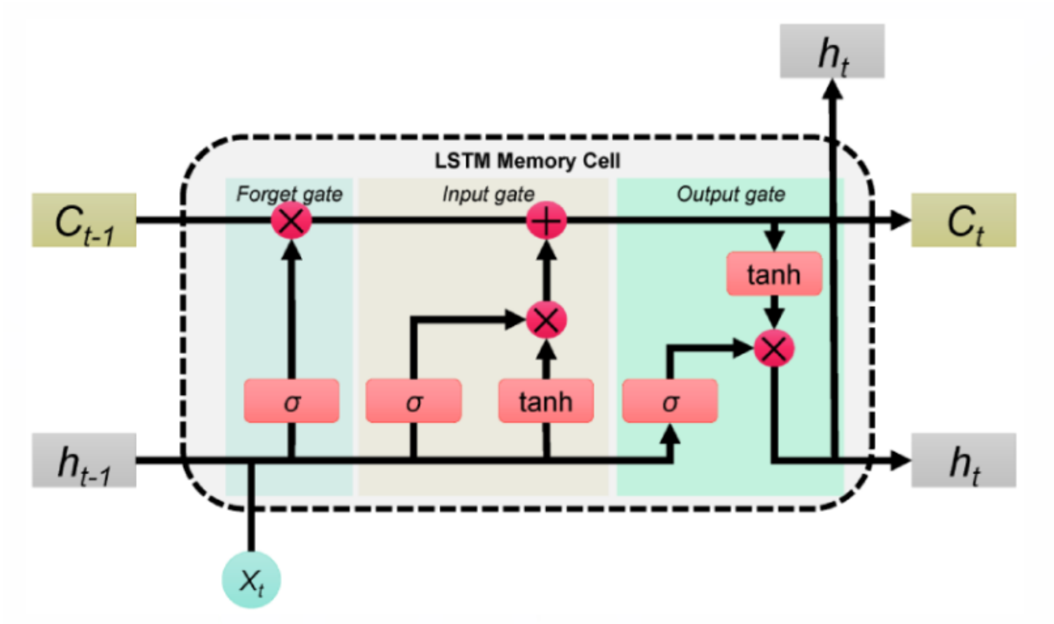

Figure 5 The LSTM cell structure (adapted from [50])

Figure 5, adapted from [50] , illustrates the internal structure of an LSTM memory cell. Each cell contains four key gates that regulate information flow:

- Forget gate: Determines which information from the previous cell state should be discarded or retained.
- Input gate: Controls how much of the new input should be added to the cell state.
- Cell state update: Combines the previous cell state (after forgetting) with the new candidate information from the input gate.
- Output gate: Decides what information from the current cell state should be output to the next hidden state.

These gates work together to maintain and update the cell state, allowing the LSTM to preserve long-term dependencies and selectively forget irrelevant information.

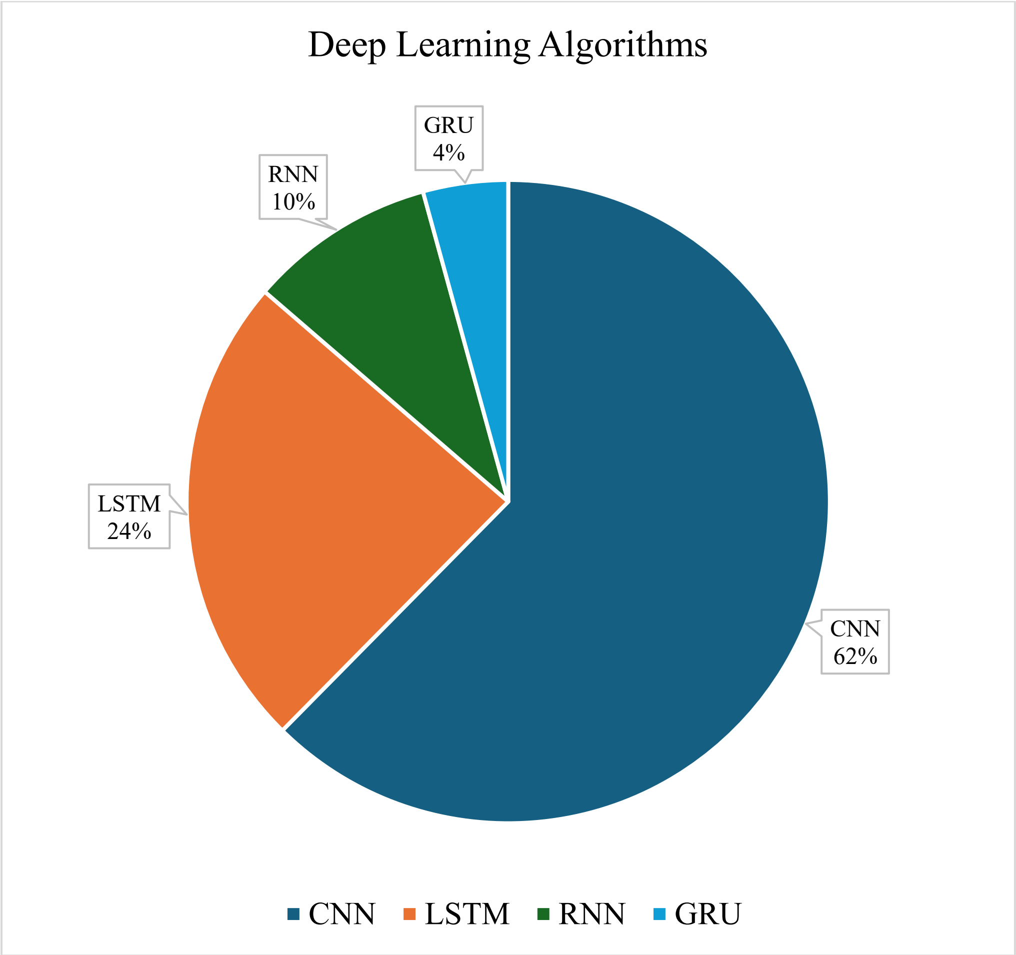

Figure 6 Number of papers that used different deep learning algorithms for device-free HAR

The results of the systematic literature review conducted for this work [32] (see Figure 6) show that a significant majority (68%) of deep learning-based device-free HAR methods utilize convolutional neural networks (CNNs), while a substantial portion (19%) employ LSTM networks. For instance, Zhang et al. [51] proposed one of the most cited methods in recent years. In their work, the authors recognized ten human activities (e.g., running, sitting, standing, lying down, bending, jumping, waving, clapping, and checking a wristwatch) with approximately 90% accuracy using one transmitter and two receivers. They proposed a WiFi-based HAR system that employs data augmentation and a Dense-LSTM model to address challenges such as inconsistent motion speeds, subject-specific variations, and small dataset sizes. Their method included synthesizing training data through eight CSI transformation techniques (e.g., dropout, time stretching, spectrum shifting) and designing a compact neural network to prevent overfitting, achieving robust performance even with limited samples.

Despite their popularity, deep learning algorithms for device-free HAR possess several drawbacks that can limit their applicability. A primary issue is their heavy dependence on large, labeled datasets. Device-free HAR is an evolving field where few large, public datasets exist, and those that do often focus on specific activities. This presents a major limitation for recognizing activities not represented in existing data. A second drawback is high computational cost. While perhaps acceptable for offline analysis, the computational demands of deep learning models can be prohibitive for edge computing and online HAR requiring real-time inference on devices with limited processing power. Finally, the inherent black-box nature of deep learning poses a significant challenge. As an emerging field, device-free HAR requires interpretability to understand the impact of various parameters and features (refer to Chapter 7), a level of transparency that deep learning models often lack.

2.4.2. Transfer learning

As delineated in the previous section, a primary challenge in device-free HAR is the scarcity of large, labeled datasets required to effectively train deep learning models from scratch. Transfer learning [52] presents a powerful strategy to mitigate this issue. This approach involves leveraging a model pre-trained on a large, related source dataset and adapting it to perform classification on a smaller target dataset. Within device-free HAR, transfer learning enables systems to utilize knowledge gained from existing activity datasets (e.g., collected in one environment) to recognize new activities or to operate effectively in a different, but related, context.

Deep learning models comprise a hierarchical structure of layers, where lower and middle layers often learn to extract general features (e.g., edges, textures, temporal patterns), while higher layers learn task-specific features. Transfer learning capitalizes on this structure by initializing a new model with the pre-trained parameters from the lower and middle layers of a source model, effectively using it as a feature extractor. Only the final layers (the classifier head) are typically replaced and trained from scratch. Subsequently, the entire network is fine-tuned on the limited target dataset. This process allows the model to adapt the general features to the specific target task, enabling accurate activity recognition even with few training samples.

A prominent example is TransferSense, a framework proposed by Bu et al.[53] for device-free HAR using CSI data. This method utilizes a deep learning model trained with a transfer learning objective. The core contribution of their work is to demonstrate that a model can be trained in one context (e.g., a specific environment or set of users) and deployed effectively in another, addressing the critical challenge of cross-domain generalization in device-free HAR.

2.4.3. Classical Learning

Classical (or traditional) machine learning models encompass a family of statistical algorithms that learn to map input features to output labels or values based on patterns in training data. Unlike deep learning, these models typically rely on carefully engineered features as input for tasks such as classification and regression. A wide variety of classical models have been applied in device-free HAR; this section details the most prevalent and impactful among them.

The results of the systematic literature review [32] (see Figure 7) indicate that a significant proportion of the reviewed papers (50%) identified the Support Vector Machine (SVM) as the best-performing model, followed by the K-Nearest Neighbors (KNN) algorithm (25%). Other frequently employed and high-performing models include Random Forest, Decision Tree, Extreme Learning Machine (ELM) [54], and Logistic Regression [55], [56] are the next frequent models.

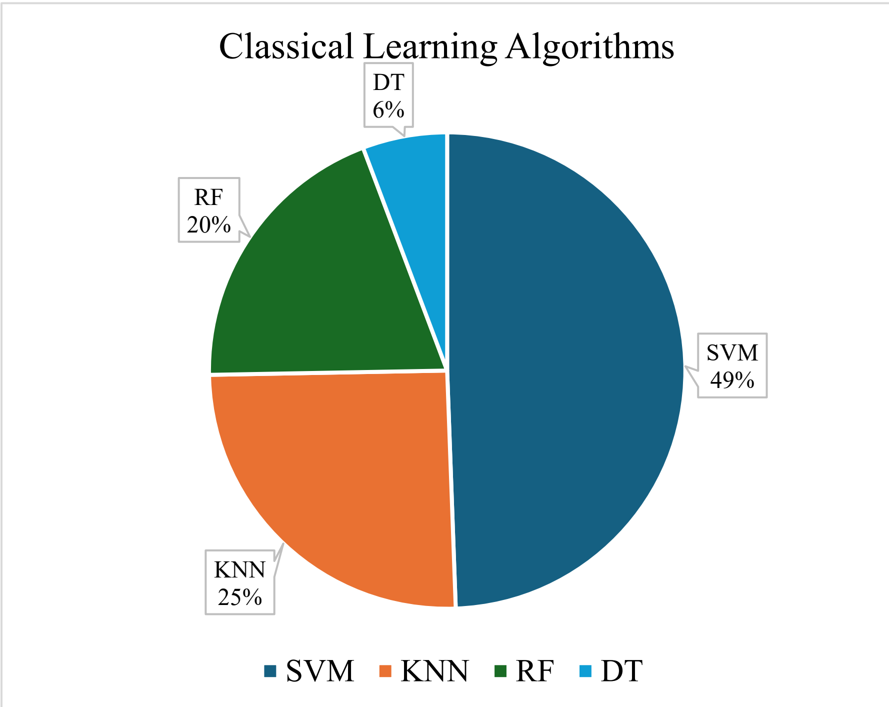

Figure 7 Number of papers that used different classical learning algorithms for device-free HAR

Although the performance of deep learning methods in device-free HAR is often superior, numerous studies demonstrate that classical learning methods can achieve highly competitive results. For instance, Atif et al. [57] compared an Artificial Neural Network (ANN) and an SVM model for HAR in a privacy-preserving edge computing environment. Their system achieved 99% accuracy with the ANN and 97.5% with the SVM in recognizing five activities: no activity, sitting, standing, walking, and picking up an object. These results demonstrate that well-optimized classical models can closely approximate the performance of neural networks. Further reinforcing this point, Natarajan et al. [58] demonstrated that traditional machine learning models can rival deep learning even in challenging through-the-wall scenarios. Using single-link Wi-Fi CSI data from low-cost ESP32 devices, their proposed time-domain and frequency-domain feature set achieved 97–99% accuracy in classifying five activities (empty room, walking, standing, sitting, lying) across diverse NLoS conditions. Notably, Linear Discriminant Analysis (LDA) outperformed deep learning models, including LSTMs and Transformers, in NLoS environments, reaching 99.6% accuracy for joint activity-location classification. This work underscores that well-designed feature extraction, paired with lightweight classical models (e.g., LDA, Random Forests), can match deep learning performance while significantly reducing computational costs—a critical advantage for edge deployment in applications like smart buildings or healthcare monitoring.

Support Vector Machine (SVM)

The Support Vector Machine (SVM) is a powerful supervised learning algorithm primarily designed for binary classification, often employed in HAR using a one-versus-rest strategy for multi-class problems. Its application in device-free HAR is well-documented:

- Hao et al. [59] boosted an SVM classifier by integrating it with a K-means algorithm, achieving 89.3% accuracy in recognizing sign language gestures.
- Zhao et al. [60] utilized an SVM with a Gaussian kernel to create user profiles, classifying atomic actions (walking, sitting down, pushing, swinging, waving) with 97% accuracy and noting its negligible impact on total recognition time due to fast response.
- Huang and Dai [61] employed SVM with sequential minimal optimization (SMO) [62] and 10-fold cross-validation to recognize actions (standing, walking, running) and differentiate motion speeds, achieving 76.7% accuracy.
- Yan et al. [63] applied an SVM to achieve 98% accuracy in classifying fall directions (falling forward, backward, left, right) and user positioning.

K-Nearest Neighbor (KNN)

The K-Nearest Neighbor (KNN) algorithm is a non-parametric, instance-based learning method that classifies a data point based on the majority class among its K closest neighbors in the feature space [64]. Its simplicity and effectiveness make it a common baseline in HAR:

- Yang et al. [65] achieved 90% accuracy in classifying actions (jumping, running, sitting down, walking, boxing, golf swinging) using a customized distance metric based on reconstruction error in activity-specific feature subspaces.

- Sheng et al. [66] employed a CNN for feature extraction and a KNN for classification, achieving 97% accuracy on a set of gestures. Their work highlighted the importance of tuning K, finding that overly large values (e.g., K=20) can introduce bias and degrade performance.
- Chen et al. [67] combined KNN with Dynamic Time Warping (DTW) distance to account for temporal variations in hand gestures, achieving 95% accuracy.

Random Forest (RF)

Random Forest (RF) is an ensemble learning method that constructs a multitude of decision trees during training and outputs the mode of their classes for classification [68]. By training each tree on a random subset of data and features, it effectively reduces overfitting. Wang and Zheng [69] conducted a comparative study of classifiers (RF, Neural Network, Decision Tree, SVM, Naïve Bayes) for recognizing atomic actions (e.g., raising/dropping a hand, walking, sitting, standing, falling). Their Random Forest model achieved the highest accuracy (93.5%), which the authors attributed to its inherent robustness against overfitting, a particularly valuable trait when working with limited training data.

Naive Bayes (NB)

Naive Bayes classifiers are based on applying Bayes' theorem with a strong (naive) assumption of conditional independence between every pair of features given the class label. While this assumption rarely holds completely in real-world data, NB classifiers are often surprisingly effective and are known to be robust to certain violations of this assumption. Sharma et al. [70] evaluated several classifiers (Gaussian NB, Multi-Layer Perceptron, KNN, Random Forest) for recognizing basic activities (standing, sitting, lying) using RF signals. Contrary to what might be expected, the Gaussian NB model achieved the highest accuracy (95%), demonstrating its potential utility in certain HAR tasks.

Decision Tree

A Decision Tree is a predictive modeling tool that recursively partitions the feature space into regions, making a sequence of hierarchical, logical decisions to arrive at a classification [71]. Zeng et al. [72] proposed WiWho, a device-free person identification framework. Their system used a decision tree classifier based on a standard set of statistical features (e.g., min, max, mean, variance, energy, entropy) to recognize basic activities (walking, sitting, standing), achieving 95% accuracy.

2.5 Statistical and Optimization Methods

In addition to machine learning methods, traditional statistical and optimization techniques play a critical role in various aspects of device-free HAR system design. Unlike machine learning, which is often model-dependent, statistical methods provide a means to analyze the inherent properties of the data itself, independent of any subsequent classification algorithm. To illustrate, one could evaluate a feature's utility for discriminating between two activities by training a Random Forest model on a subset of the data and testing its accuracy. However, this outcome confounds the quality of the feature with the efficacy of the model; a poor result does not preclude the feature being discriminative with a different classifier. In contrast, a statistical test —such as the Kruskal-Wallis test [73]— directly compares the distributions of the feature across different activities, providing an unbiased, model-agnostic assessment of its discriminatory power. This section details the specific statistical and optimization methods leveraged in this work.

2.5.1. MANOVA

Analysis of Variance (ANOVA) is a statistical method used to compare the means of a continuous variable across two or more independent groups, determining whether any of those means are statistically significantly different from each other. It operates by partitioning the total observed variance into components: variance between groups and variance within groups. A significant F-statistic indicates that the between-group variance is larger than would be expected by chance alone. The two primary versions are one-way ANOVA (one independent variable) and two-way ANOVA (two independent variables).

In HAR, one-way ANOVA can be employed to evaluate the discriminatory power of individual features. For example, to determine which of two features better distinguishes between two activities, an independent ANOVA test is conducted for each feature. A statistically significant result (typically indicated by a low p-value or a high F-statistic) suggests that the feature's distribution differs meaningfully across the activities, marking it as discriminative. Even if both features are significant, their results are comparable; the feature with the more extreme test statistic is considered more strongly discriminative.

Two-way ANOVA is utilized when a dependent variable is influenced by two independent factors. For example, in a HAR study, one might investigate the effect of both activity type (e.g., walking, sitting) and environment (e.g., room A, room B) on a specific feature. Two-way ANOVA assesses three effects: the main effect of activity, the main effect of environment, and the interaction effect between activity and environment. This analysis determines whether the feature varies significantly across activities, across environments, and whether the effect of an activity on the feature depends on the environment (or vice versa). Such analysis is crucial for identifying features that are robust across environmental changes or are specifically sensitive to certain activities, guiding the selection of reliable features for robust device-free HAR systems.

Multivariate ANOVA (MANOVA) generalizes ANOVA to scenarios with multiple, correlated dependent variables. In device-free HAR, where numerous features are extracted from sensor data, MANOVA assesses whether the entire set of features differs significantly across groups (e.g., activity types). Its key advantage over conducting multiple separate ANOVAs is that it accounts for the correlations between features, providing a more holistic and statistically rigorous evaluation. Two-way MANOVA further extends this to two independent factors and their interaction. For instance, it can examine how activity type and environment jointly influence a multivariate set of features. This allows for the identification of feature sets that are collectively discriminative for activities while being invariant to environmental changes, thereby directly contributing to the development of more robust and accurate HAR models.

2.5.2. Genetic Algorithm

Genetic Algorithm (GA) is an optimization technique inspired by the principles of natural selection and genetics. It iteratively evolves a population of candidate solutions toward an optimum by applying operators such as selection, crossover, and mutation. In the context of device-free HAR, GAs are primarily used for feature selection. The algorithm begins with an initial population where each individual represents a candidate subset of features. The fitness of each individual—a measure of solution quality—is evaluated using a predefined fitness function. A key contribution of this work is the use of a statistical measure derived from MANOVA (e.g., p-value, Pillai's trace) as this fitness function, assessing the combined discriminatory power of the feature subset. Individuals with higher fitness (e.g., lower MANOVA p-value) have a greater probability of being selected to reproduce, passing their "genes" (features) to the next generation. Over successive generations, this process evolves increasingly fit solutions, ultimately identifying an optimal or near-optimal subset of features for activity classification.

Chapter 3 Related Works

3.1 Human Activity Recognition

HAR gained significant academic and commercial attention following the publication of seminal papers on activity recognition using wearable accelerometers [74] in 2004, cell phone accelerometers [75] in 2005, and a key survey of vision-based HAR [76] in 2008. However, even in these early years, researchers noted critical limitations: privacy concerns associated with vision-based systems [77], and the discomfort and lack of social acceptability [78] of wearable and object-tagged sensors [79]. At the time, a viable solution to overcome both of these barriers simultaneously was not available.

The exploration of environmental sensors utilizing radio frequency signals led to the introduction of device-free HAR after 2010 as a direct response to the drawbacks of earlier modalities. This emerging approach was exemplified by early work such as Sigg et al. [80], which demonstrated the ability to distinguish activities like walking, lying, crawling, and standing with high accuracy using software-defined radio signals.

Table 3 Distribution of reviewed papers based on utilized sensors in device-free methods.

Category	Subcategory	# of Papers	Percentage
Radiation	WiFi	160	~ 66%
	Radar	34	~ 14%
	RFID	15	~ 6%
Acoustic	Acoustic	14	~ 6%
	Ultrasonic	2	~ 1%
Optical	Visible Spectrum	4	~ 2%
	Infrared	3	~ 1%
Hybrid	Hybrid	5	~ 2%
Other	Other	5	~ 2%

As detailed in Section 2.1.1, the systematic literature review [32] conducted for this work categorized device-free HAR methods based on sensing modality. The distribution of 242 reviewed papers (from 2017 to 2023) across these categories is presented in Table 3 and illustrated in Figure 8. The data reveals a clear dominance of radiation-based sensors, which were employed in 199 papers (~82% of the total). Within this category, WiFi-based sensing is by far the most prevalent, accounting for 160 papers (~66% of the total). In contrast, acoustic sensors were used in only 14 studies (~6%), and optical sensors in merely 7 studies (~3%). This quantitative analysis underscores the research community's strong focus on RF-based methods, and particularly WiFi, a trend that informs the focus of the research presented in this dissertation.

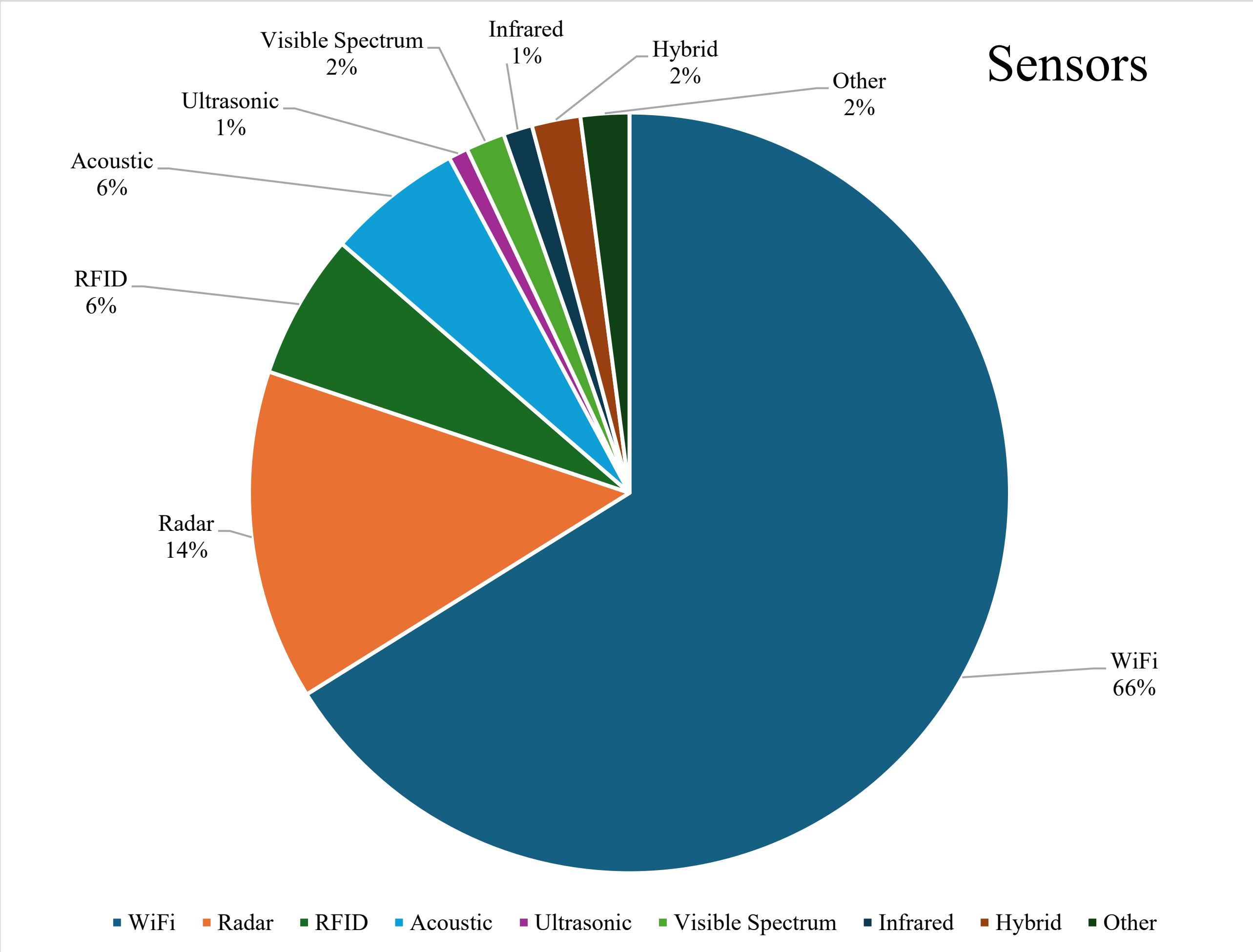

Figure 8 Visualization of the distribution of studies based on utilized sensors in proposed device-free methods

3.2 Device-free Human Activity Recognition

Device-free human activity recognition (HAR) leverages ambient signals to detect and classify human activities without requiring the user to carry or wear any sensor. In these systems, sensors are embedded in the environment (e.g., WiFi transmitters, radars, acoustic sensors) and emit signals such as radio waves, sound waves, or light waves. The systems recognize activities by analyzing how the reflections of these transmitted signals are altered by human motion. The definition of "device-free" has evolved. Initially, it described any system where the user was unencumbered, a broad definition that included vision-based methods as long as the camera was environmental [81], [82], [83]. However, as the field matured, a more precise taxonomy has been widely adopted, categorizing HAR into three distinct modalities: device-based (wearables, object-tagged), vision-based, and device-free (using non-visual ambient signals), thereby explicitly separating camera-based approaches from device-free methods [29].

A seminal early work in RF-based device-free sensing was introduced by Kosba et al. [84] , who proposed a WLAN device-free passive (DfP) indoor localization system. Although focused on localization rather than activity recognition, this work demonstrated the feasibility of using standard WiFi hardware for detecting human presence and motion without carried devices, thereby paving the way for future device-free HAR research. Building on this foundation, the work presented in this dissertation focuses specifically on methods that utilize radio frequency signals, with a primary emphasis on WiFi for recognizing human activities.

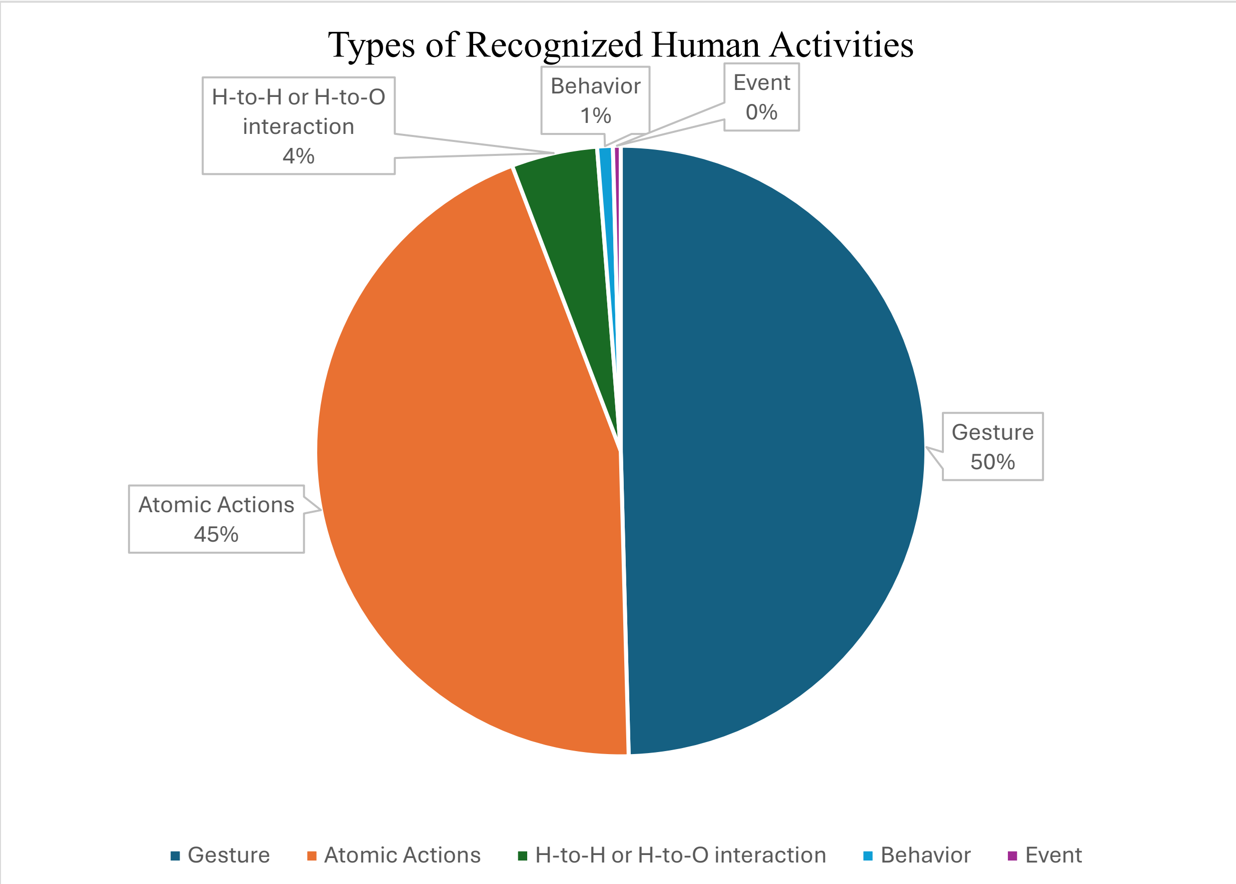

Figure 9 Visualization of the distribution of studies based on the type of recognized human activity

The systematic literature review [32] categorized the 242 studied papers based on the type of activities recognized, adopting the taxonomy proposed by Vrigkas et al. [85]. This taxonomy defines: (1) gestures (primitive body part movements, e.g., hand signs); (2) atomic actions (whole-body motions describing an action, e.g., walking, sitting); (3) human-to-object (H-to-O) or human-to-human (H-to-H) interactions; (4) group actions; (5) behaviors (actions presenting an emotion); and (6) events (high-level social actions).

As shown in Figure 9, the distribution of research is highly concentrated: 50% of studies focused on recognizing gestures (120 papers) and 45% on atomic actions (108 papers). This reveals that the vast majority of device-free HAR research has addressed a relatively narrow range of simple, context-agnostic activities (e.g., standing, sitting, walking, falling, or gestures like waving and clapping). Consequently, there is a notable scarcity of studies focused on recognizing more complex activities within specific, real-world contexts.

This gap is highlighted by the few studies that have targeted specific contexts. For instance, Zhou et al. [86] focused on office activities, recognizing ten predefined actions including whole-body movements (e.g., walking, jumping) and seated partial-body activities (e.g., typing, drinking, stretching), achieving over 97% accuracy. Similarly, Wang et al. [87] targeted in-vehicle context, utilizing RFID to recognize driver gestures (e.g., shoulder checks, answering a phone, pushing buttons) with 95% accuracy. These works demonstrate the potential for device-free HAR in specialized domains but also underscore that such focused studies are currently the exception rather than the norm.

3.2.1. Wi-Fi-Based HAR

In the nascent stages of WiFi-based sensing, the Received Signal Strength Indicator (RSSI) was the primary metric used for localization and rudimentary activity recognition, largely due to its widespread availability on commercial-off-the-shelf (COTS) devices [88]. However, the low granularity and high susceptibility of RSSI to environmental noise and multipath effects fundamentally limited its utility for fine-grained sensing tasks. A paradigm shift occurred circa 2013-2014 with the broader adoption of Channel State Information (CSI) tools, which provided unprecedented access to the fine-grained physical layer information of WiFi signals, heralding a new era for device-free HAR [89], [90]. Early pioneering work quickly demonstrated the potential of CSI for fine-grained recognition, such as identifying words from mouth motion profiles with 91% accuracy [91]. The field's evolution accelerated with the integration of deep learning; the first studies applying Convolutional Neural Networks (CNNs) to CSI data emerged around 2016, targeting activities like walking and falling [92]. This was swiftly followed by demonstrations of sophisticated applications like contactless sleep monitoring [93] and elderly fall detection [94] using commodity WiFi hardware.

In recent years, WiFi-based HAR has matured significantly [34], driven by three key developments: (1) advances in WiFi sensing technology, including higher-resolution CSI from WiFi 5/6 and mmWave [95]; (2) the adoption of advanced deep learning architectures like Transformers ; and (3) a expansion into diverse real-world applications including fall detection [96], sleep monitoring [97], and smart home control [98]. Concurrently, the field has begun to address critical challenges such as privacy preservation through techniques like federated learning [99] and system integrity via adversarial robustness research [100].

3.3 Context-Specific Human Activity Recognition

3.3.1. Importance of Context in HAR

The concept of context is fundamental to human activity recognition. Drawing from early definitions in mobile computing, context in HAR can be understood as "the location and identities of nearby people and objects, and changes to those objects" [101]. Human activities are rarely performed in isolation; they typically involve complex interactions between a user and their surrounding environment, people, and objects. Even seemingly context-independent actions, such as yawning, are influenced by situational factors, and this context is often inadvertently captured in the data collection process. This is especially true for video-based and device-free HAR methods, where the sensor data inherently encapsulates a rich signature of the environment. Consequently, the significance of context has been a subject of investigation across all HAR modalities, including device-based [102], video-based [103], and, more recently, device-free approaches [104].

3.3.2. Kitchen Activity Recognition

While traditional HAR research has largely focused on full-body activities like walking or sitting in controlled environments, less attention has been given to recognizing complex, context-specific tasks in real-world domestic settings. One such environment is the kitchen [105], an interaction-dense space central to daily routines, nutrition, and cognitive functioning.

Kitchen activities involve frequent transitions, object interactions, and fine-grained hand movements within specific spaces [106]. Unlike full-body activities, these tasks often exhibit overlapping motion patterns, subtle gestures [107], and strong dependence on contextual factors such as layout, appliance placement [73], and user behavior. Accurate recognition of such activities has potential applications in dietary monitoring [108], elderly care [109], and cognitive health assessment [110].

Pham et al. [111] developed decision tree, Bayesian net, and naïve Bayes classifiers to recognize 11 food preparation actions: chopping, peeling, slicing, dicing, coring, spreading, eating, stirring, scooping, scraping, and shaving. They put 3-axis accelerometer sensors in the utensils and asked 20 users to prepare a mixed salad. They achieved 82.9% of accuracy in recognition of the mentioned activities. Mohammad et al. [112] provided a dataset containing smart-watch data for 10 participants performing kitchen activities: mix with chopsticks, move aside, cut on board, put ingredients, peel off the onion, peel vinyl, peel the cabbage, put food ingredients, wash cookware, pick up ingredients, serve on a dish, shake, cut the ingredients, wipe the hands, and put a large object. Despite a relatively small amount of data, Convolutional Neural Networks (CNN) achieved higher accuracy compared to traditional classifiers that use statistical features. Monteiro et al. [113] proposed a vision-based HAR method using a deep neural network (DNN) to recognize nine kitchen actions, breaking, mixing, baking, turning, cutting, boiling, seasoning, peeling, and none, from videos of four participants performing five recipes, achieving an average accuracy of 90%. These kitchen activities were originally defined by Bansal et al. [114] in 2013 which used a hybrid model using a support vector machine and a hidden Markov model and achieved 72% accuracy using the same dataset. In another study, Zolfaghari et al. [115] conducted a benchmark study on recognizing six kitchen actions: pour, put, scrape, stir, take, and wait, using embedded sensors in 37 kitchen objects, along with wearable sensors and video cameras. Using a hybrid AI model, they achieved an F1-score of 81% for activity recognition. Liu et al. [116] recently designed a real-time kitchen activity recognition system using wearable and environment devices like, accelerometer, gas sensors, distance sensors, barometers, etc. They achieved 87% accuracy in recognition of 15 different activities in kitchen context including opening and closing the refrigerator, opening and closing the microwave, pouring and boiling water, stirring, drinking, using the toaster, opening and closing a cabinet, cleaning, taking food items, and null class (standing idle).

The studies summarized above rely on object-tagged, wearable, or vision-based methods. While device-free approaches have been applied to HAR since approximately 2018, their application to the kitchen context has been exceptionally limited and peripheral, often recognizing only one or two kitchen activities incidentally within a broader study, as comprehensively detailed in Table 4. This table categorizes recognized kitchen activities into four groups: eating, food preparation, cleaning, and other activities. Analysis of the table reveals two key insights: first, that activities related to food preparation represent the most diverse and complex category; second, and more importantly, that the number of device-free (DF) studies targeting these activities is vanishingly small compared to other modalities. The valuable context provided by recognizing these activities—such as cooking pasta, preparing a meal, or making a beverage—could significantly enhance context-aware recommender systems and personalized advertising. Despite a rich body of HAR research in kitchens using other methods, device-free HAR for complex kitchen activities, particularly food preparation, constitutes a significant and unexplored research gap. This gap is the primary focus of the research presented in this dissertation.

Table 4 Kitchen-related human activity recognition. VB: vision based, W: wearable, OT: object tagged, DF: device free, ND: non-dining activity (1) eat, (2) get a snack, (3) drink, (4) eat soup, (5) eat chips, (6) eat pasta, (7) eat sandwich (8) eat with hands, (9) eat with spoon, (10) eat with chopstick, (11) drink with cup, (12) drink with bottle, (13) drink with straw, (14) cook pasta, (15) make tea, (16) prepare breakfast, (17) prepare dinner, (18) prepare lunch, (19) prepare meal, (20) cook, (21) make a sandwich, (22) make a bowl of cereals, (23) prepare beverage, (24) prepare hot beverage, (25) do the dishes, (26) wash hands, (27) wash using washing machine (28) tidy the kitchen, (29) clean the table (30) set the table, (31) open & close fridge, (32) open & close dishwasher, (33) do dining room activity, (34) do kitchen activity

Study	Year	Method				Dining Activity																																		ND
						Eating													Food Preparation											Cleaning					Other					
		VB	W	OT	DF	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	
[117]	2022				✓								✓	✓	✓	✓	✓	✓																						✓
[118]	2022		✓					✓																																✓
[119]	2022				✓	✓																																	✓	✓
[120]	2022		✓					✓	✓	✓	✓	✓																												
[121]	2022		✓		✓	✓																		✓	✓					✓				✓	✓					✓
[122]	2021				✓	✓		✓											✓	✓										✓	✓									✓
[123]	2021				✓	✓	✓	✓													✓	✓	✓	✓						✓		✓								✓
[124]	2021		✓			✓																																		✓
[125]	2021				✓	✓	✓														✓	✓	✓		✓					✓		✓								✓
[126]	2021	✓	✓																	✓						✓	✓			✓			✓		✓					✓
[127]	2021				✓	✓		✓																✓						✓										✓
[128]	2020		✓		✓			✓																				✓												✓
[129]	2020				✓	✓																	✓																	✓
[130]	2020				✓																			✓																✓
[131]	2020		✓	✓	✓			✓																												✓	✓			✓
[132]	2020		✓			✓																			✓					✓						✓				✓
[133]	2020				✓																				✓				✓											✓
[134]	2020		✓			✓													✓																					✓
[135]	2020				✓	✓																			✓															✓
[136]	2020		✓			✓																																		
[137]	2020	✓				✓																																		✓
[138]	2020				✓	✓																		✓																✓
[139]	2020		✓			✓																																		
[140]	2020		✓			✓																																		
[141]	2019		✓			✓																																		✓

3.4 Positioning this Dissertation

As illustrated in Figure 10, this dissertation is positioned at the intersection of two research domains: device-free human activity recognition and context-specific HAR within kitchen environments. While significant advancements have been made in each domain independently, their integration remains a largely unexplored area. Prior kitchen HAR research has predominantly relied on wearable, object-tagged, or vision-based methods, each with inherent limitations regarding privacy, usability, or deployment cost. Conversely, device-free HAR research has largely focused on simple, context-agnostic activities like walking, falling, or gesturing. This work directly addresses this gap by investigating the application of device-free sensing, specifically using WiFi, to recognize the complex, fine-grained, and interaction-rich activities characteristic of a kitchen.

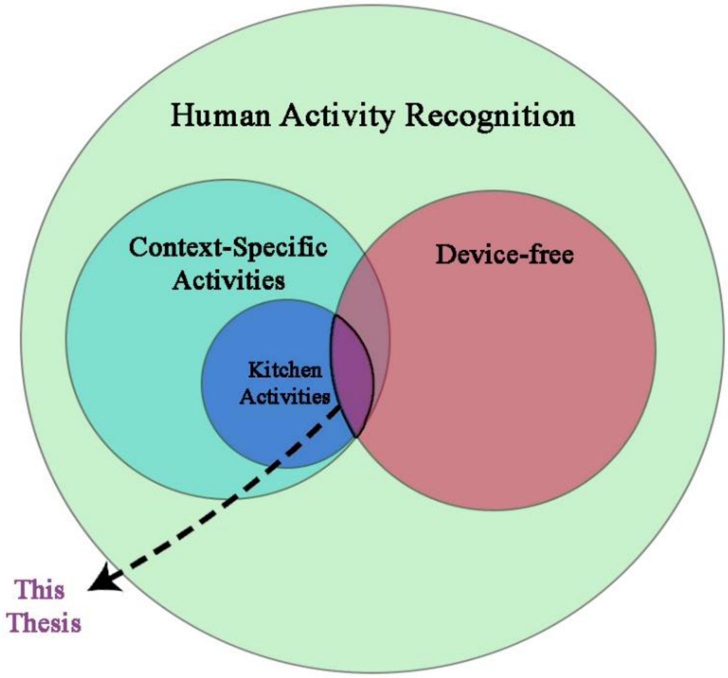

Figure 10 Positioning this dissertation

Chapter 4 Problem Definition

4.1 Motivation and Background

4.1.1. Smart Homes: From Automation to Intelligence

Smart homes are defined by ecosystems of interconnected sensors and actuators that enable the control, monitoring, and automation of household functions [142] . The foundational layer of smart home functionality involves controlling elements like lighting, heating, or appliances via explicit user commands or pre-programmed rules. These systems are largely static and possess limited adaptability to dynamic user conditions or contexts.

The field has since evolved from basic automation toward intelligent decision-making [143][153] and context-awareness [144]. Leveraging machine learning and advanced sensor technologies, modern smart homes can infer user preferences and behavior patterns by passively observing interactions within the environment. For instance, a system that learns a user consistently sets the temperature to 22°C at 10 PM could proactively execute this adjustment or issue a reminder if the user forgets, adapting its response based on its configured level of autonomy [145],[2].

This transition from automation to intelligence facilitates more personalized and efficient smart homes, enhancing their ability to support applications like aging-in-place [146], health monitoring [147], [148], and behavior recognition [149] (e.g., fall detection [150]). The integration of device-free sensing technologies is a key enabler of this intelligence, as it permits passive, continuous monitoring without requiring users to wear or interact with any device, thereby offering a privacy-preserving [151] and unobtrusive [5], [152] solution.

4.1.2. Personalized Advertisement and Data Monetization

Personalized advertising represents a significant commercial application enabled by intelligent smart homes. As illustrated in Figure 11, this can be facilitated through data trading systems that offer users compensation in exchange for access to their behavioral data, creating a ecosystem for targeted advertising and associated user benefits [153]. In this model, the system aggregates and processes user data, selling anonymized insights to service providers who then deliver contextually relevant advertisements.

Conventional approaches to advertising personalization often rely on low-level, explicit data streams such as direct sensor readings, purchase histories, and user-provided profile information. The integration of HAR introduces a paradigm shift. By inferring the specific activities a user is performing in real time, the system can deduce high-level intent and context. Sharing this rich contextual information (e.g., "user is preparing a chocolate cake") with service providers enables the delivery of profoundly more precise and timely advertisements (e.g., for a specific brand of chocolate or baking utensils), effectively translating ambient sensor data into actionable consumer insight.

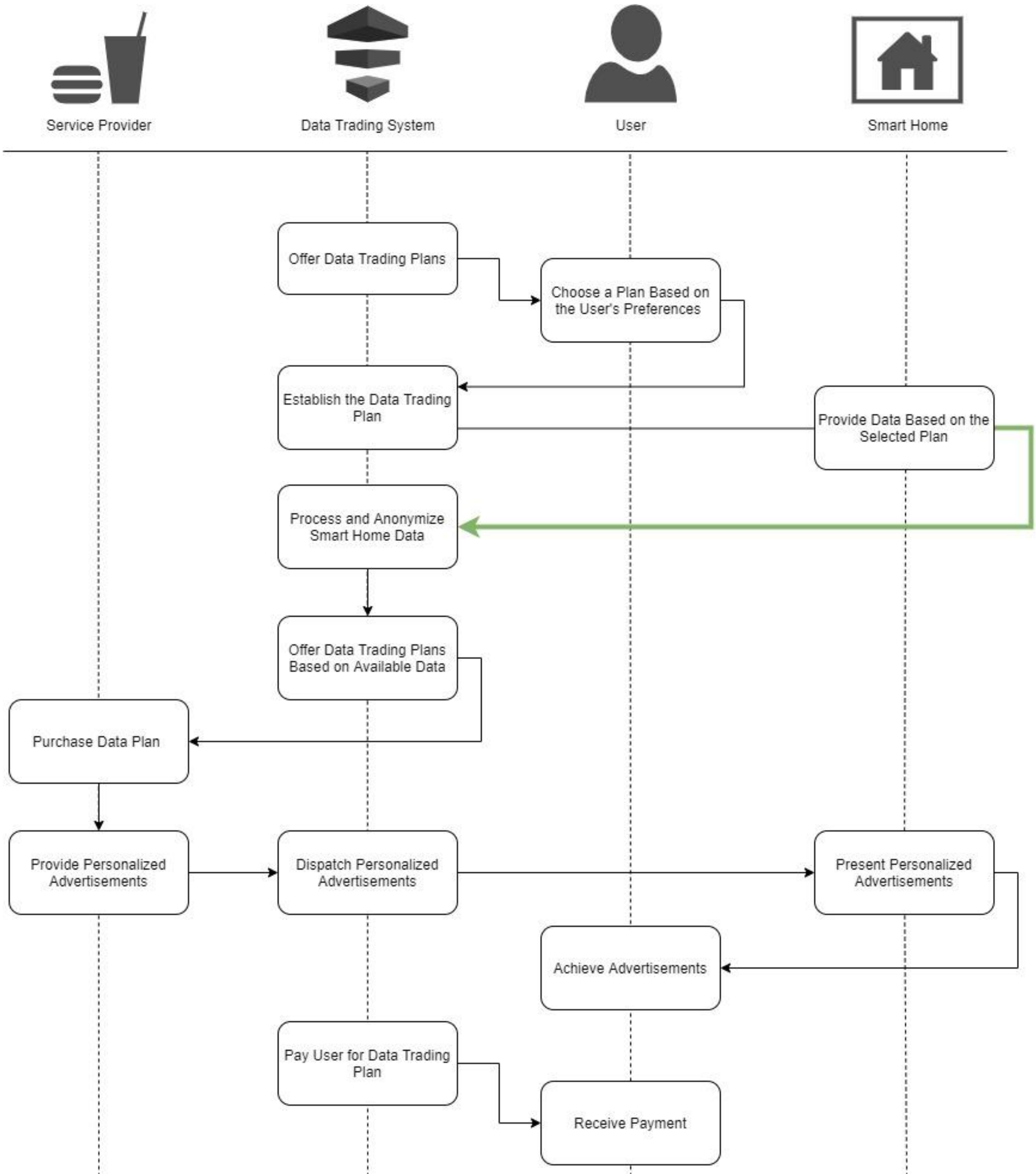

Figure 11 Personalized advertisement in smart home: a sequence diagram

4.1.3. The Role of HAR in Kitchen

Among the most valuable and activity-rich contexts within a smart home is the kitchen. This environment serves as the primary hub for food preparation and is a space where a wide variety of fine-grained, object-interactive, and complex activities occur. As established in Section 3.3.2, kitchen activity recognition underpins a wide range of applications, including health monitoring [110], dietary assessment [108], energy management, and assisted living for elderly or disabled [109].

A particularly compelling application is personalized advertising. By accurately recognizing user activities within the kitchen, smart home systems can infer immediate user needs and preferences in real time. For example, upon detecting that a user is preparing a specific meal (e.g., pizza), the system could proactively suggest related products (e.g., a premium mozzarella), complementary recipes, or limited-time promotions, offering highly timely and contextually relevant recommendations.

The utility of HAR for this application is fundamentally governed by its granularity. Coarse-grained recognition (e.g., determining that a user is active in the kitchen) enables the presentation of general, kitchen-related advertisements. Fine-grained recognition (e.g., identifying the specific activity of kneading pizza dough) allows for hyper-targeted advertisements (e.g., for a specialty olive oil or a pizza stone). This hierarchical precision enables a powerful win-win scenario: service providers achieve higher engagement and conversion rates through relevant advertising, while users benefit from offers that are genuinely useful and aligned with their immediate tasks.

4.2 System Terminology

4.2.1. Sensory System

The performance of any device-free HAR system is heavily reliant on the capabilities of its sensory system and the quality of the data it collects. Several key aspects of the sensor setup directly impact recognition performance. The most significant challenges include sensor selection, sensor placement, determining the optimal number of sensors, and ensuring real-time responsiveness.

Sensor Selection

The sensor is widely considered the most critical component in device-free HAR systems. The quality of a sensor directly affects the performance of activity recognition. Previous studies [29] comprehensively explained different sensors utilized in device-free HAR systems and compared them using various metrics like resolution, update rate, detection range, unobtrusiveness, processing complexity, calibration complexity, sensitivity, life span, weather dependency, form stability, electric noise coupling, occlusion, and power efficiency. However, a review of the literature reveals that this area remains underexplored; there is a notable lack of studies that systematically compare the effect of different sensor types on recognition performance across a large and diverse set of activities.

Sensor Placement

The effect of the environment on WiFi signal propagation is a well-established principle in the literature [154]. Our previous work [73] demonstrated that sensor placement has a direct effect on performance. In that study, all parameters affecting the signal were fixed to isolate the impact of placement. Cimdins et al. [156] designed a 3D model of an experiment layout including the sensors, direct paths, and echo paths resulting from reflections and investigated the optimal sensor placement for device-free localization. Neupane et. al [157] assumed that increasing the line-of-sight leads to the optimal WiFi transceiver placement. As an example, they theoretically and experimentally found that in a rectangular room, the best location for the WiFi transmitter and receiver is in the middle of two long edges.

Lately, Wang et al. [44] demonstrated that WiFi sensing coverage is closely related to the distance between the transmitter and receiver. Their experimental results showed that when the distance ranges from 0.5 to 4 meters, the sensing area takes the shape of an ellipse with a long axis of approximately 4 meters and a short axis of up to 3 meters. When the distance reaches 5 meters the coverage area would be like a Cassini oval and when the distance is more than 5 meters, two separated circles around the transmitter and receiver with a less than 2 meters radius would be shaped. Although their research theoretically studies the effect of LoS path length on sensing coverage and demonstrates it in two case studies, the results cannot be generalized to all environments and layouts. Finding the best sensor placement in a device-free environment to reach nearly optimal accuracy in recognizing fine-grained human activities is still an open challenge in this field.

Optimal Number of Sensors

Propagated signals in device-free HAR systems are generated and collected using transmitter and receiver sensors. Meanwhile, some sensors (like ESP32-WROVER-IE [158]) act as a transmitter and a receiver at the same time called transceivers. Previous studies have employed configurations ranging from a single transmitter-receiver pair to multiple receivers to collect propagated signals to recognize human activities [159]. Some studies claim that methods using a single receiver outperform those with multiple sensors [160]; however, such claims are often based on inequivalent comparisons and lack the scientific rigor required for a fair assessment of the impact of sensor count on performance. Nevertheless, it highlights the need for more studies in the field of device-free HAR to explore and validate such claims.

Real-time Responsiveness

Most real-world applications of device-free HAR systems, such as assistive services for elderly people [5], [161], [162] or personalized advertisement [153], require data to be analyzed in real time. Edge computing can be a solution to tackle this challenge [163]. So, it's crucial to know which sensors can provide this capability. It is possible to train a classification model on an edge device or just save and run the trained model on it. Liu et al. [164] proposed an RFID-based device-free HAR method which can be deployed on edge devices to achieve real-time gesture detection. They evaluated their method by officeroom4 [27] and Widar 3 [165] datasets. They achieved 99.1% accuracy 30 times faster than state of the art methods and experimentally showed that their model is practical and sufficient for real-time human activity recognition. A review of the literature suggests that this recent study [164] is among the few to demonstrate a practical, real-time implementation of a device-free HAR system, highlighting a significant area for further development.

4.2.2. System Users

A fundamental principle of device-free sensing is that system performance is influenced by how users perform activities. These methods operate by analyzing signal propagation, and the characteristics of any object—including its size, distance, material, and movement—affect signal reflections. As the human body reflects signals, the kinematics of a user's movements directly alter the resulting signal patterns.

User Dependency

One of the key challenges in device-free HAR is that many proposed methods are user-independent. To investigate the user-dependency of a recognition method, researchers evaluate classification models on a set of subjects different from those used in the training phase. For example, Huang et al. [166] proposed a user-independent HAR method to recognize granular human activities and tested their method on a separate set of users. Their results showed that the approach was user-independent, achieving over 98% accuracy in recognizing self-harvested actions such as clapping, hand waving, clicking, walking, sitting, bending, jumping, and kicking. Similarly, Kim et al. [167] achieved an average accuracy of 90% in recognizing 10 gestures from new users who were not involved in the training process. Despite some efforts to create user-independent models, the results of the systematic literature review [32] (Figure 12) indicate that the majority of studies in the field have not robustly addressed or evaluated this critical issue.

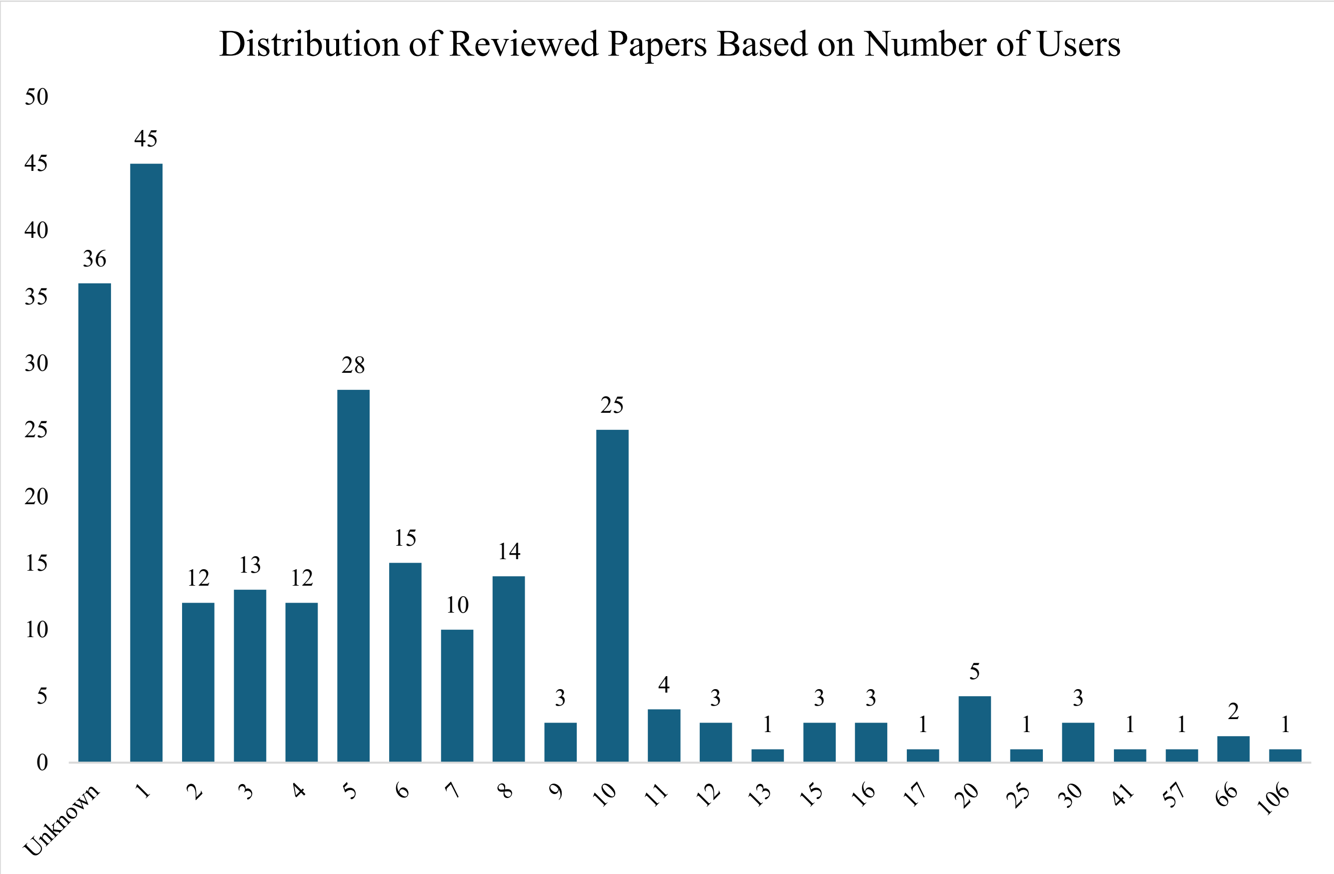

Figure 12 Distribution of reviewed papers based on number of users

Multi-subject Activities

In many real-world applications of device-free HAR systems, more than one user is performing activities at the same time; including activities that need to be done by more than one user [16] (e.g., playing chess, shaking hands, hugging) or individual activities that are performing at the same time in the same environment by two subjects [168] (e.g., one user doing the dishes while the other one cutting the bread). As Figure 13 from our systematic literature review [32] indicates, few researchers have addressed supporting multi-subject activities (14%) in device-free HAR that shows a significant knowledge gap.

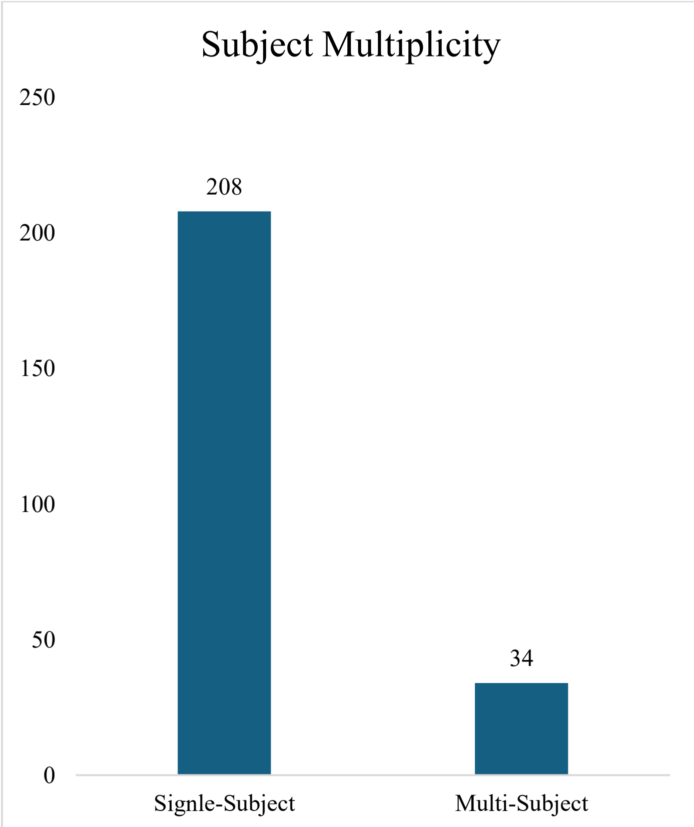

Figure 13 Subject multiplicity in reviewed papers

Interferer Presence

Some real-world scenarios contain more than one human subject present in the environment but only one or some of them are the subjects of the device-free HAR method. In these situations, although other human subjects are not the targets of the system, they should not be ignored. These interferers still affect the signal propagation more than a static object (e.g., furniture). Previous studies have attempted to address this challenge by considering the impact of interferers during the denoising step [169]. Lately, Jin et al. [170] could achieve 93.01% accuracy in recognition of six human gesture including waving, greeting, passing by, clapping, handshaking, and hugging in presence of three type of interfere activities including walking, running, and behavior (like clapping and passing) using mmWave radars. However, interference in HAR systems is still underexplored.

4.2.3. System Environment

The deployment environment is a defining characteristic of a device-free HAR system, directly influencing signal propagation. Therefore, it is crucial to investigate environmental characteristics that impact system performance. For instance, previous studies have successfully identified different surface materials and textures using WiFi signals [155]. By measuring WiFi signal reflections, they achieved 95%

accuracy in classifying five materials—copper, aluminum, plywood, birch, and human—demonstrating that surface type significantly influences WiFi signal behavior. Therefore, the material composition of walls, appliances, and furniture is an important characteristic of the environment to consider when placing sensors in WiFi-based HAR systems. Consequently, researchers strive to conduct experiments in realistic settings to ensure their results are generalizable and not artifacts of a specific, controlled lab condition.

Cross Environment Performance

The operational principle of device-free HAR is predicated on the physics of signal propagation. This process involves the transmission of a signal and the analysis of its reflections from all entities within a space, including walls, furniture, and humans. Consequently, the resulting signal pattern is exquisitely sensitive to the environment; any alteration in furniture type, position, orientation, or room layout will alter the propagation paths and change the received signal. This means the signal signature will differ between two rooms with identical layouts but different furniture, and between two rooms with similar furniture but different layouts. This inherent sensitivity renders device-free HAR methods highly dependent on their deployment environment.

A core challenge arising from this dependency is cross-environment generalization—the ability of a model to perform accurately in a new, unseen environment. While no single metric directly quantifies this capability, the systematic literature review conducted for this dissertation [32] reveals that researchers have attempted to demonstrate it by training and testing models in distinct environments (see Figure 14). The scarcity of robust solutions in this area underscores the magnitude of the challenge. Figure 14 illustrates that 65% of the papers (157 out of 242) reviewed in our SLR [32] evaluated their methods in only a single environment. Furthermore, among the studies conducted in multiple environments, a significant numbers failed to adopt a cross-environment validation protocol (e.g., training in one environment and testing in another). This prevalence of non-genralizable evaluation methods highlights a knowledge gap regarding cross-environment generalization of device-free HAR methods.

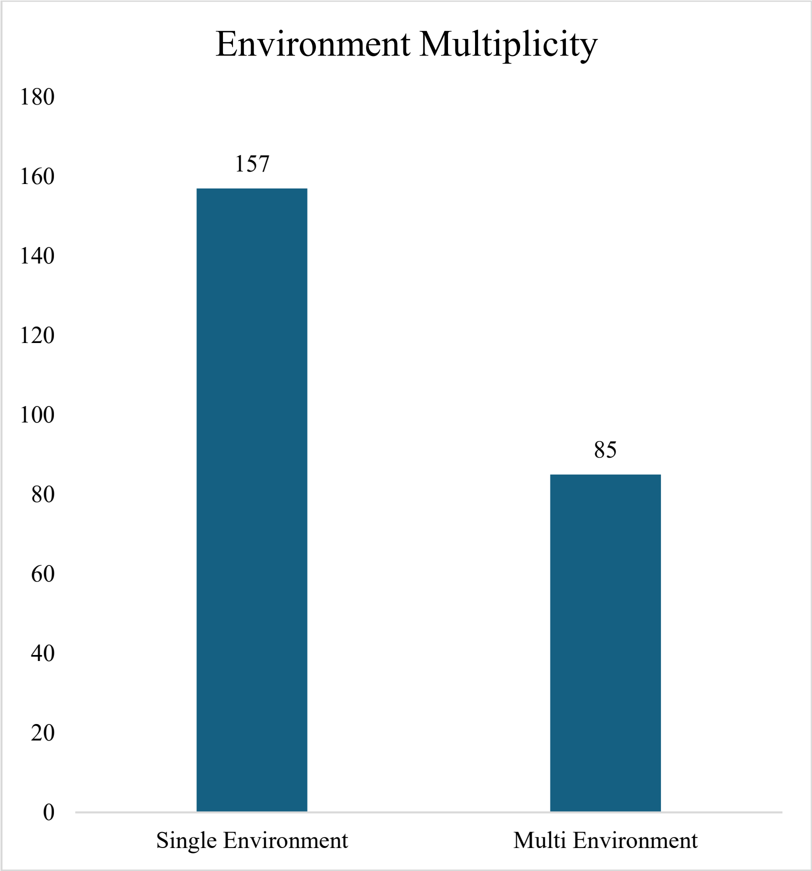

Figure 14 Environment multiplicity in reviewed papers

Real-world data gathering

Given that signal propagation is so intimately tied to the environment, evaluating the performance of a device-free HAR system under realistic conditions is paramount for assessing its real-world applicability. The literature shows that experiments are typically conducted in one of three tiers of environmental realism:

- Controlled Laboratory Environment: A space designed explicitly for experimentation, where the placement of objects and sensors is optimized for data capture rather than realism. Users perform scripted activities solely for data-gathering purposes [16].
- Realistic Simulated Environment: An environment designed to closely mimic a real-world setting (e.g., a mock kitchen or living room built in a lab), though data collection remains the primary goal. Users still perform activities prompted by an experiment protocol [171].
- In-Situ Environment: A genuine real-world setting (e.g., an actual home kitchen). Data is collected during the course of users' normal, daily routines, meaning activities are performed naturally and are not scripted for the experiment [172].

Gather data in a Simulated environment

The results of our SLR [32] indicated that only 13% of studies utilized existing datasets, highlighting a critical field-wide challenge: the scarcity of large, diverse, and publicly available datasets. The primary impediment is the high cost and logistical complexity of collecting high-fidelity data in real-world settings.

Simulation-based data generation [210] presents a promising yet underexplored solution to this data scarcity problem. By accurately modeling electromagnetic wave propagation within 3D simulated environments, it is possible to synthetically generate vast, comprehensive datasets. These virtual datasets could encompass a wide range of fine-grained and coarse-grained activities, performed by virtual humans with different kinematics, across countless environmental contexts and layouts. This approach represents a significant potential paradigm shift for the field, offering a path to overcome its fundamental data limitations and serving as a crucial area for future research.

4.3 Problem Formulation

This dissertation focuses on developing novel methods for accurate, device-free human activity recognition (HAR) in kitchen environments using Wi-Fi signals. The research specifically addresses core challenges in environmental robustness, recognition granularity (from coarse to fine-grained activities), and specific-context activity recognition for real-world application.

4.3.1. Statement of the Research Problem

Despite significant advances in device-free HAR, a critical gap exists between its theoretical potential and its practical, reliable application in real-world kitchen environments. This gap is characterized by four interconnected core problems:

Problem 1. The Context-Specific Granularity Problem

While device-free HAR has proven effective for coarse-grained activities (e.g., walking, sitting, falling) in general settings, its ability to reliably distinguish the fine-grained and subtle activities specific to a kitchen context (e.g., chopping vs. stirring, pouring vs. shaking) remains unproven and underexplored. Determining how existing methods can be adapted or extended to recognize such subtle activities remains an unresolved challenge.

Problem 2. The Environmental Robustness Problem

The performance of WiFi-based HAR is notoriously sensitive to the environment.

A) Most existing systems are trained and evaluated in controlled, single-environment labs, leading to models that fail to generalize.

B) There is a profound lack of understanding of how to achieve cross-environment robustness, ensuring a model developed in one kitchen layout performs accurately in another with different furniture, appliance placement, and material compositions. This lack of generalizability is a major barrier to real-world deployment.

To expand on the notion of generalizability, it is important to examine whether models can consistently recognize the same type of activity (with same intentional action and performed with comparable motions) across environments with differing spatial and material configurations.

Problem 3. The Data Scarcity and Benchmarking Problem

Progress in any machine learning domain is fueled by large, public, and well-defined datasets. The field of device-free kitchen HAR suffers from an absence of inclusive public datasets that encompass a wide range of fine-grained activities collected in realistic settings. This scarcity impedes the development, fair comparison, and benchmarking of new algorithms. This problem is closely tied to Problem 2, as creating inclusive datasets that capture activities across multiple real-world environments would not only enable fair benchmarking but also foster the development of models with greater environmental robustness.

Problem 4. The Optimization Problem

The optimal approach for kitchen HAR is unknown. It is unclear whether classical machine learning or deep learning is best suited to handle the high-dimensional, noisy nature of WiFi signals (CSI/RSSI) for this specific task. Furthermore, key system parameters, most notably the optimal placement of transceivers within the complex RF environment of a kitchen, have not been systematically investigated to maximize recognition quality.

Therefore, the overarching research problem targeted by this dissertation is the development and validation of a cross-environment robust, accurate, and generalizable system for fine-grained, device-free human activity recognition in real-world kitchen environments using WiFi signals. This work tackles the problems of granularity, environmental dependency, data scarcity, and feature selection optimization to bridge the gap between laboratory proof-of-concepts and practical smart home applications.

4.3.2. Research Questions

This dissertation seeks to address the following research questions to tackle the problems introduced in previous section:

- RQ1: What are the key steps and processes involved in recognizing human activities in a device-free context? (to address Problem 4)
- RQ2: Can WiFi-based device-free sensing be utilized to recognize kitchen activities with high accuracy without requiring users to carry or wear any device? (to address Problem 1)
- RQ3: To what extent can WiFi Channel State Information (CSI) and Received Signal Strength Indicator (RSSI) signals be used to recognize fine-grained human activities in a kitchen context within real-world environments? (to address Problem 1)
- RQ4: How can the performance of kitchen activity recognition in real-world environments be improved? (to address Problem 2)
    - RQ4.1: How does the type of activities in the kitchen context influence HAR performance, and how does recognition accuracy vary across different sets of activities? (to address Problem 3)
    - RQ4.2: How does sensor placement in the kitchen context affect HAR performance, and what configurations yield the highest recognition accuracy? (to address Problem 4)
- RQ5: How can classical machine learning and deep learning approaches be optimized for device-free kitchen activity recognition? (to address sProblem 4)

4.3.3. Research Approach

To answer the research questions and address the identified research problems, this dissertation follows a structured, stepwise research approach. The approach begins with a systematic literature review to establish foundational knowledge, identify methodological patterns, and highlight key knowledge gaps in device-free HAR. Building on these insights, an initial device-free HAR method is proposed and tested using a publicly available dataset, followed by its extension and evaluation for fine-grained kitchen activity recognition (Problem 1). Subsequent steps involve adapting the method for real-world environments (Problem 2), designing and collecting and inclusive kitchen activity dataset (Problem 3), and finally developing a multi-method HAR system that integrates classical, deep learning, and transfer learning approaches (Problem 4).

Each step is explicitly designed to target one or more of the four research problems and is mapped to specific sections of the dissertation. This mapping is summarized in Table 5, which illustrates how each step contributes to the overall research and ensures coverage of all identified challenges.

Step 1. Systematic Literature Review

Given the vast and fragmented body of work in device-free HAR, a systematic literature review of more than 2,000 research papers was conducted to identify foundational principles, common steps, dominant research focuses, potential research gaps, and reusable methods applicable to each stage of the HAR pipeline. This review aimed to uncover best practices for recognizing fine-grained human activities in specific contexts, such as kitchens, with a particular focus on addressing Problem 4 by highlighting available methods and knowledge gaps in device-free HAR. The results of this review are integrated throughout the dissertation and can be found in Sections 2.43.1, 3.2, 4.2, and 7.1.1

Step 2. Initial Device-Free Kitchen HAR Study

Building on existing public HAR datasets and state-of-the-art methods, this step involves replicating and refining device-free HAR models as an initial step toward kitchen-specific activity recognition, thereby assessing feasibility and identifying potential methodological improvements prior to developing a context-specific system. A publicly available dataset containing fine-grained human activities was selected to evaluate the proposed approach. The outcomes of this feasibility study are presented in Chapter 5, that is a base method to address Problem 1.

Step 3. Fine-Grained Device-Free Kitchen HAR

Given the unique challenges of kitchen environments and the need for fine-grained activity recognition, adapt and extend existing device-free HAR methods to ensure they can accurately distinguish between closely related kitchen activities. In addition, this step investigates how variations in activity granularity and sensor placement influence the performance and reliability of device-free HAR methods in kitchen environments. In this step, Problem 1 is mostly targeted by investigating the feasibility of an extended version of proposed initial device-free HAR method (Step 1) in recognition of fine-grained kitchen activities as presented in 5.2. Investigation of the effect of granularity on quality of the device-free HAR method is also performed in 5.3 to enforce the importance of Problem 1. At the end of this step, 5.4 investigate the quality of the device-free HAR method on different sensor placements to suggest the base sensor placement for collecting inclusive dataset highlighted in Problem 3.

Step 4. Kitchen Activity Dataset Collection

Given the absence of large-scale, publicly available datasets for device-free kitchen activity recognition, design and collect a comprehensive RF-based dataset covering a diverse range of kitchen activities, ensuring it supports both coarse- and fine-grained recognition tasks and can be used to benchmark different methods. This step addresses Problem 1, Problem 2, and Problem 3 as it includes an inclusive dataset containing granular activities from various environments. Subsections of 0 explains how this inclusive dataset includes a range of granular activities (Section 6.1.3) and various real-world environments (Section 6.1.2).

Step 5. Multi-Method HAR System

Given RF signal measurements of kitchen activities, develop and evaluate a hybrid recognition system that integrates deep learning, transfer learning, and classical machine learning approaches to identify the most effective method, and optimize the best-performing approach for recognizing an extended range of kitchen activities. While this step is primarily targets Problem 4, the experiments designed after optimization in Section 6.4.4 and Section 6.4.5, targets Problem 1 and Problem 2 by recognizing granular activities and recognizing environment-activity pairs respectively.

Table 5 Mapping of Research Steps to Addressed Problems and Relevant Sections

	Problem 1	Problem 2	Problem 3	Problem 4
Step 1	Section 3.2 Device-free Human Activity Recognition	Section 4.2.3 System Environment	Section 4.2.3 System Environment	Section 2.4 Machine Learning Methods
Step 2	Section 5.1 Feasibility Study 1	-	-	-
Step 3	Section 5.2 Feasibility Study 2, Section 5.3 Feasibility Study 3	-	Section 5.4 Feasibility Study 4	-
Step 4	Section 6.1.3 Activity	Section 6.1.2 Environment	Section 6.1 Dataset Collection	-
Step 5	Section 6.4.4 Experiment 2	Section 6.4.5 Experiment 3	-	Section 6.4 Learning Optimization

Chapter 5 Feasibility Study

This chapter presents four feasibility studies (Step 2 and Step 3 of research approach) conducted to establish the potential of device-free HAR methods to recognize fine-grained kitchen activities, directly addressing address Problem 2 and Problem 3 outlined in this dissertation.

- Feasibility Study 1 (Section 5.1)validates the core technical pipeline by applying a novel CSI-RSSI fusion method to a general human interaction dataset. Its success demonstrates the method's technical feasibility and its potential to be extended for granular kitchen activity recognition, thereby addressing Problem 2.
- Feasibility Study 2 (Section 5.2) investigates the method's transferability to the kitchen domain. The results confirm the possibility of recognizing fine-grained kitchen activities while highlighting specific challenges, providing critical insights for Problem 2.
- Feasibility Study 3 (Section 5.3) quantitatively isolates the impact of activity granularity. By demonstrating that high performance on coarse-grained activities does not guarantee success with fine-grained ones, it underscores a key challenge central to Problem 2.
- Feasibility Study 4 (Section 5.4) evaluates the critical role of sensor placement within a kitchen environment. Its findings directly motivated the optimal sensor configuration used in Dataset Collection step, a decision that addresses Problem 3.

5.1 Feasibility Study 1: Validation on an Existing Dataset

5.1.1. Study Design and Objectives

This feasibility study was conducted to develop a robust, device-free HAR method [47] using a public dataset [173]. The primary objective was to create a competitive benchmark model that effectively fuses Channel State Information (CSI) and Received Signal Strength Indicator (RSSI) features for recognizing fine-grained human-to-human interactions. Success in this general domain would validate the proposed signal processing and feature engineering pipeline (Figure 15) before its application to the specific context of kitchen activity recognition.

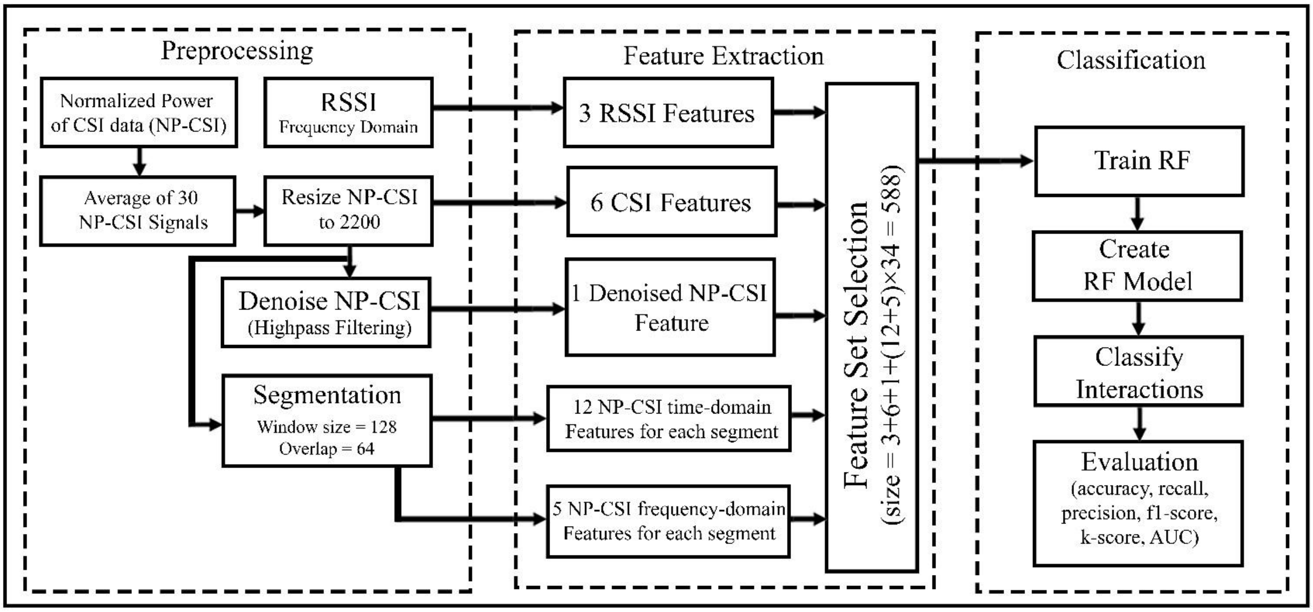

Figure 15 Feasibility Study 1 - A novel method for fine-grained human activity recognition by combining CSI and RSSI features

5.1.2. Dataset

The study utilized the dataset [173], which captures Wi-Fi signals (CSI and RSSI) during interactions between two human subjects. The dataset comprises 12 interactions performed by 40 different subject pairs, with each pair repeating each interaction 10 times. From these, seven distinct interactions were selected for this study: approaching, departing, handshaking, hugging, pointing (right hand), punching (right hand), and pushing. Each data sample consists of CSI data (30 subcarriers) from a 2x3 MIMO system and RSSI data, with varying packet lengths up to 2,200.

5.1.3. Preprocessing Pipeline

The raw data underwent a multi-stage preprocessing pipeline:

1. Averaging: To reduce dimensionality, the 30 subcarriers of the CSI data were averaged into a single time-series vector for each transmitter-receiver antenna pair.
2. Normalized Power Calculation: The averaged CSI data was converted to normalized power (NP-CSI) using the formula $\text{Normalized Power} = \frac{|H_i|^2}{N}$, where H_i is the CSI value and N is the packet length. This emphasizes signal variations caused by movement.
3. Length Standardization: All NP-CSI sequences were resized to a fixed length of 2,200 packets

(the maximum in the dataset) using linear interpolation to ensure uniform feature extraction.

4. Noise Reduction: A high-pass filter was applied in the frequency domain to remove low-amplitude noise components (Algorithm 1).
5. Segmentation: The preprocessed NP-CSI data was segmented using a sliding window (size=128 packets, overlap=64 packets) for temporal feature extraction. RSSI data was used without segmentation for coarse-grained features.

Algorithm 1 Removing Noise using High Pass Filter

Input: Average of 30 subcarriers NP-CSI data
Output: Average of 30 subcarriers NP-CSI data without noise
1: $FT = Fourier\ Transform\ of\ data\ signal$
2: for i = 0 to $length(FT)$ do
3: if $FT[i] < 20$ then
4: $FT[i] = 0$
5: end if
6: end for
7: $IFT = Inverse\ Fourier\ Transform\ of\ FT$
8: return IFT

5.1.4. Feature Engineering

A comprehensive set of 588 features was extracted per sample, falling into two categories:

- Global Features (10): Seven frequency-domain features were extracted from the entire NP-CSI sequence (F1-F7), capturing overall signal characteristics like dominant frequencies and entropy. Three features were extracted from the RSSI data (F8-F10), capturing the base energy level from each receiver antenna. The detailed algorithms for extracting these global features are described in our previously published work [47].
- Windowed Features (578): For each of the 34 segments, 12 time-domain features (F11-F22) and 5 frequency-domain features (F23-F27) were extracted. These included statistical measures (e.g., kurtosis, mean absolute deviation), signal properties (e.g., number of peaks), and spectral properties (e.g., median frequency, occupied bandwidth). The complete list, mathematical definitions, and implementation details for all windowed features are provided in [47].

The entire feature extraction pipeline, along with a comprehensive evaluation of each feature's contribution, has been peer-reviewed and published in [47]. This manuscript can be consulted for full algorithmic details and implementation code.

5.1.5. Classification and Hyperparameter Tuning

The resulting 588-dimensional feature vectors were used to train seven classical machine learning classifiers (SVM, GNB, DT, LR, LDA, KNN, RF). A generalized model was trained on data from all subject pairs to ensure real-world applicability. A rigorous grid search was employed to find the optimal hyperparameters for each classifier (Table 6).

Table 6 Classification hyperparameters

Classifier	Hyperparameters		
	Parameter	Search Space	Best Value
Support Vector Machine (SVM)	kernel	'rbf', 'sigmoid'	’rbf’
	C (regulation)	0.1,1, 10, 100	0.1
	gamma	1,0.1,0.01,0.001	0.001
Gaussian Naïve Bayes (GNB)	Smoothing variable	1e-2, 1e-3, 1e-4, 1e-5, 1e-6, 1e-7, 1e-8, 1e-9, 1e-10, 1e-11, 1e-12, 1e-13, 1e-14, 1e-15	1e-14
Decision Tree (DT)	criterion	'gini', 'entropy'	'entropy'
	splitter	'best', 'random'	'best'
	max_depth	range (1, 10)	6
	min_samples_split	range (1, 10)	2
	min_samples_leaf	range (1, 5)	8
Logistic Regression (LR)	Penalty	l1, l2	l2
	C	0.001, 0.01, 0.1, 1, 10, 100, 1000	0.001
	Solver	'newton-cg', 'lbfgs', 'liblinear'	'newton-cg'
Linear Discriminant Analysis (LDA)	solver	'svd', 'lsqr', 'eigen'	'svd'
	store_covariance	True, False	True
	tolerance	0.0001, 0.001,0.01, 0.1	0.1
K Nearest Neighbors (KNN)	n_neighbors	range(3,10)	6
	weights	'uniform', 'distance'	'distance'
	algorithm	'auto', 'ball_tree', 'kd_tree', 'brute'	'auto'
	p	1, 2	1
Random Forest (RF)	criterion	'gini', 'entropy', 'log_loss'	'log_loss'
	max_depth	range(1, 10)	9
	min_samples_split	range(1, 10)	5
	min_samples_leaf	range(1, 5)	1

5.1.6. Evaluation Method

Model performance was evaluated using an 80/20 train-test split. Six metrics were used: Accuracy, Precision, Recall, F1-Score, Cohen's Kappa, and Area Under the Curve (AUC) [174]. These metrics provide a holistic view of performance, balancing per-class effectiveness (Precision, Recall) and overall agreement (Kappa).

5.1.7. Results and Discussion

Initial experiments on all 12 interactions yielded a maximum accuracy of 67.81% with Random Forest (RF), with confusion analysis revealing inherent ambiguities between certain activities (e.g., left vs. right-hand pointing). By focusing on the seven most distinguishable interactions, performance improved dramatically. The RF classifier achieved the best performance on the refined 7-class problem, with an accuracy of 94.16% (see Figure 17). The confusion matrix (see Figure 16) shows high per-class accuracy, with minimal confusion between classes.

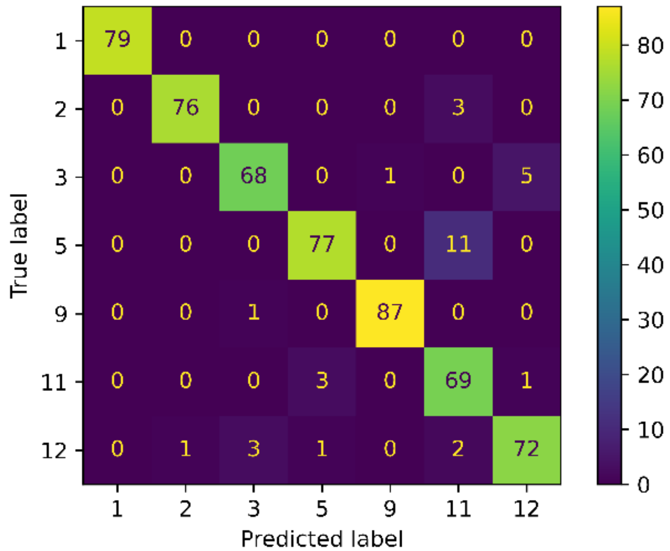

Figure 16 Feasibility Study 1 - Confusion matrix using the RF method for recognizing 7 interactions

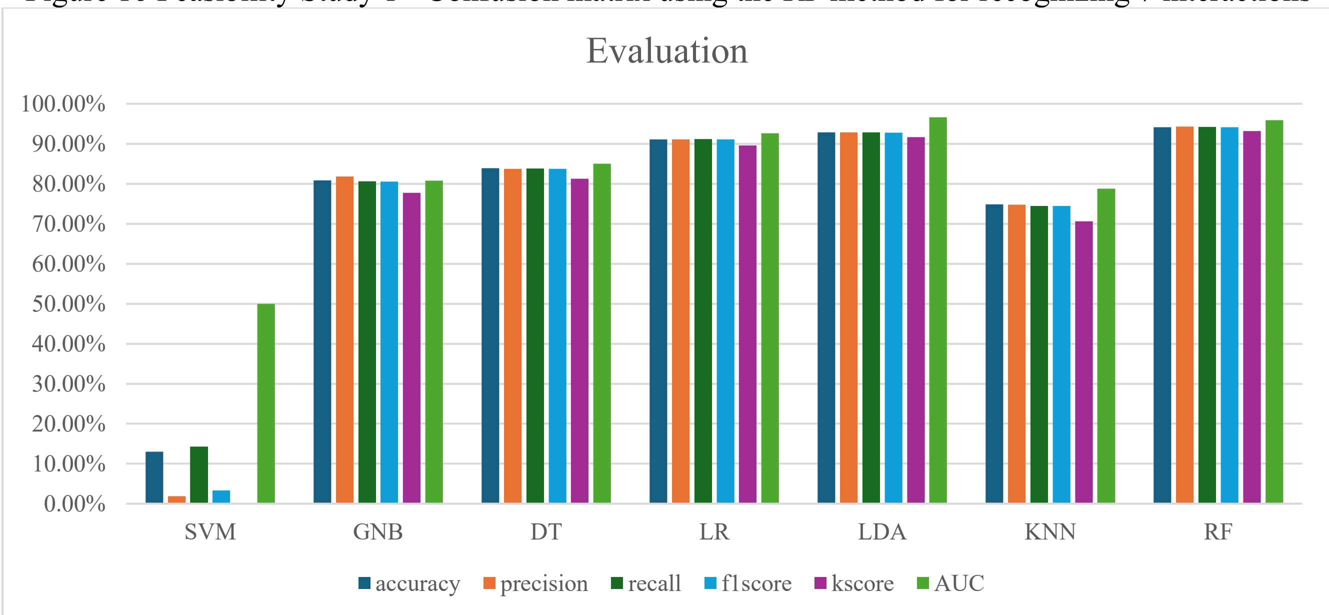

Figure 17 Feasibility Study 1 - Evaluation of different classification methods for recognizing 7 interactions

5.1.8. Conclusion of Feasibility Study 1

This study successfully developed and validated a high-accuracy device-free HAR method using a fusion of CSI and RSSI features. The achieved performance of 94.16% is competitive with state-of-the-art results on this dataset. Crucially, this result demonstrates the feasibility of the core technical approach, sophisticated preprocessing and multi-domain feature extraction from Wi-Fi signals for fine-grained activity recognition. The effectiveness of this pipeline provides a strong foundational justification and a method blueprint for the subsequent research presented in this dissertation, which applies and adapts this approach to the specific challenges of kitchen activity recognition.

5.2 Feasibility Study 2 – Performance on Fine-Grained Kitchen Activities

5.2.1. Study Design and Objectives

This study aimed to evaluate the transferability and limitations of the device-free HAR method developed in Feasibility Study 1 [47] by applying it to the recognition of fine-grained activities in a kitchen context [106]. The primary objective was to determine whether a model designed to recognize human-to-human interactions could be effectively repurposed for distinguishing complex, object-centric kitchen activities, specifically six basic knife skills: chop, French cut, cube, slice, julienne, and mince (see Figure 18). A peer-reviewed publication based on this feasibility study can be found in [106].

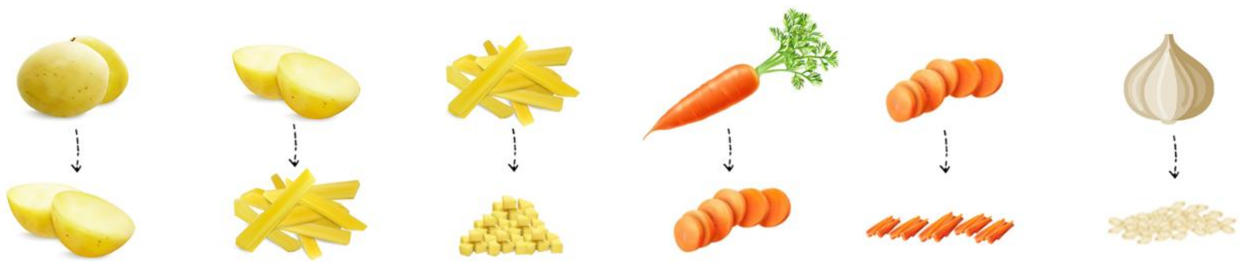

Figure 18 Feasibility study 2 - six basic knife activities in cooking. From left to right: chop, French cut, cube, slice, julienne, and mince.

5.2.2. System Architecture and Method

The system architecture for data collection and activity recognition is illustrated in Figure 19. A user performed activities within the propagation path between an ESP32 microcontroller (functioning as a WiFi transceiver) and an iPhone 12 mini (functioning as a receiver). The received WiFi Channel State Information (CSI) and Received Signal Strength Indicator (RSSI) data were streamed to a laptop for recording.

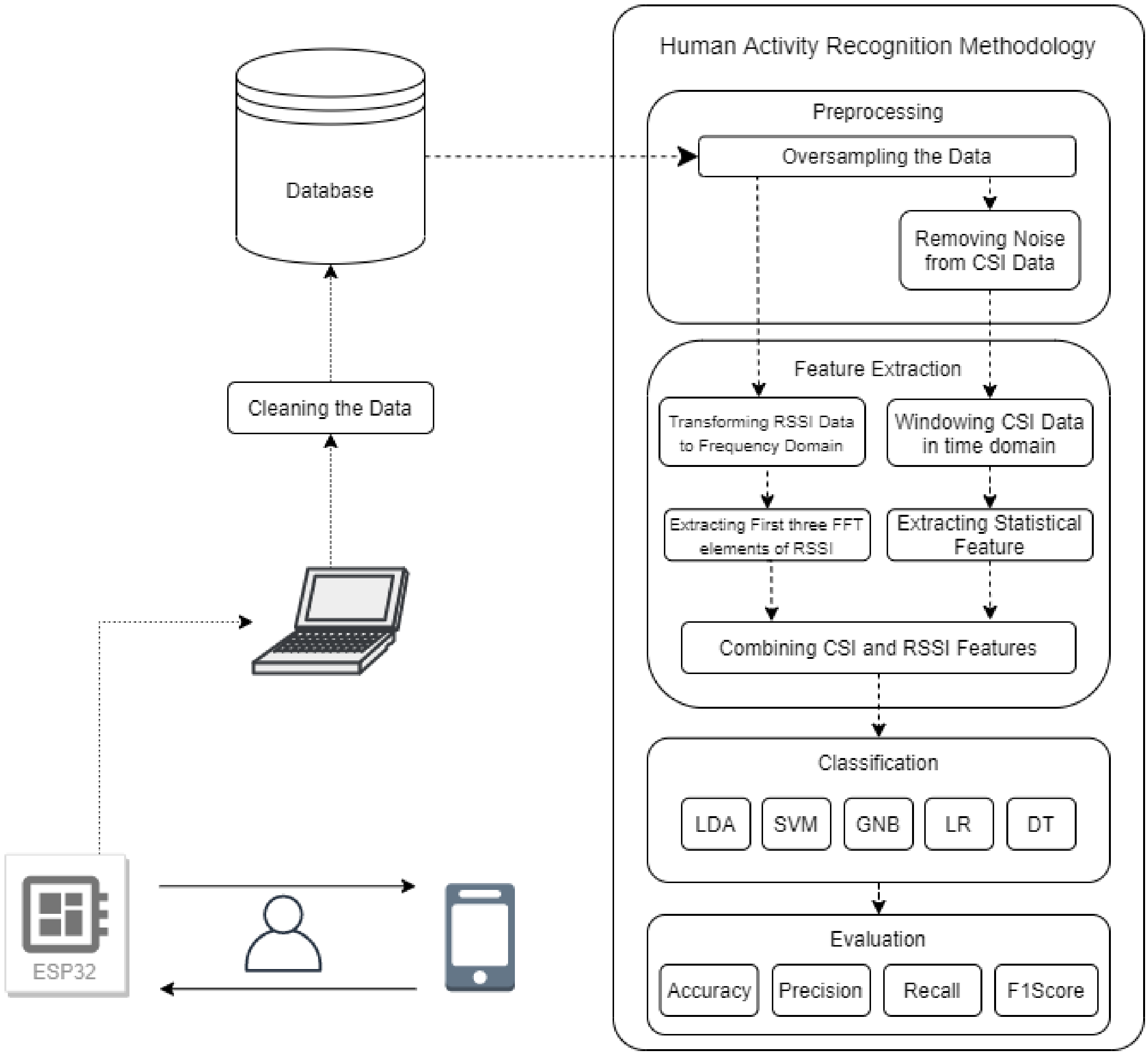

Figure 19 Feasibility study 2 - system architecture

The core HAR method from Feasibility Study 1 [47] was employed, consisting of:

1. Data Oversampling: To augment the limited dataset, a novel oversampling technique was applied. For each recorded data sample, ten new samples were generated by iteratively removing the first five packets from the previous sample. This approach is valid for these activities as they consist of sequential, repetitive motions; removing initial repetitions does not alter the fundamental activity signature.
2. Preprocessing and Feature Extraction: The preprocessing pipeline (noise removal, length standardization via linear interpolation) from Study 1 was applied. The feature set was also replicated, extracting:

- Global Features: Seven CSI features (F1-F7) and three RSSI features (F8-F10) from the entire sample [48].
- Windowed Features: Seventeen statistical features (e.g., number of peaks, kurtosis, mean frequency) were extracted from windowed segments of the CSI data (window size=64, overlap=32).

3. Classification: The resulting feature vectors were used to train and evaluate multiple classical machine learning classifiers (LDA, SVM, GNB, LR, DT).

5.2.3. Experimental Setup and Data Collection

The physical setup comprised an ESP32 transceiver and an iPhone receiver (Figure 20), positioned in a controlled room (Figure 21) with a specific layout (Figure 22). Two participants were asked to perform each of the six knife activities (see Figure 23) five times, resulting in an initial dataset of 60 samples. Following the oversampling procedure, the dataset was expanded to 660 samples for analysis. The cleaned dataset has been made publicly available [175].

Figure 20 Feasibility study 2 - system instruments: iPhone 12 mini (right) as a WiFi receiver and ESP32 (left) as a WiFi transceiver.

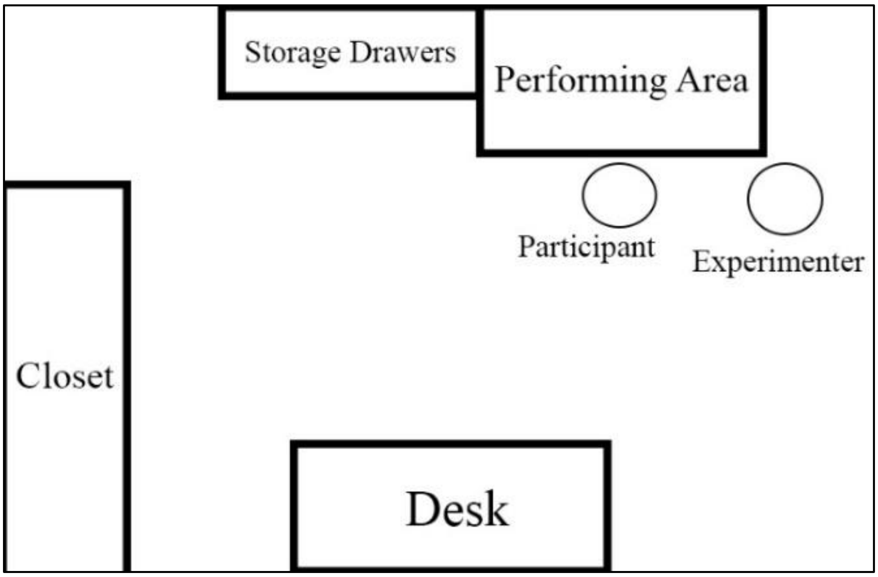

Figure 21 Feasibility study 2 - experiment room

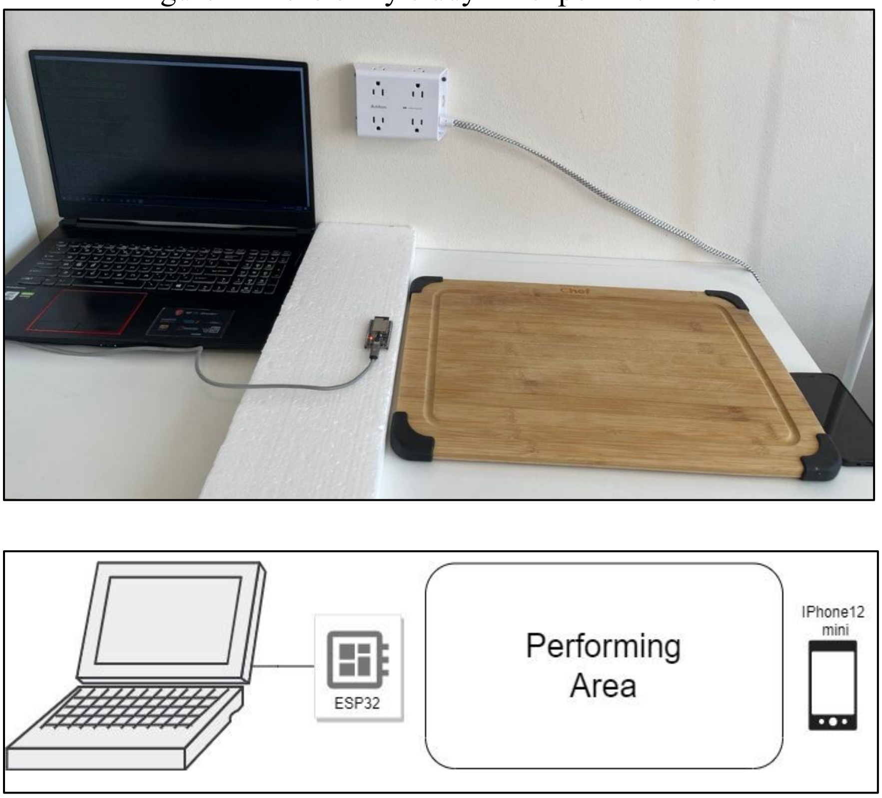

Figure 22 Feasibility study 2 - layout of the system experiment

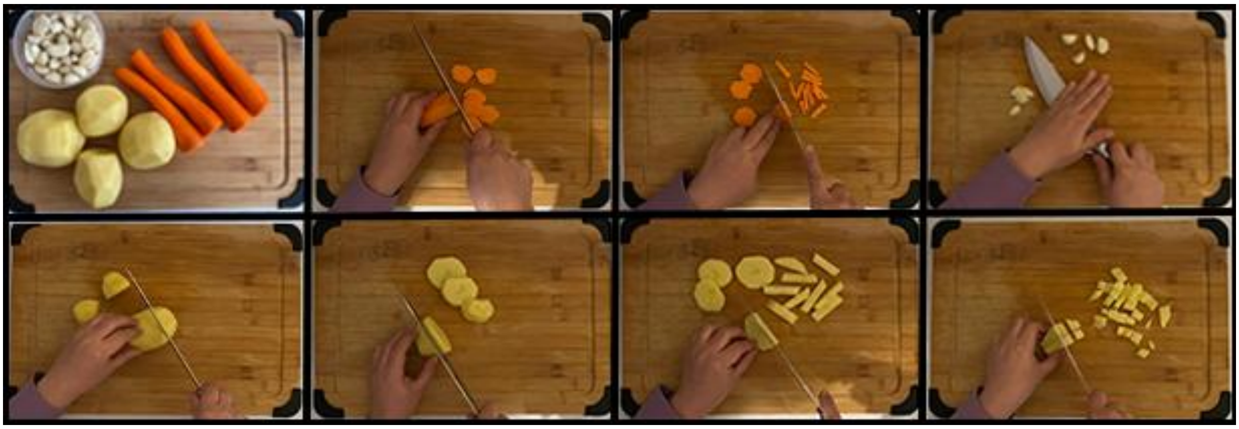

Figure 23 Feasibility study 2 - experiment photos

5.2.4. Results and Discussion

The study evaluated performance on two tasks:

- Task 1: recognition of three activities (chop, French cut, slice)

  The SVM classifier achieved the highest accuracy of 74% (Figure 24). The confusion matrix (Figure 25) shows good distinguishability between these three activities, demonstrating that the method generalized reasonably well from human interactions to this specific subset of kitchen activities.

- Task 2: recognition of all six activities (i.e., chop, French cut, cube, slice, julienne, and mince)

  Performance dropped significantly to an accuracy of 36%, which is unacceptable for practical application.

The high performance on three activities confirms the feasibility of using WiFi sensing for some fine-grained kitchen tasks. However, the sharp performance decline when including all six activities highlights the fundamental challenge and a key limitation of the transferred method: CSI and RSSI data alone appear to lack the discriminative power to distinguish between very similar, fine-grained object manipulations (e.g., cubing vs. julienne). This suggests that the current feature set, while effective for larger body movements, may be insufficient for this more complex domain. Furthermore, the dataset size, even after augmentation, may still be too small for the complexity of the problem.

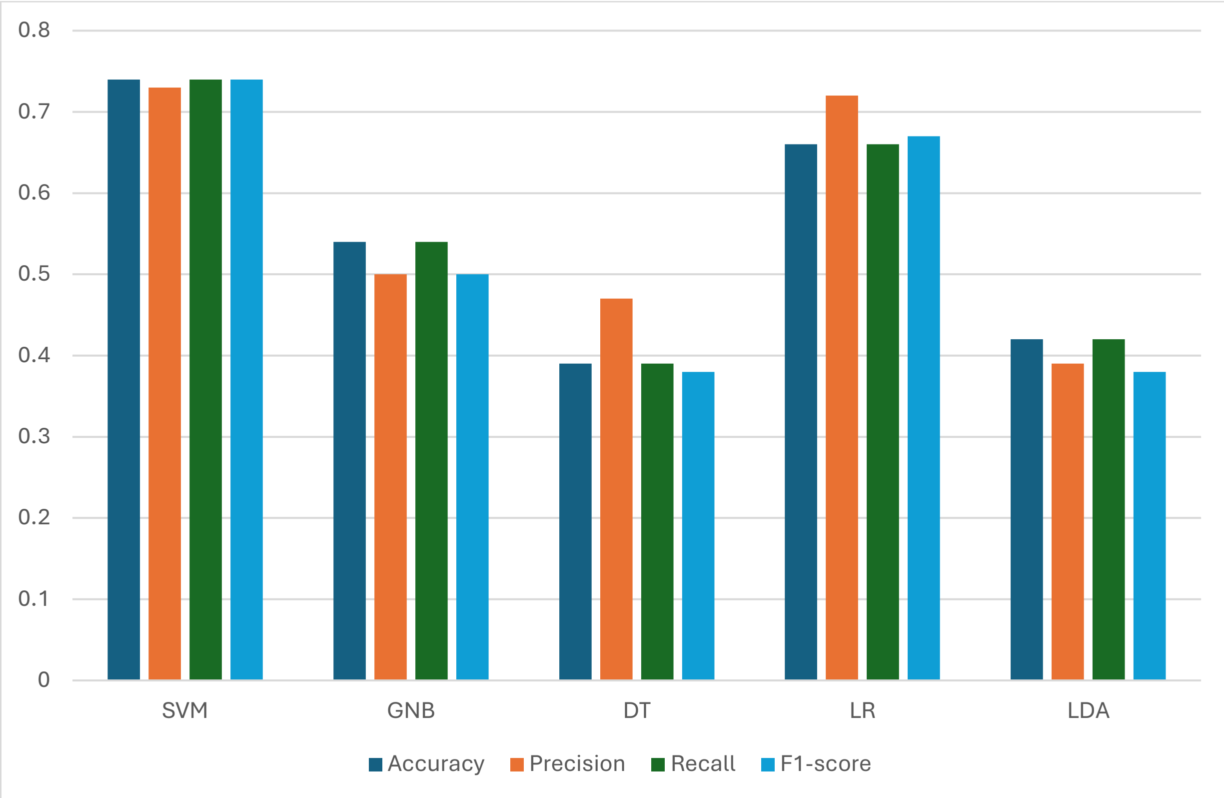

Figure 24 Feasibility study 2 - performance evaluation of classification methods in recognition of three basic knife activities in cooking (chop, French cut, and slice)

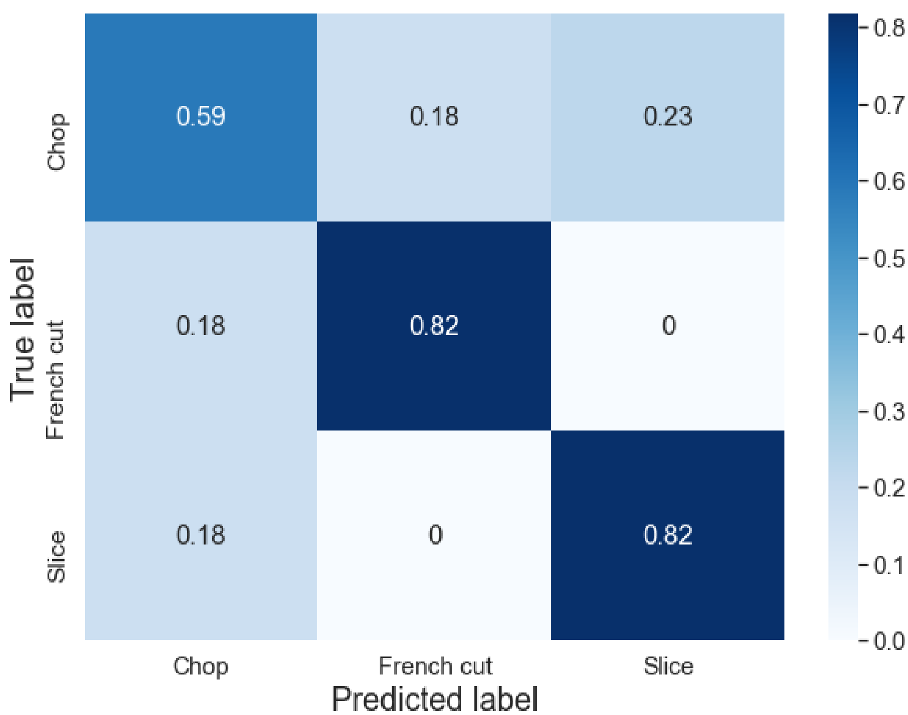

Figure 25 Feasibility study 2 - confusion matrix of using SVM to classify three basic knife activities; i.e., chop, French cut, and slice.

5.2.5. Conclusion and Implications

As detailed in [106], This study demonstrated that a device-free HAR method designed for human interactions can achieve modest accuracy (74%) when applied to a limited set of three fine-grained kitchen activities. However, its performance degrades severely when applied to a fuller range of six highly similar activities (36% accuracy).

The findings lead to several critical conclusions for the future direction of this research:

- Transferability is limited: Methods for coarse-grained activities do not automatically translate to fine-grained tasks.
- New features are needed: The solution likely requires novel feature engineering or deep learning approaches specifically designed to capture the subtle nuances of kitchen activities.
- Multi-Modal sensing may be necessary: Fusing WiFi data with other contextual cues (e.g., object presence via RFID or computer vision) could be essential for high accuracy.
- Larger datasets are crucial: significantly larger and more varied datasets are required to train robust models for this complex domain.

Thus, while proving initial feasibility, this study primarily serves to clearly define the specific challenges that must be addressed to achieve robust device-free HAR in kitchen environments, which will be the focus of the subsequent chapters of this dissertation.

5.3 Feasibility Study 3 – The Impact of Activity Granularity

5.3.1. Study Design and Objectives

While the performance gap between coarse and fine-grained activity recognition is well-documented for wearable and vision-based systems [179], it remains quantitatively unexplored in device-free HAR. It is often assumed that fine-grained activities are more challenging to recognize, but the magnitude of this effect and its implications for method transferability are unknown. This study aimed to scientifically investigate whether a device-free HAR method that performs well on coarse-grained activities can maintain its performance when applied to fine-grained activities, under otherwise identical conditions [107]. The objective was to isolate and measure the effect of activity granularity on recognition accuracy.

5.3.2. System Architecture and Method

The system architecture is shown in Figure 26. The experimental setup was held constant: the same kitchen from Study 5.4, the optimal sensor placement (Placement 2), the same user, and the same data collection apparatus (ESP32 transceiver, iPhone receiver) were used.

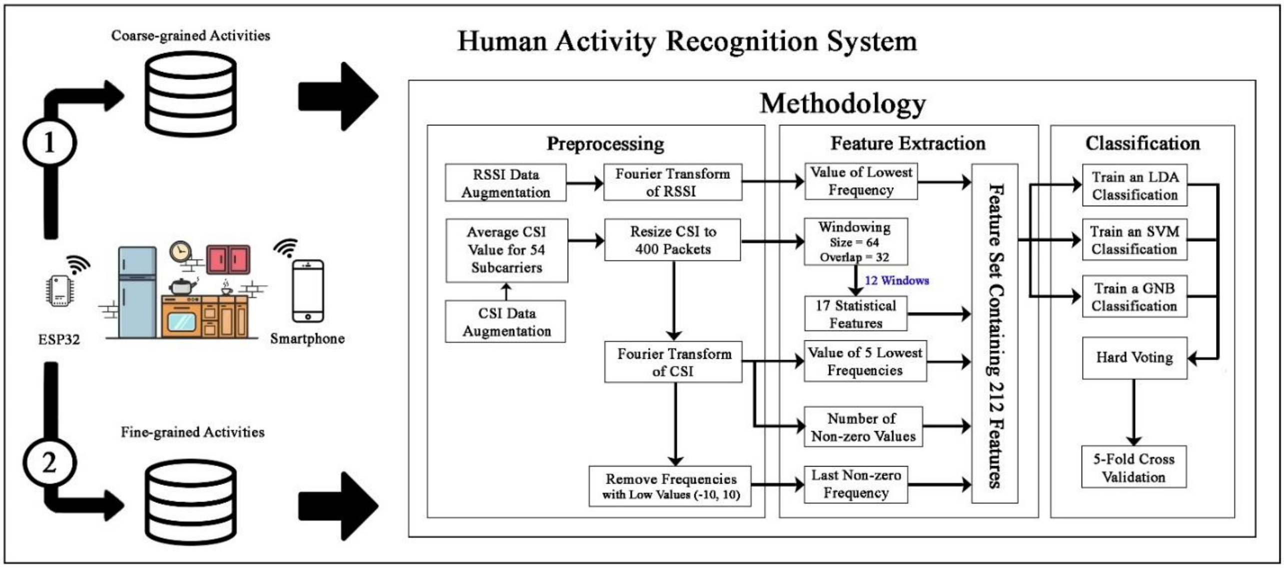

Figure 26 Feasibility study 3 - system architecture

Data Collection and Augmentation

The user performed two distinct activity sets:

- Coarse-grained Activities: Stir-frying potatoes, moving cans from a fridge to a cabinet, and filling/turning on an electric kettle (Figure 27).
- Fine-grained Activities: Chopping, slicing, and French cutting potatoes (Figure 35).

Each activity was performed five times, creating an initial dataset. To mitigate the small sample size and improve model generalization, a data augmentation algorithm (Algorithm 2) was employed. This algorithm applied random noise injection ($\varepsilon = \pm 0.05$) to the CSI and RSSI values of each packet, generating 99 new synthetic samples from each original sample, resulting in a final dataset 100 times larger.

Feature Extraction and Classification

The feature extraction pipeline from [47] was applied, resulting in a 212-dimensional feature vector for each sample. A hard-voting ensemble classifier was implemented, combining the predictions of three individual classifiers (Gaussian Naïve Bayes, Linear Discriminant Analysis, and Support Vector Machine) that demonstrated the highest preliminary accuracy. The final classification decision was made by majority vote.

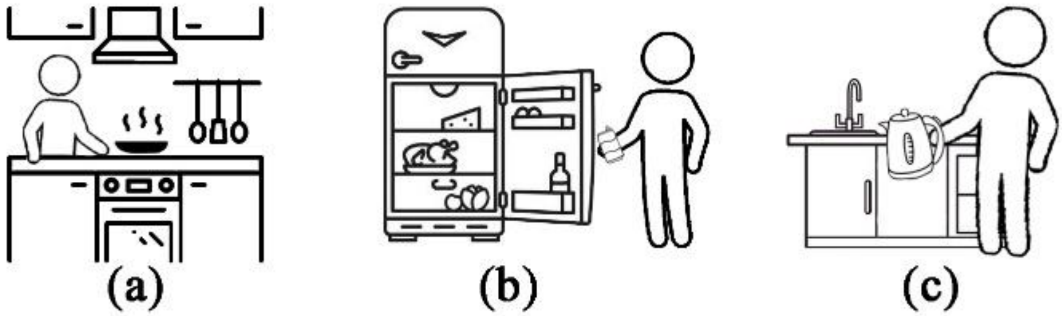

Figure 27 Feasibility study 3 - three coarse-grained activities. from left to right: a) stir-frying cubed potato, b) taking several cans from the fridge and putting them on a cabinet, c) filling an electric kettle with water and turning it on

Algorithm 2 Data Augmentation

Input: Data Entry
Output: 99 Augmented Data Entry
$data = Input\ Data\ Entry$
$listOfAugmentedData = [\,]$
for i = 0 to 99 do
  $packetlist_{aug} = [\,]$
  for $packet$ in $data.packets$:
    $\varepsilon_1 = random\ number\ in\ [-0.05, 0.05]$
    $RSSI_{aug} = (1 + \varepsilon_1) \times packet.RSSI$
    $CSI_{aug} = [\,]$
    for CSI in $packet.CSI$:
      $\varepsilon_2 = random\ number\ in\ [-0.05, 0.05]$
      $CSI_{aug}.add((1 + \varepsilon_2) \times CSI)$
    end for
    $packet_{aug} = Packet\ (RSSI_{aug}, CSI_{aug})$
    $packetlist_{aug}.add(packet_{aug})$
  end for
  $data_{aug} = Data\ (packetlist_{aug})$
  $listOfAugmentedData.add(data_{aug})$
end for
return $listOfAugmentedData$

5.3.3. Evaluation and Results

Model performance was rigorously evaluated using 5-fold cross-validation for both activity sets independently.

- Coarse-grained activities: The voting ensemble classifier achieved a mean accuracy of 92.27% (Figure 28). The confusion matrix (Figure 29) shows high per-class accuracy with minimal confusion.

- Fine-grained activities: Using the identical method, feature set, and classifier, the mean accuracy dropped significantly to 51.53% (Figure 30). The confusion matrix (Figure 31) shows substantial confusion between all three activities.

The dramatic 40% drop in accuracy for fine-grained activities can be attributed to two fundamental challenges intrinsic to the device-free sensing paradigm. First, fine-grained activities (e.g., chopping vs. slicing) involve smaller, more localized, and more similar limb motions, primarily in the hands and forearms. These subtle movements generate correspondingly smaller and less distinctive perturbations in the ambient WiFi multi-path profile compared to the large, full-body displacements of coarse-grained activities (e.g., moving between fridge and cabinet). Second, the signal features extracted from CSI and RSSI data, while effective for distinguishing gross motor patterns, lack the discriminative resolution to capture the nuanced temporal and spectral signatures of near-identical hand kinematics. Consequently, the feature vectors for different fine-grained activities become statistically inseparable in the model's learned space, leading to near-chance-level classification performance with a standard activity recognition pipeline.

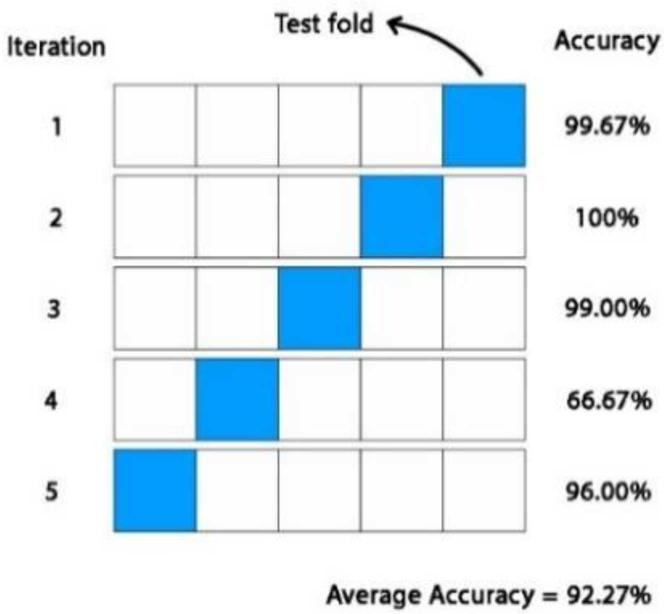

Figure 28 Feasibility study 3 - fold cross-validation of presented method for coarse-grained activity recognition.

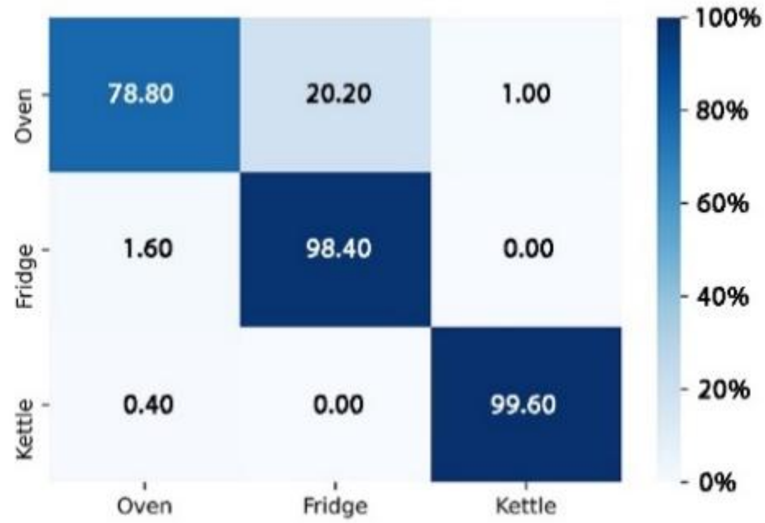

Figure 29 Feasibility study 3 - recognition accuracy of coarse-grained activities i.e., stir-frying cubed potato, taking several cans from the fridge and putting them on a cabinet, and filling an electric kettle with water and turning it on.

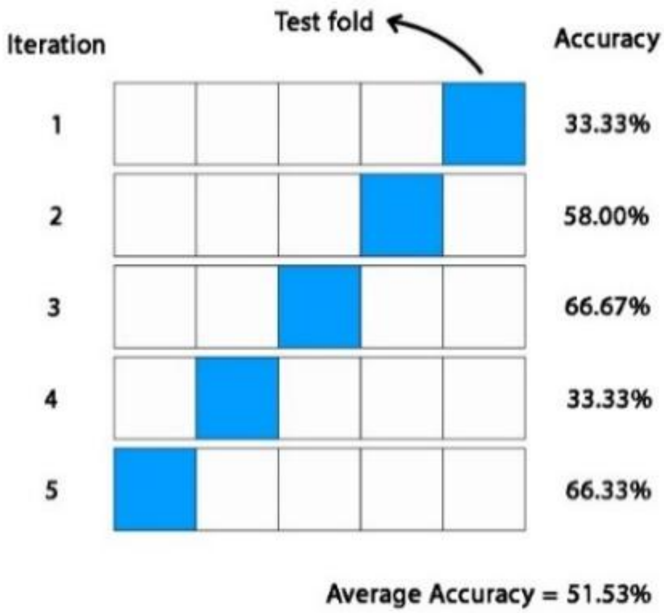

Figure 30 Feasibility study 3 - five-fold cross-validation of presented method for fine-grained activity recognition.

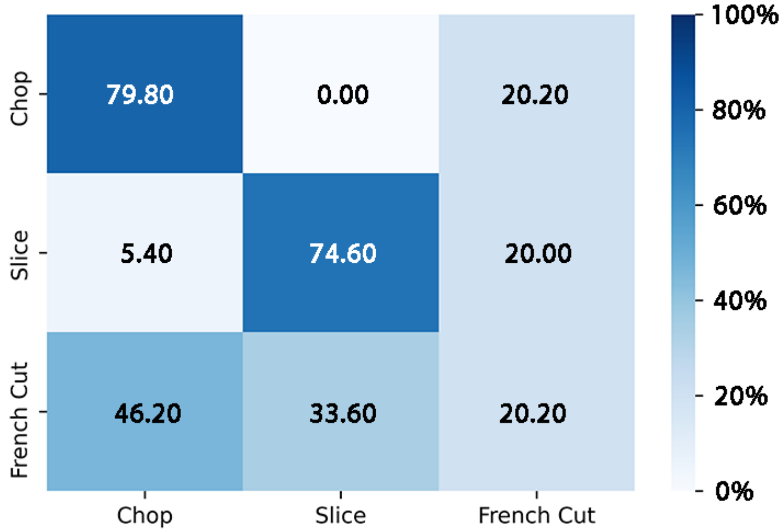

Figure 31 Feasibility study 3 - accuracy of fine-grained activities i.e., chopping, slicing, and French-fries cutting

5.3.4. Conclusion

This study provides conclusive evidence that activity granularity is a dominant factor influencing the performance of device-free HAR systems. The results demonstrate a stark performance gap: a system achieving high accuracy (92.27%) on coarse-grained activities failed to maintain comparable performance on fine-grained activities (51.53%) when all other variables were rigorously controlled. This indicates that methods developed for coarse-grained activities are not automatically transferable to fine-grained tasks. The challenge of fine-grained recognition is fundamental and necessitates specialized techniques in feature engineering, model architecture, and learning paradigms, which defines the core problem addressed in the remainder of this dissertation.

5.4 Feasibility Study 4 – Sensor Placement in Kitchen Environment

5.4.1. Study Design and Objectives

The performance of device-free Human Activity Recognition (HAR) systems is inherently tied to the propagation of wireless signals, which is heavily influenced by the physical environment. While previous research has optimized sensor placement for maximum coverage or signal strength [44], [157], [176], a critical gap exists in understanding how strategic placement within a complex, cluttered environment like a kitchen directly affects the discriminative power of the extracted data for activity classification. The underlying hypothesis is that placement governs the Line-of-Sight (LoS) conditions between the radio path, the user, and the activity area, which in turn determines the nature and strength of the multi-path signal propagation caused by human motion. This study aimed to scientifically evaluate the effect of sensor placement on device-free HAR performance in an authentic kitchen. The primary objective was to develop and evaluate an empirical method for optimizing sensor placement by determining which sensor placement provides the most distinguishing data for recognizing fine-grained kitchen activities [73]. The optimization criterion was the statistical separability of activity classes in the extracted feature space.

5.4.2. System Architecture and Method

The research method followed a three-tiered pipeline illustrated in Figure 32: data acquisition, feature extraction, and statistical evaluation.

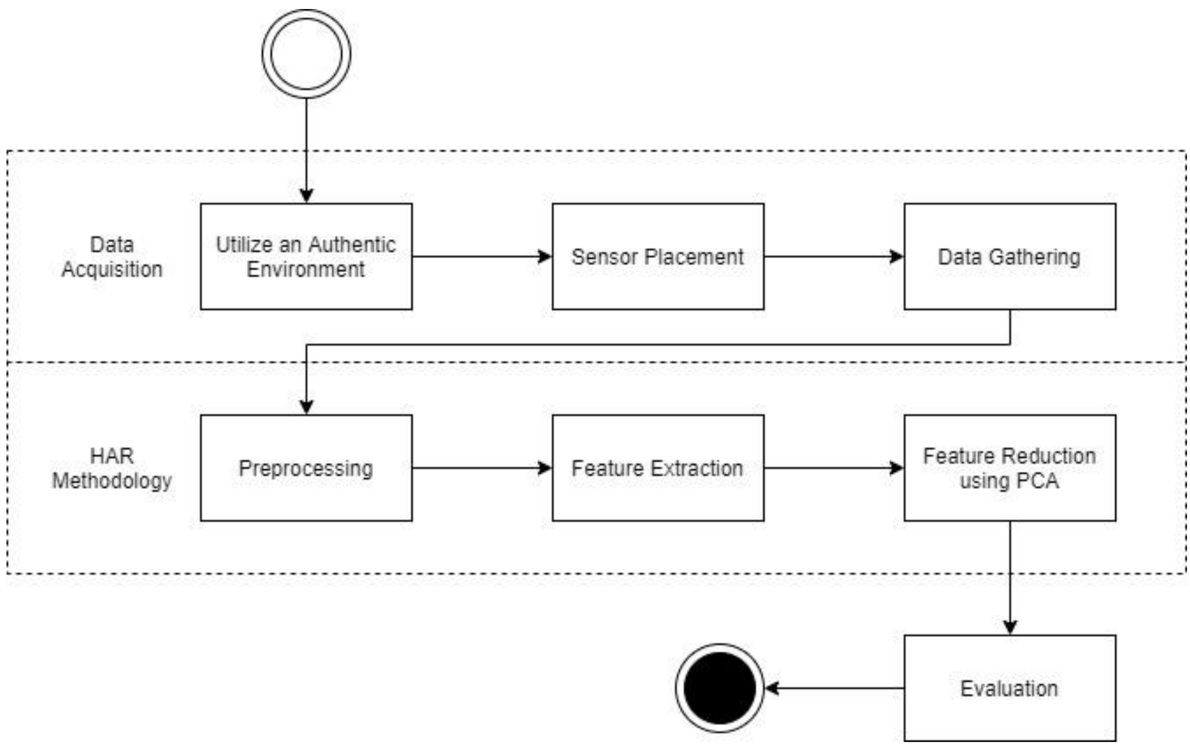

Figure 32 Feasibility study 4 - research method

Data Acquisition

Data was collected in a fully furnished, authentic kitchen (237 cm x 226 cm x 244 cm), containing standard appliances, cabinets filled with utensils, and various obstacles (Figure 33). Three strategic sensor placement configurations were deployed using an ESP32 WiFi transceiver and an iPhone 12 mini receiver (see Figure 34 and Table 7):

- Placement 1: The performing area was in the Line-of-Sight (LoS) of the sensors, while the user was in Non-Line-of-Sight (NLoS).
- Placement 2: Both the user and the performing area were in LoS.
- Placement 3: Both the user and the performing area were in NLoS.

A single participant performed three fine-grained knife activities (chopping, slicing, and French cutting potatoes; Figure 35) five times each for every sensor placement, following a strict timeline (Figure 36). This procedure resulted in 45 raw data samples (3 activities × 5 trials × 3 placements).

Table 7 Feasibility study 4 - the sensor placement strategies

Sensor Placement	User	Performing Area
Placement 1	NLoS	LoS
Placement 2	LoS	LoS
Placement 3	NLoS	NLoS

Figure 33 Feasibility study 4 - the authentic kitchen used in the experiment.

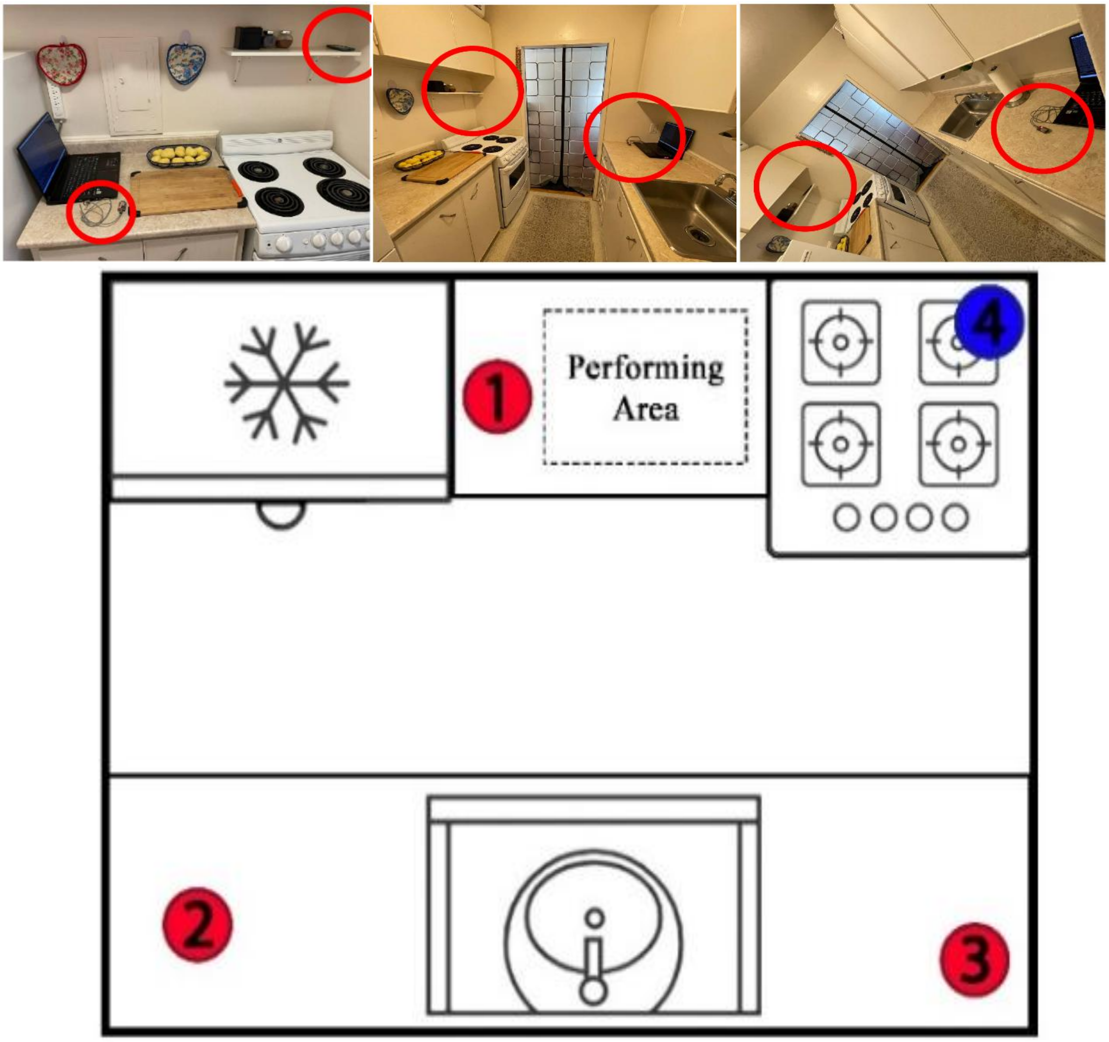

Figure 34 Feasibility study 4 - three sensor placements in the experiment layout.

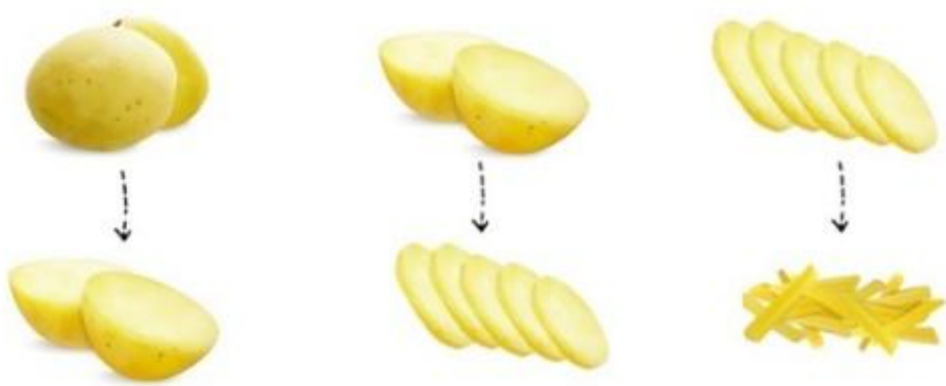

Figure 35 Feasibility study 4 - three basic knife activities in kitchen: chop, slice, and French cut (. from left to right).

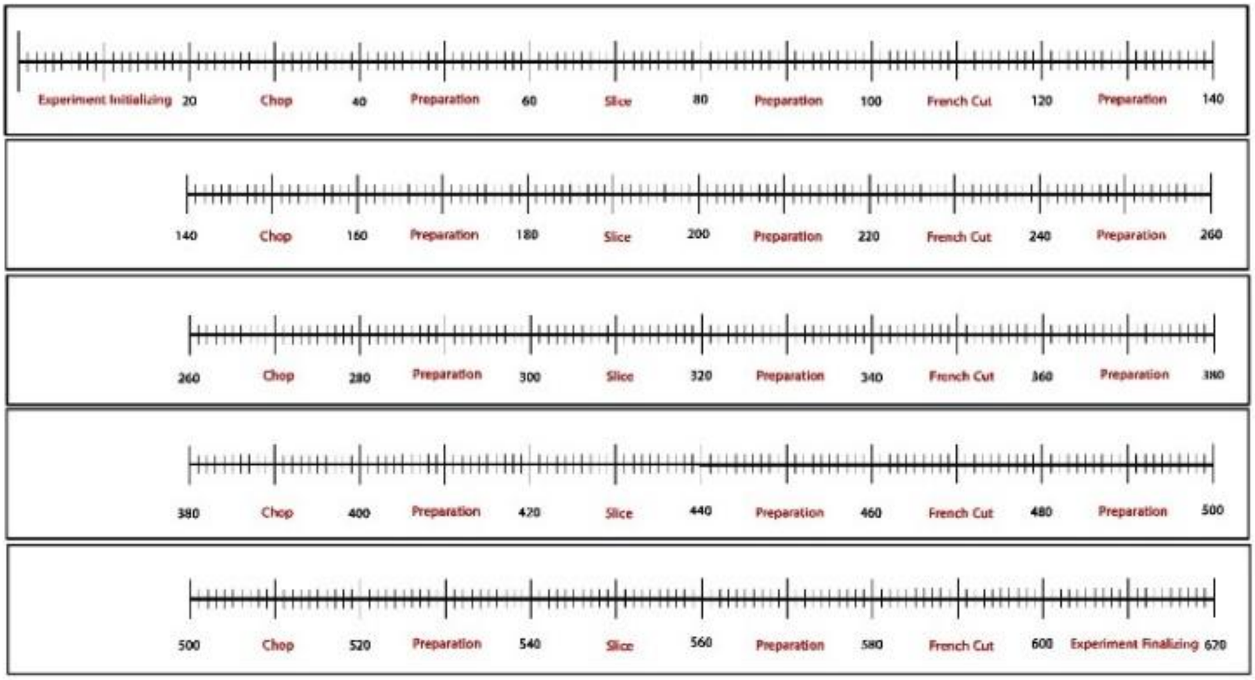

Figure 36 Feasibility study 4 - user activity timeline for each sensor placement.

Feature Extraction and Dimensionality Reduction

The preprocessing and feature extraction pipeline from prior work [47], [220] was applied to the collected CSI and RSSI data, yielding 49 features per sample. To assess the inherent separability of the data for each placement without the confounding factor of a classifier's bias, a statistical approach was employed. Principal Component Analysis (PCA) was applied to the 49-dimensional feature space to reduce it to the three most discriminative components for visualization and analysis. The first principal component (PC1), which explains the greatest variance, was selected for statistical hypothesis testing.

5.4.3. Evaluation and Results

The distribution of data points for each activity within the reduced PCA space was visualized using pair plots for each sensor placement (Figure 37, Figure 38, and Figure 39). To quantitatively evaluate the separability, a Kruskal-Wallis H-test was conducted on the values of PC1 to determine if the distributions of the three activities were statistically distinct for each sensor placement.

- Placement 1 (user NLoS, area LoS): The test did not reveal statistically significant differences among the activity groups ($p > 0.08$). The null hypothesis could not be rejected, indicating no significant separabi
- Placement 2 (user LoS, area LoS): A statistically significant difference was found ($\chi^2(2) = 5.18$, p

< 0.08). The null hypothesis was rejected, indicating that the activities were distinguishable in this placement.

- Placement 3 (user NLoS, area NLoS): The test did not reveal statistically significant differences among the activity groups ($p > 0.08$). The null hypothesis could not be rejected, indicating no significant separability.

5.4.4. Conclusion

This study demonstrated that sensor placement has a profound and measurable impact on the performance of a device-free HAR system in a kitchen environment. The results indicate that the optimal configuration is Placement 2, where both the user and the performing area are in the line-of-sight of the transceiver and receiver. This placement yielded the only statistically significant separability between the fine-grained activities. This finding provides a critical, empirically-derived design guideline for deploying device-free HAR systems, establishing that strategic sensor placement is a prerequisite for achieving high recognition accuracy in real-world environments.

Furthermore, this work highlights a critical direction for future research: large-scale, automated sensor placement optimization. It's clear that testing all possible positions in an environment is impractical, but the LoS principle established here can inform the selection of candidate positions. A comprehensive optimization framework could then involve deploying sensors in a larger set of these feasible, practical locations, collecting data for target activities, and comparing system performance to identify the global optimum for a given space and activity set. This would move beyond heuristic placement towards systematic, performance-driven deployment.

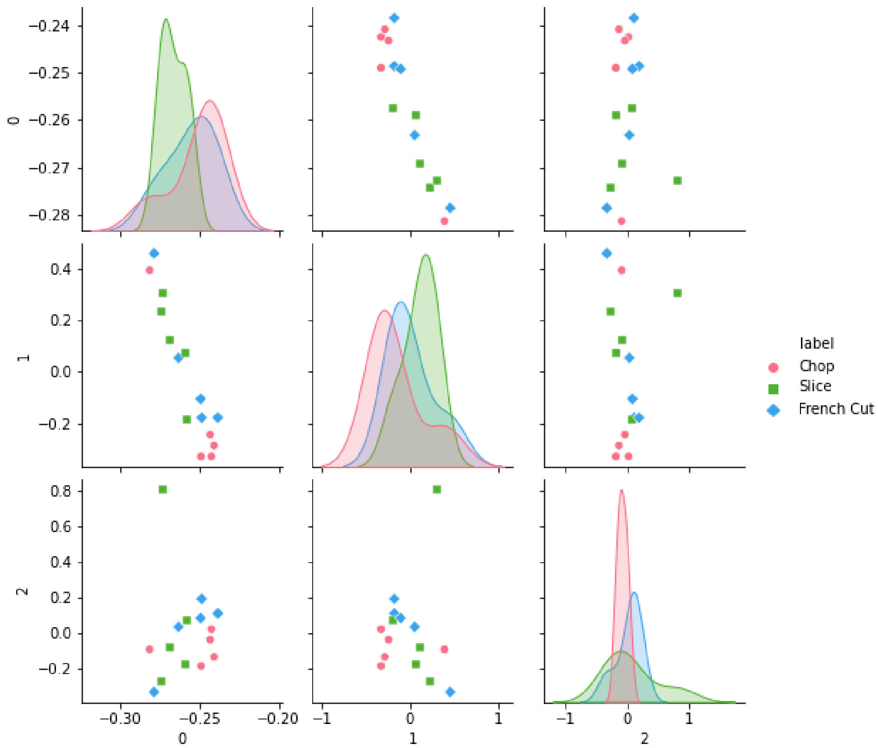

Figure 37 Feasibility study 4 - Iris pair plot for sensor placement 1

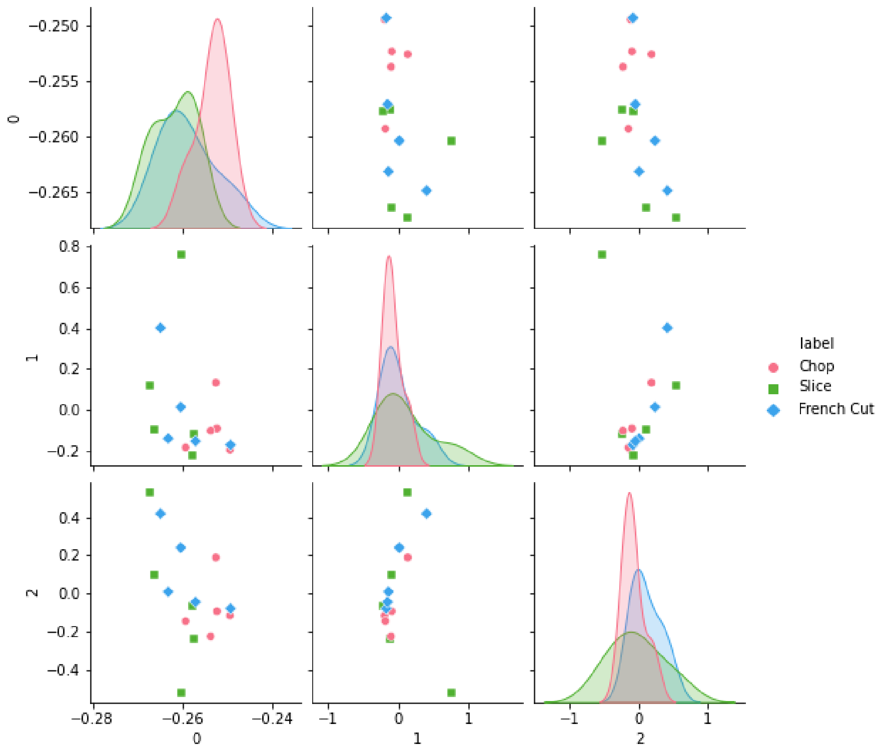

Figure 38 Feasibility study 4 - Iris pair plot for sensor placement 2

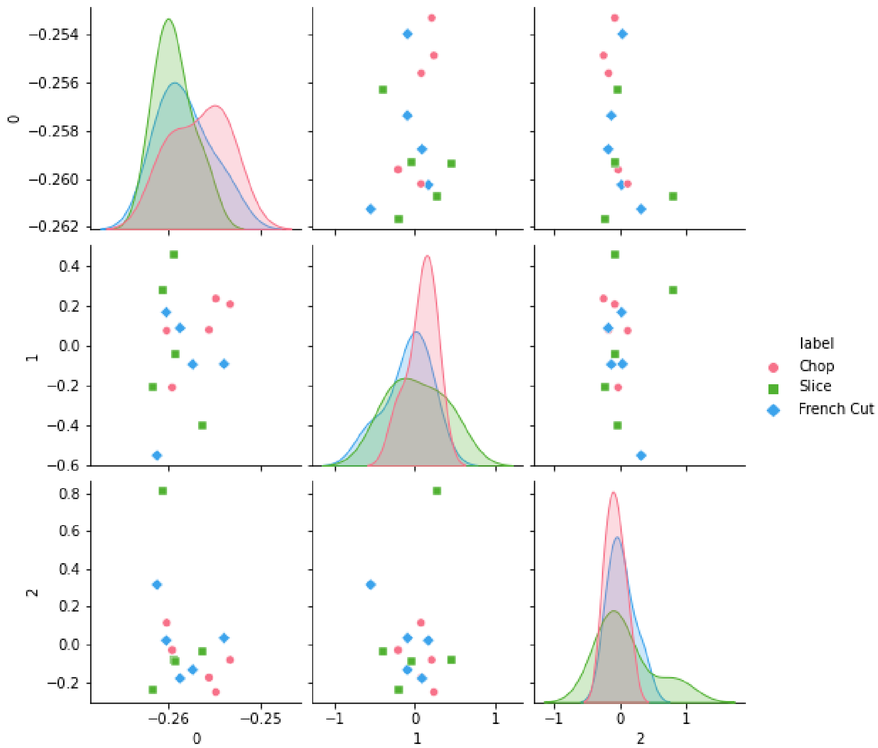

Figure 39 Feasibility study 4 - Iris pair plot for sensor placement 3

Chapter 6 Proposed Method

This chapter presents the proposed method for device-free HAR in kitchen environments, corresponding to Step 4 and Step 5 of the research approach outlined in 4.3.3. The chapter is structured in two main parts. First, it details the inclusive method for collecting a novel, multi-granularity dataset of kitchen activities using WiFi sensing hardware in three distinct real-world kitchens to address the Problem 2 and Problem 3. Second, it introduces a system for evaluating three distinct HAR approaches including deep learning, transfer learning, and classical machine learning, culminating in the development of an optimized model based on a genetic algorithm for MANOVA-driven feature selection to address Problem 4. The performance of this optimized method is then rigorously evaluated through a series of experiments designed to assess its capability in recognizing activities of varying granularity (addressing Problem 1) and its robustness across different environments (addressing Problem 2).

6.1 Dataset Collection

6.1.1. Sensor

The data collection for this research [190] was performed using two ESP-WROOM-32 modules [180]. These boards are equipped with a built-in 2.4 GHz Wi-Fi chip and were configured to function as a transceiver and a receiver. In Access Point (AP) mode, the ESP32 acts as a Wi-Fi hotspot by creating its own wireless network. Other Wi-Fi-enabled devices, such as smartphones, laptops, or another ESP32, can connect directly to this network using the SSID and password defined on the ESP32. In Station (STA) mode, the ESP32 behaves like a regular Wi-Fi client by connecting to an existing Wi-Fi network, such as a home or office router. Once connected, it can communicate with other devices on the same network or access cloud services and the internet.

The selection of the ESP-WROOM-32 module as the sole sensing platform for this research was driven by several pragmatic and methodological criteria critical for foundational research in device-free HAR:

- Ubiquity and Low Cost: The ESP32 is a widely available, low-cost platform, which enhances the reproducibility of this research and its potential for real-world deployment.
- Integrated WiFi Capability: Its integrated 2.4 GHz Wi-Fi transceiver with native support for CSI data extraction provides the essential raw signal for device-free sensing without requiring external,

specialized hardware.

- Programmability and Community Support: The platform is highly programmable via the Arduino framework or ESP-IDF, with extensive community and documentation support, allowing for full control over the data collection protocol.
- Focus on Methodology: Using a single, well-understood sensor type allows the study to isolate and investigate the core challenges of activity recognition, such as feature extraction, granularity, and placement, without the confounding variable of heterogeneous hardware performance.

The AP and STA mode allows the ESP32 to function as both an Access Point and a Station at the same time. This dual-mode capability enables the ESP32 to connect to a router (for internet or network access) while simultaneously hosting its own Wi-Fi network for direct connections from other devices.

To collect this dataset, we configured one ESP32 in Access Point (AP) mode to create a wireless network, and another ESP32 in Station (STA) mode to connect to this network as a Wi-Fi client.

Figure 40 ESP32 microcontroller

6.1.2. Environment

Data was collected in three distinct real-world kitchens (Figure 41, Figure 42, and Figure 43) to ensure the dataset reflects natural scenarios and allows for robustness testing. All objects and appliances remained in their natural positions throughout the experiments. As established in Feasibility Study 3 (section 5.4) [73], the sensors were placed in opposite corners to maintain both the user and the performing area within the Line-of-Sight (LoS) of the transceiver and receiver.

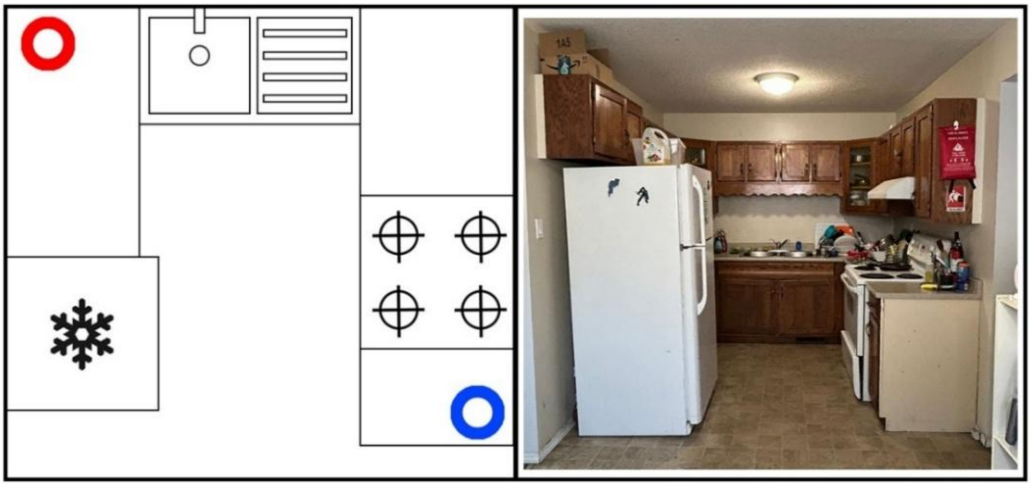

Figure 41 Kitchen 1 – the blue and red circle represent the transceiver and receiver, respectively.

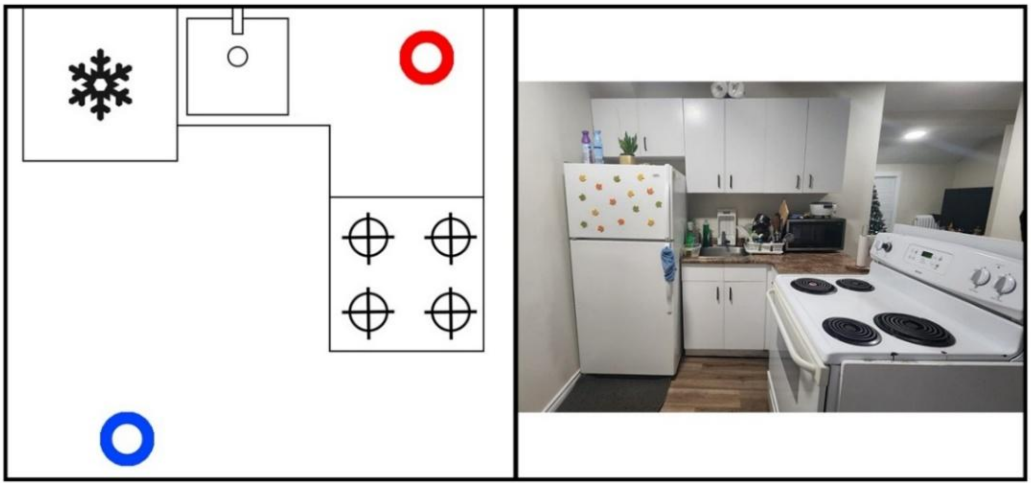

Figure 42 Kitchen 2 – the blue and red circle represent the transceiver and receiver, respectively.

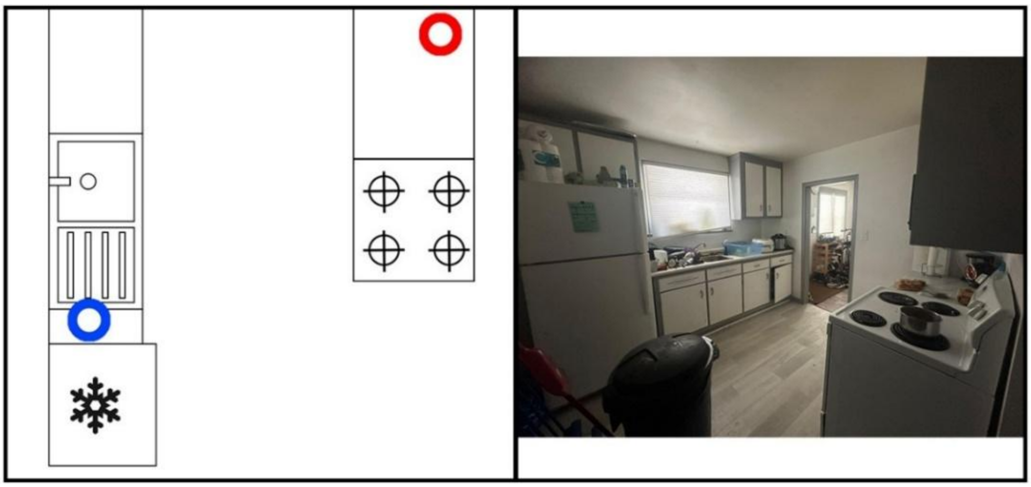

Figure 43 Kitchen 3 – the blue and red circle represent the transceiver and receiver, respectively.

6.1.3. Activity

The set of activities for the dataset was deliberately chosen to address the core research question of how activity granularity affects device-free HAR performance. The selection process followed a structured, scenario-driven methodology:

- Establishing a Real-World Context: A common domestic task, preparing and serving a simple soup, was selected as the overarching scenario. This provided a coherent, logical sequence for the activities.
- Sampling Across Granularity Levels: To create a balanced testbed, we sampled three representative activities from each of three granularity levels (i.e., fine, mid, and coarse) from within the soup-making scenario. This tri-level structure allows for controlled comparisons and isolates the variable of granularity.
- Applying Defined Kinematic Criteria: Activities at each level were chosen based on explicit kinematic and contextual definitions:
    - Fine-grained: Minimal body movement, high similarity between activities (hand/forearm motions).
    - Mid-grained: Moderate movement within a confined space, clear functional differences.
    - Coarse-grained: Significant full-body displacement across the environment, distinct functional goals.

This method resulted in a strategically composed dataset of nine activities, designed not to be an exhaustive list of kitchen tasks, but rather a principled experimental testbed to systematically evaluate the challenges of granularity in device-free sensing.

Fine-grained Activities

Fine-grained activities are those that closely resemble one another and involve minimal movement of body parts during execution. As illustrated in Figure 44, the participants performed the following fine-grained activities: Chop – cutting potatoes into irregular pieces; Slice – cutting potatoes into regular oval pieces; and French Cut – cutting potatoes into regular, thin strips.

Mid-grained Activities

Mid-grained activities are those that differ from one another and involve moderate movement of body parts within a confined space. As illustrated in Figure 45, the participants performed the following mid-grained activities: Pour – transferring water from the kettle to the pot; Stir – mixing the water in the pot by moving a spoon; and Ladle – transferring soup from the pot or cooking vessel to a serving bowl.

Figure 44 Fine-grained activities from left to right: chop, slice, and French cut.

Figure 45 Mid-grained activities from left to right: pour, stir, and ladle.

Figure 46 Coarse-grained activities from left to right: store, wash, and place.

Coarse-grained Activities

Coarse-grained activities are distinct activities that involve significant movement of body parts within a broader area. As illustrated in Figure 46, the participants performed the following coarse-grained activities: Store – storing leftover food into the fridge; Wash – washing the pot; and Place – putting plates into the cabinet.

6.1.4. Participants

To enable future analysis of user-specific effects, data for the same activity set in the same environment was collected from three different participants. However, the primary focus of this dissertation is to establish foundational performance baselines and investigate core challenges such as activity granularity and sensor placement under controlled conditions. Therefore, for the core experiments reported in Chapters 6, the data from all three participants were pooled. The evaluation (e.g., 5-fold cross-validation) was performed on this combined dataset. In this design, data from the same participant can and does appear in both the training and testing splits. This approach is appropriate for this foundational stage, as it provides a stable estimate of the method's performance by maximizing the use of available data, prior to investigating the more complex variable of cross-user generalization in future work.

Figure 47 Participant performing a chopping activity in the presence of the experimenter.

6.1.5. Collection Procedure

As described earlier, each participant was asked to perform every activity in each environment ten times. As visualized in Figure 48, the data collection procedure began with the participant turning on the access point (AP) by pressing a button on the nearby laptop and waiting for five seconds. The experimenter observed the participant and activated the STA as soon as the AP was turned on. The participant then performed the activity as instructed and waited five seconds after completing it. Finally, the participant turned off the AP and the experimenter turned off the STA. After a short processing period, a `.txt` file was saved on both laptops connected to the AP and STA.

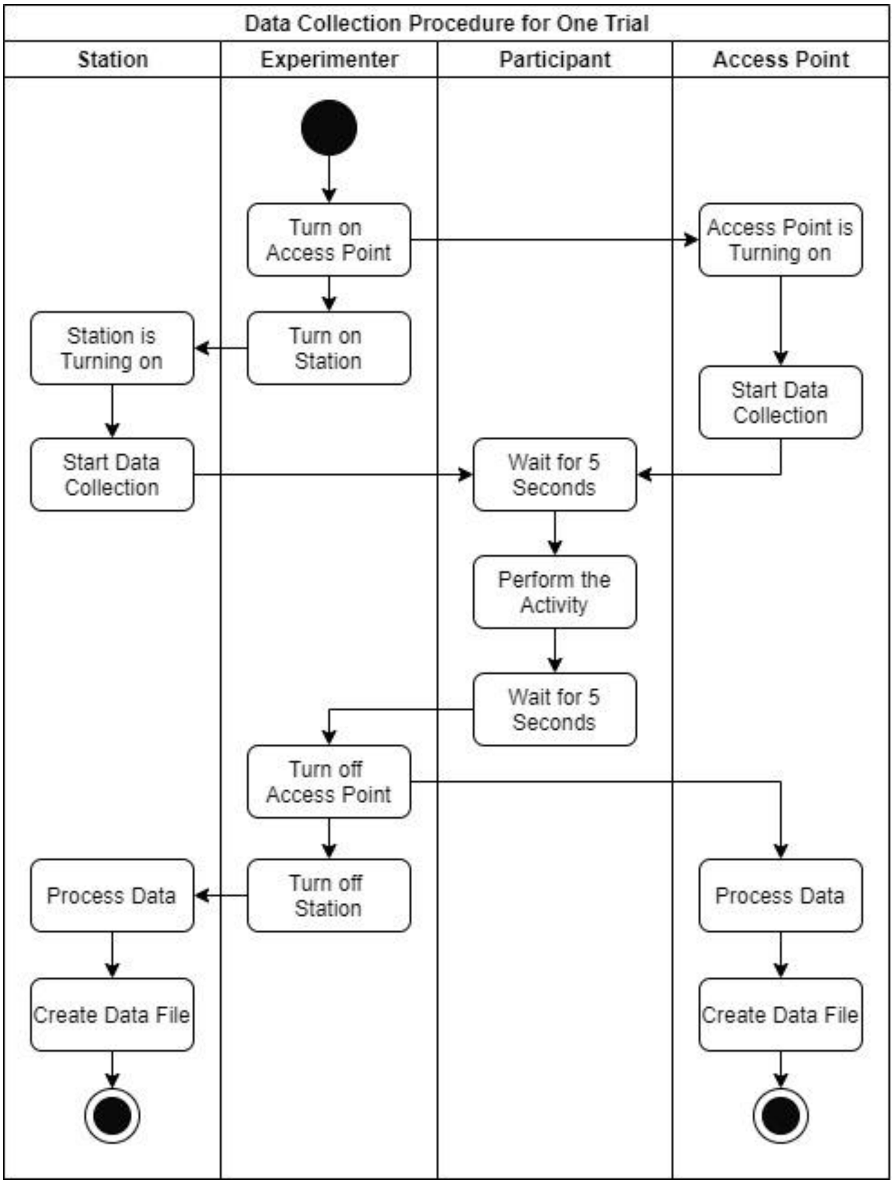

Figure 48 Data collection procedure – a sequence diagram

6.1.6. Data Acquisition

We used the ESP32-IDF v3.3.1 firmware [181] installed on both the STA and AP nodes to collect data. This firmware enables the ESP32 to actively capture Wi-Fi packets. It intercepts each received packet at the physical layer, extracts CSI and RSSI before the data reaches higher network layers (e.g., IP or TCP), and transmits the extracted information to a connected PC via the Universal Asynchronous Receiver/Transmitter (UART) serial interface in real time.

6.1.7. Wi-Fi Channel and Subcarrier Structure

The ESP32 sensors operated on the 2.4 GHz Wi-Fi band using the IEEE 802.11b/g/n standard with a 20 MHz channel bandwidth, a high serial baud rate of 921600, and a data packet rate of 100. In this configuration, the 20 MHz bandwidth is divided into 128 discrete frequency bins, known as subcarriers, using a 64-point Inverse Fast Fourier Transform (IFFT). This process is fundamental to Orthogonal Frequency Division Multiplexing (OFDM), a modulation technique used in 802.11g and 802.11n (but not in 802.11b). OFDM improves spectral efficiency and robustness to multipath fading by transmitting data in parallel over multiple orthogonal subcarriers. Among these 64 subcarriers, 52 are actively used: 48 for data transmission and 4 as pilot subcarriers for synchronization and channel estimation. The remaining subcarriers serve as null subcarriers, some placed at the edges to act as guard bands and others (such as the central DC subcarrier) to mitigate interference around the carrier frequency. This design enables reliable and efficient wireless communication in noisy and dynamic environments like indoor settings.

These 128 bins represent the division of a 20 MHz Wi-Fi channel into 64 subcarriers using 64 inverse fast Fourier transform (IFFT), each subcarrier spaced 312.5 kHz apart. The subcarriers are indexed from −32 to +31 and are symmetrically centered around the channel's center frequency (Fc). Since we are using Wi-Fi channel 11 in the 2.4 GHz band, the center frequency Fc is 2462 MHz. The frequency of each subcarrier is calculated using the formula:

$$f_n = Fc + (n \times 312.5\, kHz)$$

Where $n \in [-32, 31]$ is the subcarrier index and f_n is the frequency of subcarrier n. For instance, the frequency of subcarrier -32 is

$$f_{-32} = 2462\ \mathrm{MHz} + (-32 \times 312.5\, kHz) = 2452\, MHz$$

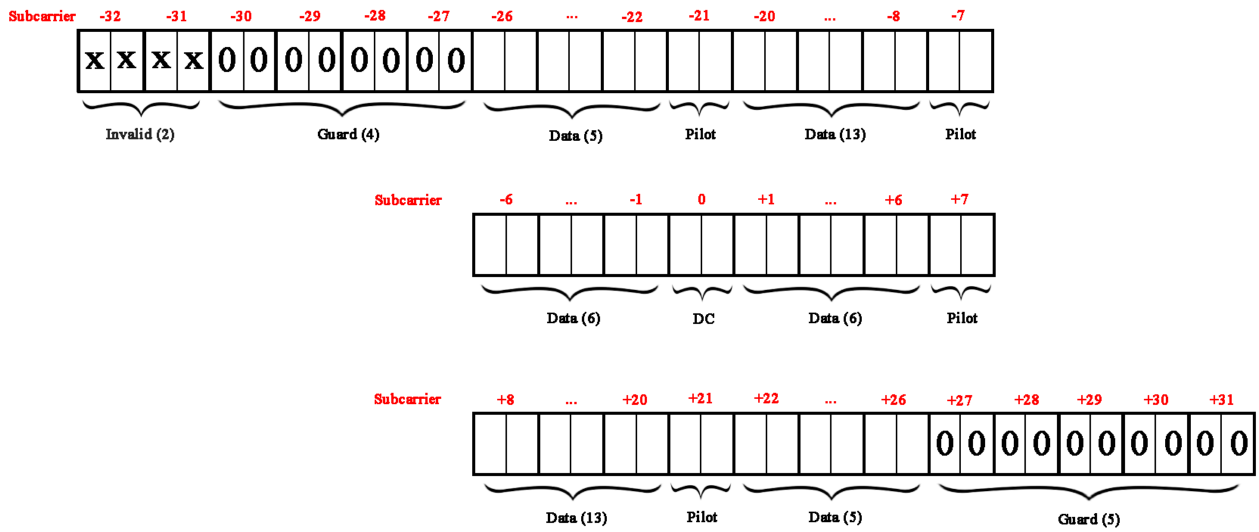

Figure 49 Structure of CSI including data, pilot, and guard subcarriers

Figure 49 illustrates the structure of CSI subcarriers extracted by ESP32-IDF v3.3.1 firmware [181]. The first four bytes of CSI data are invalid due to a hardware limitation in ESP32. The next eight bytes, like the last ten, represent guard subcarriers at the edges of the Wi-Fi channel (e.g., 2.412 GHz, 2.417 GHz) to minimize spectral leakage [182]. In addition, this structure defines four pilot subcarriers at position -21, -7, +7, and +21 to help with channel estimation, phase tracking, and frequency offset correction. A DC subcarrier also defined at subcarrier 0 to prevent interference at the center frequency. In overall, 48 subcarriers are available and contains the actual CSI data. Position of these subcarriers is illustrated in Figure 49. As you can see, each subcarrier is a complex number including imaginary and real part. The first number is the imaginary part, and the second one is the real part. Algorithm 3 outlines the procedure for parsing CSI data.

Algorithm 3 Parse CSI

Input: complex numbers of CSI subcarriers

Output: amplitudes and phases

$rawCSI$ = CSI complex numbers

$imaginary, real, amplitude, phase = []$

for i in $range(rawCSI)$:

    if $i \% 2 == 0$:

        $imaginary.add(rawCSI[i])$

    else:

        $real$.add($rawCSI$[i])

for i in $range(rawCSI)/2$:

    $amplitude.add\left(\sqrt{imaginary^2[i] + real^2[i]}\right)$

    $phase.add(arc\ tangent(imaginary[i], real[i]))$

return $amplitude, phase$

6.1.8. Records and Storage

The provided dataset contains one .txt file for each trial. The .txt file includes several parts: the initial system boot messages showing firmware version, memory, and partition setup; Wi-Fi configuration details like SSID, channel, and MAC address; and most importantly, CSI data packets. Each CSI data packet contains a timestamp, metadata (such as RSSI, noise floor, rate, bandwidth, modulation coding scheme, and guard interval), and a CSI matrix. Table 8 and A structured naming convention was used for all dataset files, as defined in Table 10. P_E_S_A _T where P describes the participant that performs the activity, E specifies the performing environment, S specifies the sensors, A shows the performed activity, and T specifies 10 trials for each set of user, activity, environment, sensor data. These files are provided in a folder named "All in One" that contains all 1620 files with explained naming format. This dataset is also presented in a hierarchical approach illustrates in Figure 50.

$P \in \{p1, p2, p3\}$

$$E \in \{e1, e2, e3\}$$

$$S \in \{s1, s2\}$$

$$A \in \{a1, a2, a3, a4, a5, a6, a7, a8, a9\}$$

$$T \in \{t1, t2, t3, t4, t5, t6, t7, t8, t9, t10\}$$

Table 9 describe the format of the collected data and a sample data entry.

Table 8 CSI entry format with sample data

Data Format	CSI_DATA, <device role>, <MAC>, <RSSI>, <channel>, <secondary channel>, <bssid>, <antenna>, <aggregation>, <stbc>, <fec_coding>, <sgi>, <noise_floor>, <rate>, <sig_mode>, <mcs>, <bandwidth>, <smoothing>, <not_sounding>, <aggregation_flag>, <space_time_blocking>, <timestamp>, <csi_len>, <payload_len>, <rx_state>, <real_time>, <CSI values count>, [<CSI values>]
Data Sample	CSI_DATA,AP,48:E7:29:A0:68:70,-59, 11, 0, 0, 0, 0, 0, 0, 0, 0, 0, -97, 0, 6, 1, 1147185, 0, 37, 0, 0, 1.38942, 128, [37 80 2 0 0 0 0 0 0 0 0 0 17 -11 17 -11 18 -12 19 -11 18 -9 17 -8 17 -9 16 -7 18 -8 17 -8 19 -8 19 -8 18 -6 18 -6 16 -7 19 -7 17 -7 16 -6 15 -6 16 -6 15 -6 16 -6 15 -5 16 -6 17 -7 16 -6 0 0 15 -8 16 -8 15 -8 14 -9 14 -8 14 -9 14 -9 13 -8 13 -10 15 -6 15 -9 13 -8 14 -11 14 -10 14 -10 15 -10 14 -11 13 -10 13 -10 12 -11 13 -11 12 -12 13 -12 13 -12 12 -13 12 -13 0 0 0 0 0 0 0 0 0 0]

A structured naming convention was used for all dataset files, as defined in Table 10. P_E_S_A _T where

P describes the participant that performs the activity, E specifies the performing environment, S specifies the sensors, A shows the performed activity, and T specifies 10 trials for each set of user, activity, environment, sensor data. These files are provided in a folder named “All in One” that contains all 1620 files with explained naming format. This dataset is also presented in a hierarchical approach illustrates in Figure 50.

$$P \in \{p1, p2, p3\}$$

$$E \in \{e1, e2, e3\}$$

$$S \in \{s1, s2\}$$

$$A \in \{a1, a2, a3, a4, a5, a6, a7, a8, a9\}$$

$$T \in \{t1, t2, t3, t4, t5, t6, t7, t8, t9, t10\}$$

Table 9 CSI data packet field definitions

CSI_DATA	A static identifier that denotes the start of a CSI data record.
<device role>	Role of the ESP32 device (e.g., AP for Access Point, STA for Station).
<MAC>	MAC address of the sender or receiver, depending on the role (usually the AP's MAC).
<RSSI>	Received Signal Strength Indicator, measured in dBm (e.g., -59).
<channel>	WiFi channel number used for transmission (e.g., 11).
<secondary channel>	Offset for secondary channel in HT40 mode: 0: HT20, 1: secondary channel above, 2: secondary channel below
<bssid>	The BSSID field is usually placeholder or 0, since it's not extracted by default.
<antenna>	Antenna index used to receive the packet (ESP32 typically uses only 1 antenna, so often 0).
<aggregation>	Whether packet was aggregated: 0: no aggregation, 1: part of an aggregated packet
<stbc>	Space-Time Block Coding flag (diversity technique): 0: not used, 1: used
<fec_coding>	Forward Error Correction status: 0: off, 1: on (LDPC used)
<sgi>	Short Guard Interval flag: 0: not used (standard GI), 1: used (improves throughput)
<noise_floor>	Estimated background noise level in dBm (e.g., -97).
<rate>	Raw data rate of the received packet (in Mbps or a modulation index, depending on mode).
<sig_mode>	Signal mode: 0: legacy (non-HT), 1: high throughput, 802.11n (HT)
<mcs>	Modulation and Coding Scheme index (defines modulation type and data rate).
<bandwidth>	Channel bandwidth used: 0: 20 MHz, 1: 40 MHz

<smoothing>	Smoothing flag (CSI processing detail): 0: off, 1: on
<not_sounding>	Indicates if the packet is a sounding packet: 0: sounding 1: not sounding
<aggregation_flag>	Another flag for MPDU aggregation status (redundant with field 9).
<space_time_blocking>	Same as STBC flag (possibly redundant); kept for completeness.
<timestamp>	Internal timestamp when the CSI data was recorded (e.g., 1147185).
<csi_len>	Length of the CSI data array in bytes (e.g., 0 if empty, otherwise non-zero).
<payload_len>	Length of the WiFi packet payload (not always relevant for CSI use).
<rx_state>	Internal status flag of the received packet.
<real_time>	Timestamp in seconds with fractional part (e.g., 1.38942); useful for synchronization.
<CSI values count>	Number of CSI values reported (e.g., 128).
[<CSI values>]	The CSI amplitude and phase values in I/Q (In-phase and Quadrature) pairs.

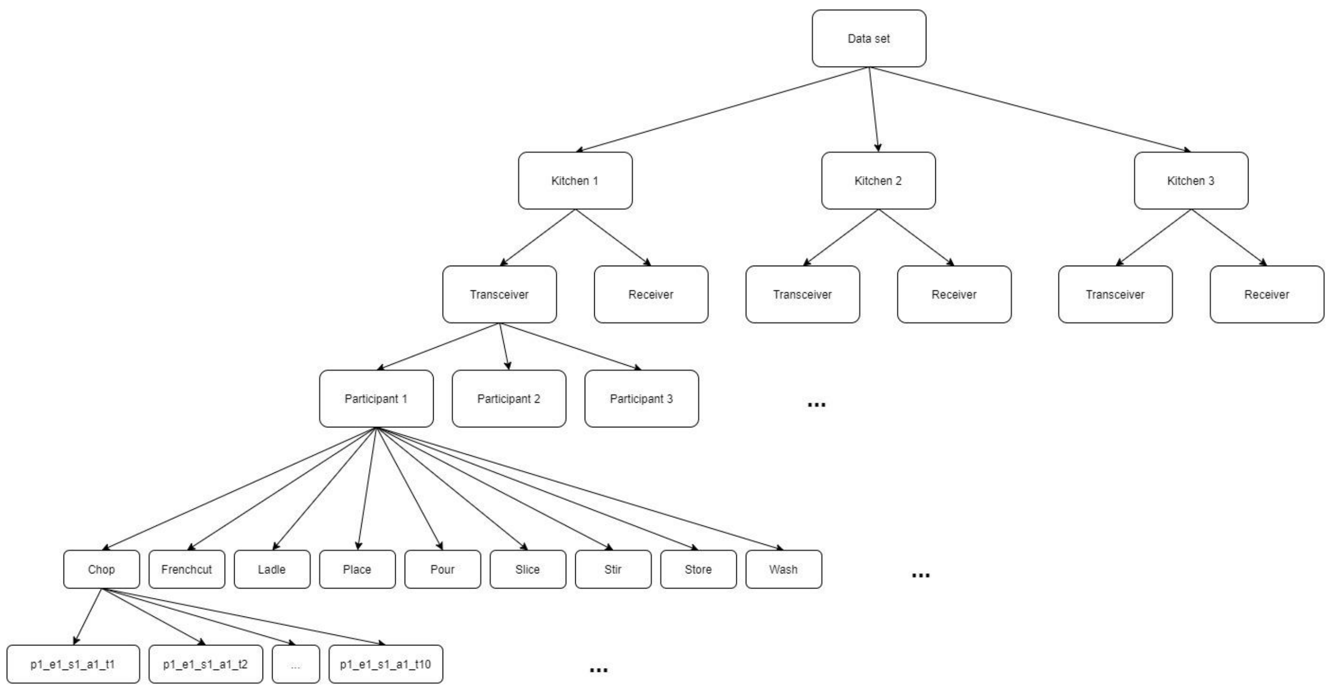

Figure 50 Dataset file hierarchy overview

Table 10 Dataset code reference

Category	Code	Meaning
Participant	$p1$	Participant 1
	$p2$	Participant 2
	$p3$	Participant 3
Environment	$e1$	Kitchen 1
	$e2$	Kitchen 2

	$e3$	Kitchen 3
Device Role	$s1$	AP
	$s2$	STA
Activity	$a1$	Chop
	$a2$	Slice
	$a3$	French Cut
	$a4$	Pour
	$a5$	Stir
	$a6$	Ladle
	$a7$	Store
	$a8$	Wash
	$a9$	Place
Trial Number	$t1$	Trial 1
	...	…
	$t10$	Trial 10

6.2 Model Structure

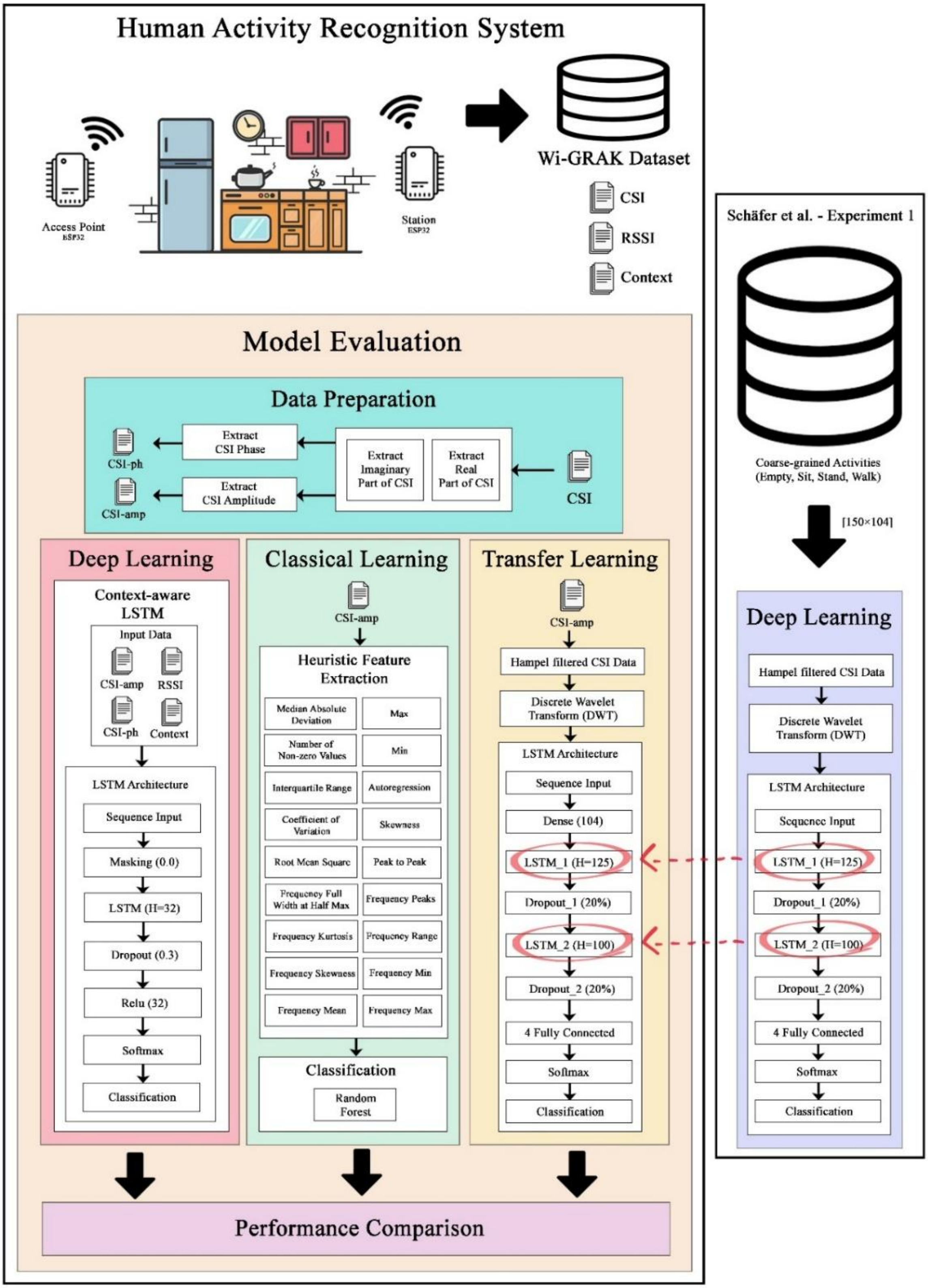

Figure 51 Model evaluation method

As shown in Figure 51, this dissertation proposes a method for human activity recognition (HAR) centered on a comparative model evaluation approach [177]. Data were first collected using an HAR system with ESP32 sensors installed in real-world kitchens. From this data, the phase and amplitude of CSI were extracted and, along with RSSI, used in three distinct recognition approaches: deep learning, transfer learning, and classical machine learning. The performance of these methods was then evaluated, and the best-performing approach was selected for further refinement in the model optimization step.

6.2.1. Data Preparation

The proposed method operates on CSI data that captures signal measurements across various human activities and environmental settings. As shown in Figure 51, the process begins by extracting the real, imaginary, amplitude, and phase components from the complex CSI data. CSI data is typically represented as a sequence of complex numbers:

$$CSI = [H_1, H_2, \dots, H_n] \quad where \quad H_i \in C$$

$$H_i = a + bj$$

where a and b are the real and imaginary parts, respectively. The amplitude A_i of each CSI value, which indicates signal strength, is calculated using Equation 1. The phase φ_i, which provides information about the signal's propagation characteristics, is computed using Equation 2. These two components are critical for extracting meaningful features from raw wireless signals for activity recognition.

$$A_i = |H_i| = \sqrt{a^2 + \mathrm{b}^2} \qquad (1)$$

$$\emptyset_i = \arg(\mathrm{H_i}) = arctan2(b, a) \qquad (2)$$

6.2.2. Deep Learning

Due to the limited availability of large datasets in the device-free kitchen context, training deep learning models like standard Long Short-Term Memory Models (LSTMs) for kitchen activity recognition is prone to overfitting and poor generalization [183]. To address this challenge, a Context-Aware LSTM [184] offer a promising solution. By integrating contextual information [185], such as the environment, sensor, or user, into the learning process, the model can better adapt to domain differences without requiring massive training data. This additional context helps the LSTM disambiguate similar patterns across domains and improves recognition accuracy by compensating for the lack of generalization that a large dataset would otherwise provide. In summary, because the dataset size in the device-free kitchen context is limited and cross-domain variation exists, using a context-aware LSTM may improve performance and robustness in recognizing kitchen activities.

For the model input, the amplitude and phase from only one CSI subcarrier were selected, as subcarriers typically exhibit high correlation. Although including more subcarriers could provide additional detail, it would significantly increase the input size and, consequently, the number of trainable parameters, which could lead to overfitting due to the lack of a large dataset. In addition to CSI features, RSSI values and contextual information were included. Each time step in the input sequence consists of 11 features: two CSI values (amplitude and phase), one RSSI value, and eight binary indicators representing the environment, participant, and sensor type. The process pipeline for each trial involved mapping categorical context attributes into one-hot vectors and combining them with signal features to create fixed-length input vectors. A sequential LSTM architecture was defined, beginning with a Masking layer to ignore padded values, followed by a 32-unit LSTM layer for learning temporal dependencies. A Dropout layer mitigates overfitting, and two Dense layers with ReLU and SoftMax activations were used to perform final classification of kitchen activities.

6.2.3. Transfer Learning

Transfer learning is a machine learning technique that can be applied to human activity recognition (HAR). In this approach, a Convolutional Neural Network (CNN) model trained on one set of activities is reused as the starting point for recognizing a second, related set of activities [186]. In this module, a deep network classifier from [187] was first replicated. This model was originally designed for recognizing coarse-grained activities such as walking, standing, sitting, and an empty room, and trained on a large dataset. The trained variables from two layers of this network are then used as the initial values for the corresponding layers in a similar deep network aimed at recognizing kitchen activities.

In [187], Schäfer et al. proposed a CSI-based HAR method to recognize activities such as walking, standing, sitting, and detecting an empty room. They gathered CSI data using the Nexmon tool and denoised the data with a Hampel filter followed by a discrete wavelet transform (DWT) [188]. The denoised data was then input to an LSTM-CNN classifier, as specified in the Deep Learning module in Figure 51. In an experiment referred to as Experiment 1.2 [187], they trained their model on a dataset collected from one user performing the activities of sitting, standing, and walking, including an empty room condition as a fourth activity. The dataset [187], consisting of 1,800 entries, was used to train the deep neural network. Although they did not report the method's overall accuracy, they achieved a precision of 97% in this experiment.

In this research, the dataset presented in [187] was utilized and MATLAB code was implemented to successfully reproduce the results. The parameter weights of two LSTM layers were then extracted into .mat files, which were subsequently imported into Python code to initialize the weights for the LSTM layers in the proposed LSTM-CNN method. A similar LSTM-CNN architecture was implemented in Python, initializing the LSTM layers with the learned parameters.

In the Transfer Learning Module (Figure 51), we applied a Hampel filter to denoise the CSI data. Then, we employed DWT to denoise the data (see Figure 52). A similar LSTM-CNN architecture was then implemented in Python, with the LSTM layers initialized using the learned parameters imported from the source model. As shown in Figure 51, the main difference between our LSTM-CNN architecture and that in reference [187] is the addition of a Dense layer at the beginning, which adjusts the input dimensions to 104 to ensure compatibility with the source model's expected input size. The final layer before the SoftMax was also modified; instead of four fully connected units (for the four source activities), n fully connected units were used, where n represents the number of target kitchen activities to be recognized.

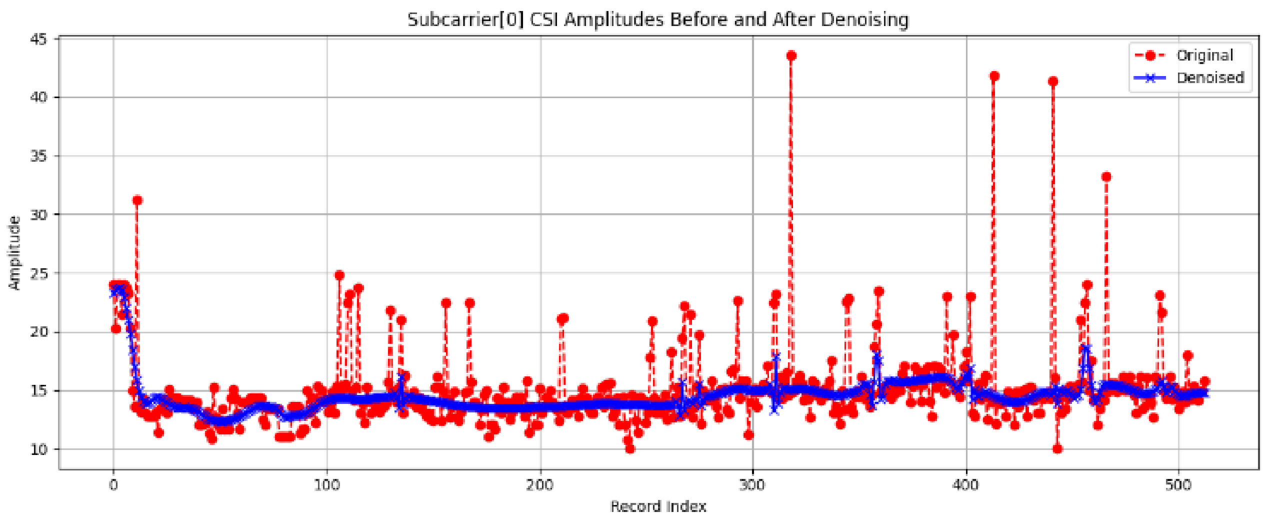

Figure 52 A sample subcarrier[0] CSI before and after denoising with Hampel filter

6.2.4. Classical Learning

As illustrated in Figure 51, for the classical learning approach, the amplitude of the CSI signal was used as input. A set of features was heuristically selected from the amplitude data [107], including median absolute deviation, maximum, minimum, interquartile range, root mean square, coefficient of variation, skewness, peak-to-peak, autoregression frequency, maximum frequency, minimum frequency, mean frequency, full width at half maximum frequency, peak frequency range, frequency skewness, and frequency kurtosis. These extracted features were then used to train a Random Forest classifier to recognize kitchen activities

6.2.5. Preliminary Comparative Evaluation

To evaluate the performance of the three proposed approaches, a series of preliminary experiments was conducted. The objective was to assess the comparative effectiveness of the deep learning, transfer learning, and classical machine learning methods in recognizing kitchen activities.

For this initial evaluation, the dataset recently gathered for this work [189] was utilized. This comprehensive dataset includes data from a transmitter and a receiver, encompassing nine kitchen activities (three fine-grained, three mid-grained, and three coarse-grained), performed by three different users across three distinct environments, with 10 trials conducted for each activity. Each data entry in the dataset comprises a sequence of data records, where each record contains CSI information for 64 subcarriers, of which 48 are data subcarriers represented by complex numbers [190].

To establish a strong and clear performance baseline for this comparative analysis, the evaluation focused only on the three coarse-grained kitchen activities (storing food, washing the pot, and putting plates away). This focus was chosen because coarse-grained activities are generally easier to recognize [107]. This strategy provides a controlled context to compare the fundamental capability of each approach on a simpler task before progressing to the more complex challenge of distinguishing fine-grained activities.

6.3 Pre-Experimental Results and Analysis

The performance of the three proposed methods—deep learning, transfer learning, and classical machine learning—was evaluated in a series of separate pre-experiments. The results demonstrated that the classical learning method (Random Forest) achieved the highest accuracy, significantly outperforming the deep learning and transfer learning approaches in recognizing the coarse-grained kitchen activities. The following subsections detail the setup and results for each pre-experiment.

6.3.1. Pre-Experiment 1: Deep Learning (Context-Aware LSTM)

The deep learning module was tested using all available data samples to maximize input data size. The input data comprised recordings from two sensors, with three users each performing three activities in three environments, for 10 trials each, resulting in 540 data entries ($2 \times 3 \times 3 \times 3 \times 10$).

To manage dimensionality, each data record was reduced to an 11-dimensional feature vector per time step, containing:

- The amplitude ($A_{0,t}$) and phase ($\emptyset_{0,t}$) of the first CSI subcarrier.
- The RSSI value ($RSSI_t$).
- An 8-element one-hot encoded vector representing context (3 environments, 3 participants, 2 sensors).

Variable-length sequences were zero-padded to a consistent length for input into the LSTM model. The architecture of the proposed context-aware LSTM is detailed in Table 11.

Table 11 Context-aware LSTM model architecture

Layer (type)	Output Shape	Param #
Masking (Masking)	(None, None, 11)	0
LSTM (LSTM)	(None, 32)	5,632
Dropout (Dropout)	(None, 32)	0
Dense (Dense)	(None, 32)	1,056
Dense (Dense)	(None, 3)	99

Results: The model significantly underperformed, with training and validation accuracies remaining between 28% and 43% over 20 epochs. The loss values stagnated around 1.09–1.11, indicating a failure to learn or converge. This underfitting is attributed to the fundamental limitation of the dataset size (270 unique activities); deep learning models typically require substantially larger amounts of data to learn complex temporal patterns effectively, which was not available in this context.

6.3.2. Pre-Experiment 2: Transfer Learning (LSTM-CNN)

This experiment utilized data from only one environment and one sensor (the transceiver unit) to isolate the effect of activity type, excluding variables like environmental layout. The dataset included three users performing three activities with 10 trials each, resulting in 90 data samples.

A pre-trained LSTM-CNN model from [187], originally trained on activities like walking and sitting, was fine-tuned on this new dataset. The goal was to adapt the model's pre-learned temporal representations to the kitchen activity domain.

Results: The transfer learning approach also demonstrated poor performance, with accuracy stagnating between 26% and 39% over the training epochs. The loss failed to decrease consistently. This indicates a fundamental domain mismatch. The pre-trained knowledge from gross motor movements in one environment did not transfer effectively to the subtle, object-interactive actions in a kitchen setting. WiFi CSI signals are highly sensitive to specific environmental geometries and activity characteristics, limiting the viability of transfer learning for this specific cross-domain application.

6.3.3. Pre-Experiment 3: Classical Machine Learning (Random Forest)

This experiment also used a controlled setting (Environment 1) but included data from both sensors, resulting in 180 data samples (3 users × 3 activities × 2 sensors × 10 trials).

18 heuristic features (e.g., median absolute deviation, interquartile range, spectral kurtosis) were extracted from the amplitude of the first CSI subcarrier for each data sample. A Random Forest classifier (100 trees) was trained on 80% of this data, with feature values standardized using a StandardScaler fit on the training fold.

Results: Evaluated via 5-fold cross-validation, the Random Forest classifier achieved an average accuracy of 68.3%. Performance varied by activity: the "wash" activity (a8) achieved an F1-score of 0.76, while "store" (a7) and "place" (a9) achieved F1-scores of 0.66 and 0.63, respectively. The confusion matrix for this experiment is presented in Figure 53. These results demonstrate that the classical learning approach, with its robust feature-based design, significantly outperformed the deep learning methods for this specific task and dataset, providing a viable baseline for recognizing coarse-grained activities.

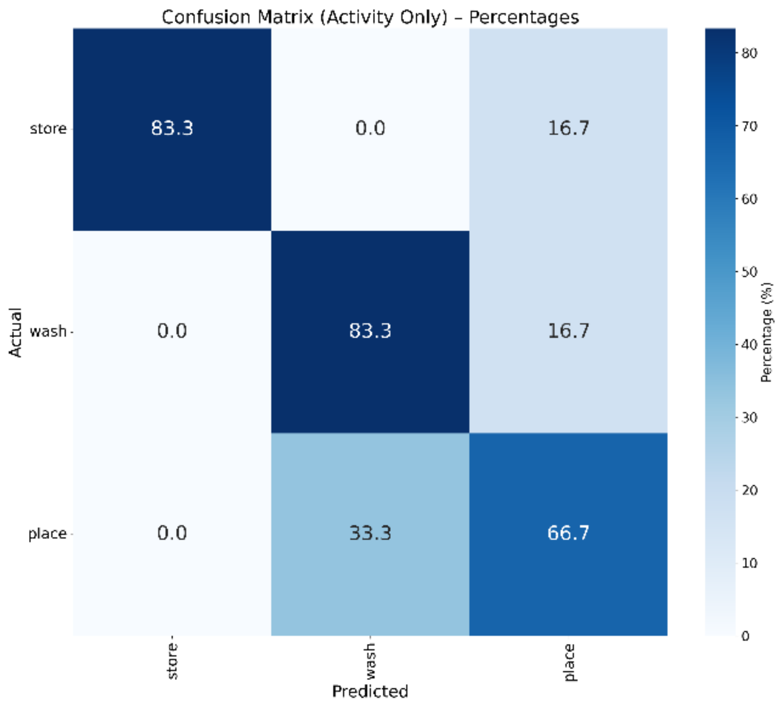

Figure 53 Confusion matrix for the classical learning method (Random Forest) recognizing coarse-grained activities.

6.3.4. Conclusion of Pre-Experiments

The pre-experiments clearly demonstrate that for the available dataset size and specific task of recognizing coarse-grained kitchen activities, a classical machine learning approach (Random Forest) is the most effective strategy. The deep learning models were hampered by data scarcity and domain mismatch issues that could not be overcome within the scope of these pre-experiments. Consequently, the classical learning approach was selected for further optimization and evaluation on a more extensive set of activities in the main experiments (detailed in Section 6.4)

6.4 Learning Optimization

Based on the results of the pre-experiments (Section 6.3), the classical machine learning approach was selected for further optimization to improve performance across a broader range of activities. The optimization strategy targeted enhancements at both the feature extraction and classification stages. This involved extracting a comprehensive set of features reported in the literature and developing a novel method to identify an optimal feature subset. Furthermore, rather than relying on a single classifier, multiple classifiers were evaluated to select the one that delivered the highest performance.

Recent work by Nguyen et al. [191] demonstrates the effectiveness of MANOVA for real-time malware detection through feature selection [192], while ANOVA-based approaches have shown promise for camera-based HAR systems [193]. In device-based HAR, researchers have developed optimal feature selection methods for recognizing long-term activities through atomic activity decomposition [194]. However, feature selection for device-free, WiFi-based HAR remains a developing area of research [211][212]. The proven efficacy of statistical parametric tests like MANOVA and ANOVA in other domains provided a foundational justification for proposing a MANOVA-based feature selection approach for device-free HAR using WiFi signals.

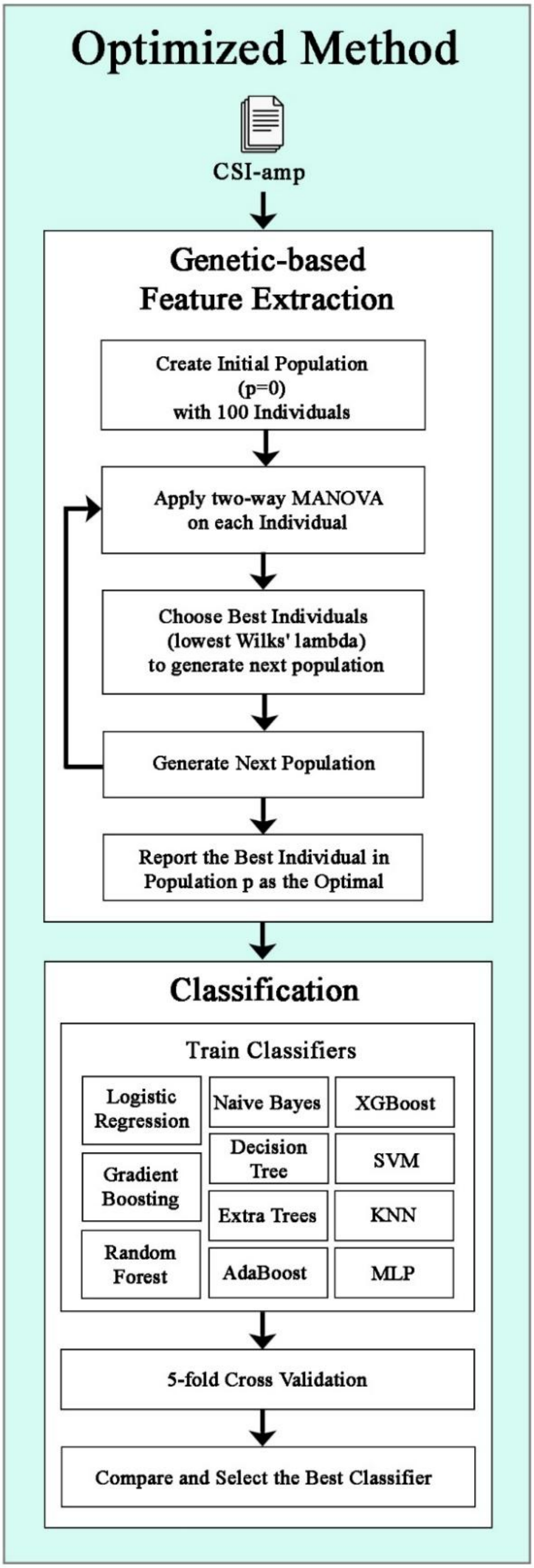

Figure 54 Optimized feature selection for classical learning methods

The optimized model (Figure 54) utilizes a genetic algorithm guided by MANOVA to select an optimal subset of features from WiFi CSI data for kitchen activity recognition. First, CSI amplitude is extracted. Then, a set of statistical and signal-domain features is computed. A genetic algorithm is employed to explore combinations of these features, evaluating each subset using Wilks' lambda from a MANOVA test with “Activity” as the factor. This approach is designed to select features that best discriminate between different activities. The selected features are then used to train multiple traditional classifiers, and the classifier with the highest accuracy on the test data is selected.

Table 12 Genetic algorithm hyperparameters

Parameter	Value	Description
Population size (n)	100	Number of individuals in each generation
Number of generations (ngen)	40	Number of generations (iterations) to evolve the population
Crossover probability (cxpb)	0.5	Probability of mating two individuals (crossover)
Mutation probability (mutpb)	0.2	Probability of mutating an individual
Bit mutation probability (indpb)	0.05	Probability of flipping each gene (bit) during mutation
Selection method	Tournament	Method for selecting individuals for the next generation
Tournament size (tournsize)	3	Number of individuals competing in each tournament selection
Initial gene activation rate	0.6	Probability a feature is selected (1) in the initial individual

6.4.1. Genetic-based Feature Extraction

To identify the most effective feature set for activity recognition in the dataset, the algorithm leverages a list of candidate features commonly employed in device-free HAR methods [32]. The initial population of the genetic algorithm consists of 100 individuals, where each individual represents a unique subset of features. A MANOVA is applied to each individual to evaluate the separability of activities based on the selected features. Individuals yielding the lowest Wilks’ lambda values, indicating greater group separability, are considered more fit and are more likely to be selected for reproduction in subsequent generations. After 40 generations, the best individual of the final population is identified as the optimal feature set that provides the highest discriminatory power for recognizing activities. This optimal set is then used to train a machine learning classifier for final activity recognition. Table 12 details the hyperparameters of the implemented genetic algorithm.

6.4.2. Multiple Classifier Classification

To evaluate the effectiveness of various machine learning models in recognizing kitchen activities from WiFi-based features, a diverse set of traditional classifiers was employed. Each model was trained using a standardized pipeline that included feature scaling (StandardScaler) followed by model fitting. This multi-classifier approach was adopted to identify the most suitable algorithm for the task, as different models have varying inductive biases and strengths depending on data characteristics.

The evaluation utilized stratified 5-fold cross-validation to ensure balanced class representation across all folds. Accuracy scores were averaged over the five folds to obtain a reliable estimate of each model's performance. The classifier with the highest mean accuracy was selected as the optimal model.

Table 13 Description of the classifiers used in optimized method

Classifier Name	Python Library	Key Hyperparameters & Values
Logistic Regression	sklearn.linear_model	max_iter=1000
SVM	sklearn.svm	Default kernel (RBF), default regularization C=1.0
KNN	sklearn.neighbors	n_neighbors=5 (default), uses Euclidean distance
Decision Tree	sklearn.tree	Default parameters, uses Gini index for splitting
Naive Bayes	sklearn.naive_bayes	GaussianNB with default settings
Gradient Boosting	sklearn.ensemble	n_estimators=100, learning_rate=0.1, max_depth=3
AdaBoost	sklearn.ensemble	n_estimators=50, base estimator is a decision stump
Extra Trees	sklearn.ensemble	n_estimators=100, criterion='gini'
Random Forest	sklearn.ensemble	n_estimators=100, random_state=42
MLP	sklearn.neural_network	max_iter=1000, default architecture (one hidden layer of 100 units)
XGBoost	xgboost	use_label_encoder=False, eval_metric='mlogloss'

Each classifier evaluated in this work was selected for its unique strengths in handling different data patterns and structures. Logistic Regression and Support Vector Machines (SVM) were chosen as effective linear models that perform well when the data is linearly separable. K-Nearest Neighbors (KNN) excels at capturing local patterns and non-linear relationships by relying on the proximity of data points. Tree-based models such as Decision Trees, Random Forests, and Extra Trees were included because they can model complex interactions between features and can handle non-linear decision boundaries. Ensemble methods like Gradient Boosting and AdaBoost were selected as they combine multiple weak learners to create a stronger, more robust model, often leading to improved accuracy. The Multi-Layer Perceptron (MLP), a type of neural network that learns non-linear mappings through layered transformations, was also evaluated. Finally, XGBoost, a highly optimized gradient boosting library, was included because it is well-suited for structured data and often outperforms traditional models in classification tasks. This diverse set ensures a comprehensive evaluation framework and increases the likelihood of identifying the most suitable model for the complex task of device-free human activity recognition. Table 13 provides a description of the classifiers used, along with their key hyperparameters.

6.4.3. Experiment 1 – Optimized Model to Recognize Coarse-grained Kitchen Activities

Experiment 1 aimed to classify coarse-grained kitchen activities using the proposed optimized model and to compare its performance against the baseline heuristic model from Pre-experiment 3. To ensure a direct comparison, the same portion of the dataset was reused, comprising data from a single environment (Environment 1) with three users performing three coarse-grained kitchen activities (10 trials each), collected by two sensors.

Subsequently, 43 common features [32] (Table 14) were extracted from the CSI amplitude of all 180 data samples. These features were provided as input to the genetic algorithm, which was tasked with selecting an optimal subset to maximize group separability, as measured by the MANOVA test (using Wilks' lambda as the fitness function). The goal was to identify the combination of handcrafted features that best distinguishes human activities based on their WiFi signal characteristics. The algorithm evolved the population over 40 generations via selection, crossover, and mutation. The best individual identified a subset of 28 features and achieved a maximum fitness (i.e., a minimal Wilks' lambda) of 36.04, which indicates a high degree of separability between the activity classes.

The selected 28-feature set was used to train the panel of 11 classifiers (Table 13). The top-performing classifier (Gradient Boosting, as detailed in the next paragraph) achieved an accuracy of 80.0% in recognizing the three coarse-grained kitchen activities. Performance was evaluated using stratified 5-fold cross-validation to ensure reliability. The confusion matrix for the optimal model is shown in Figure 55.

The 5-fold cross-validation results demonstrate the strong performance of the genetic algorithm-optimized model. Among all classifiers tested, Gradient Boosting emerged as the top performer with a mean accuracy of 73.33%, with Extra Trees, Random Forest, and MLP showing comparable performance. Comprehensive evaluation metrics reinforced the model's robustness, with macro-averaged precision, recall, and F1-score of 73.49%, 73.33%, and 73.04%, respectively, indicating balanced classification performance across all three activity classes. Collectively, these results validate that the feature selection via genetic algorithm enables accurate, generalized activity recognition while eliminating dependence on heuristic assumptions.

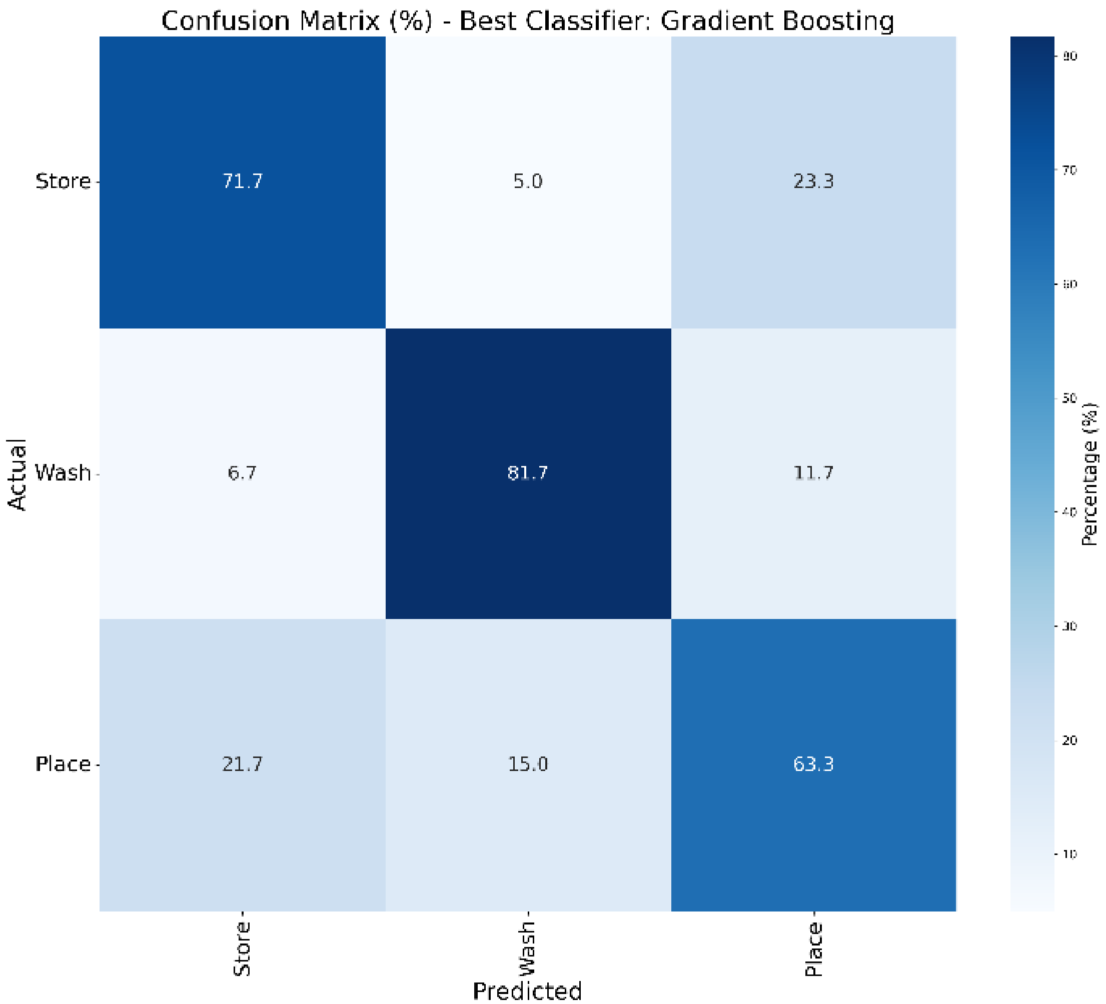

Figure 55 Confusion matrix for optimized model recognizing coarse-grained kitchen activities using Gradient Boosting algorithm and 5-fold cross validation

This experiment successfully fulfilled its objective to classify coarse-grained kitchen activities using the proposed optimized model and to benchmark it against the baseline heuristic approach. The results conclusively demonstrate the superiority of the MANOVA-guided genetic algorithm for feature selection. By identifying an optimal 28-feature subset from a pool of 43 candidates, the optimized model achieved a mean accuracy of 73.33% with the Gradient Boosting classifier, significantly outperforming the baseline heuristic model from Pre-Experiment 3 (68.3% accuracy, Section 6.3.3). This performance, validated through rigorous 5-fold cross-validation and supported by strong macro-averaged precision, recall, and F1-scores, confirms that the proposed method effectively addresses a key aspect of Problem 4 (The Optimization Problem). It provides a systematic, data-driven alternative to heuristic feature engineering, thereby enhancing the accuracy and generalizability of device-free HAR for coarse-grained activities and establishing a robust foundation for the more complex experiments that follow.

Table 14 Common features utilized in device-free HAR and their application in the genetic and heuristic algorithms.

#	Feature	Genetic	Heuristic
0	Mean	✓	
1	Standard Deviation	✓	
2	Median Absolute Deviation	✓	✓
3	Max		✓
4	Min	✓	✓
5	Interquartile Range	✓	✓
6	1st Quartile	✓	
7	3rd Quartile	✓	
8	Variance	✓	
9	Entropy		
10	Range	✓	
11	Median	✓	
12	Duration of Motion	✓	
13	Root Mean Square	✓	✓
14	Absolute Mean	✓	
15	Coefficient of Variation		✓
16	Skewness	✓	✓
17	Mean Crossing Rate	✓	
18	Peak to Peak	✓	✓
19	Above Mean Ratio		
20	Energy	✓	
21	Signal Magnitude Area	✓	
22	Autoregression		✓
23	Peak Factor	✓	
24	Wave Factor	✓	
25	Autocorrelation Coefficient	✓	
26	Mode		
27	Frequency Energy		
28	Frequency Entropy	✓	
29	Frequency Max		✓
30	Frequency Min		✓
31	Frequency Mean	✓	✓
32	Frequency Standard Deviation	✓	
33	Dominant Frequency Ratio	✓	
34	Frequency Left to Right Ratio	✓	
35	Frequency Full Width at Half Max		✓
36	Frequency Peaks		✓
37	Discrete Cosine Transform		
38	Frequency Range		✓
39	Frequency Skewness		✓
40	Frequency Kurtosis	✓	✓
41	Frequency Velocity		
42	Frequency Weighted Mean		

6.4.4. Experiment 2 – Optimized Model to Recognize Granular Kitchen Activities

Experiment 2 focused on evaluating the performance of the optimized method on fine-grained kitchen activity recognition within a single environment. The portion of the dataset from Environment 1 was reused, in which three users each performed nine distinct granular kitchen activities: Chop, Slice, French Cut, Pour, Stir, Ladle, Store, Wash, and Place. Each activity was performed 10 times per user and captured by two sensors, resulting in a total of 540 data samples:

$$\text{Number of data samples} = 3 \times 9 \times 2 \times 10 = 540$$

The same set of 43 features from both the time and frequency domains (as listed in Table 14) was extracted. To optimize the feature set for this more complex task, the MANOVA-based Genetic Algorithm was employed. It iteratively evaluated feature combinations based on their statistical separability across the nine activity classes. The optimization process ran for 40 generations and converged on an optimal subset of 25 features:

['F0', 'F1', 'F3', 'F4', 'F5', 'F8', 'F11', 'F13', 'F14', 'F15', 'F17', 'F18', 'F19', 'F21', 'F23', 'F24', 'F25', 'F26', 'F27', 'F29', 'F31', 'F32', 'F34', 'F37', 'F40']

The subsequent evaluation utilized a comprehensive set of 11 classical classifiers, including linear, tree-based, ensemble, and neural models. Each model was integrated into a pipeline with a StandardScaler and assessed using 5-fold stratified cross-validation to ensure robustness and fair comparison. The results showed that the Extra Trees classifier achieved the highest mean accuracy of 35.9%, slightly outperforming Random Forest (34.6%) and Gradient Boosting (32.6%).

The classification report for the best-performing model (Extra Trees) indicated that recognition performance varied significantly across activities. The highest F1-scores were achieved for a6 (Ladle) and a9 (Place) at 0.59 and 0.51, respectively. However, activities such as a3 (French Cut) and a4 (Pour) showed much lower F1-scores (below 0.30), highlighting the difficulty of distinguishing between kinematically similar actions using RF signals alone.

Despite careful preprocessing and feature engineering, the best-performing classifier (Extra Trees) achieved a mean accuracy of 35.9% across the 5-fold cross-validation. This result underscores the significant challenge of distinguishing fine-grained, kinematically similar activities using device-free RF-based sensing and suggests the need for further refinement in both feature design and model architecture. The confusion matrix for the nine granular kitchen activities is shown in Figure 56.

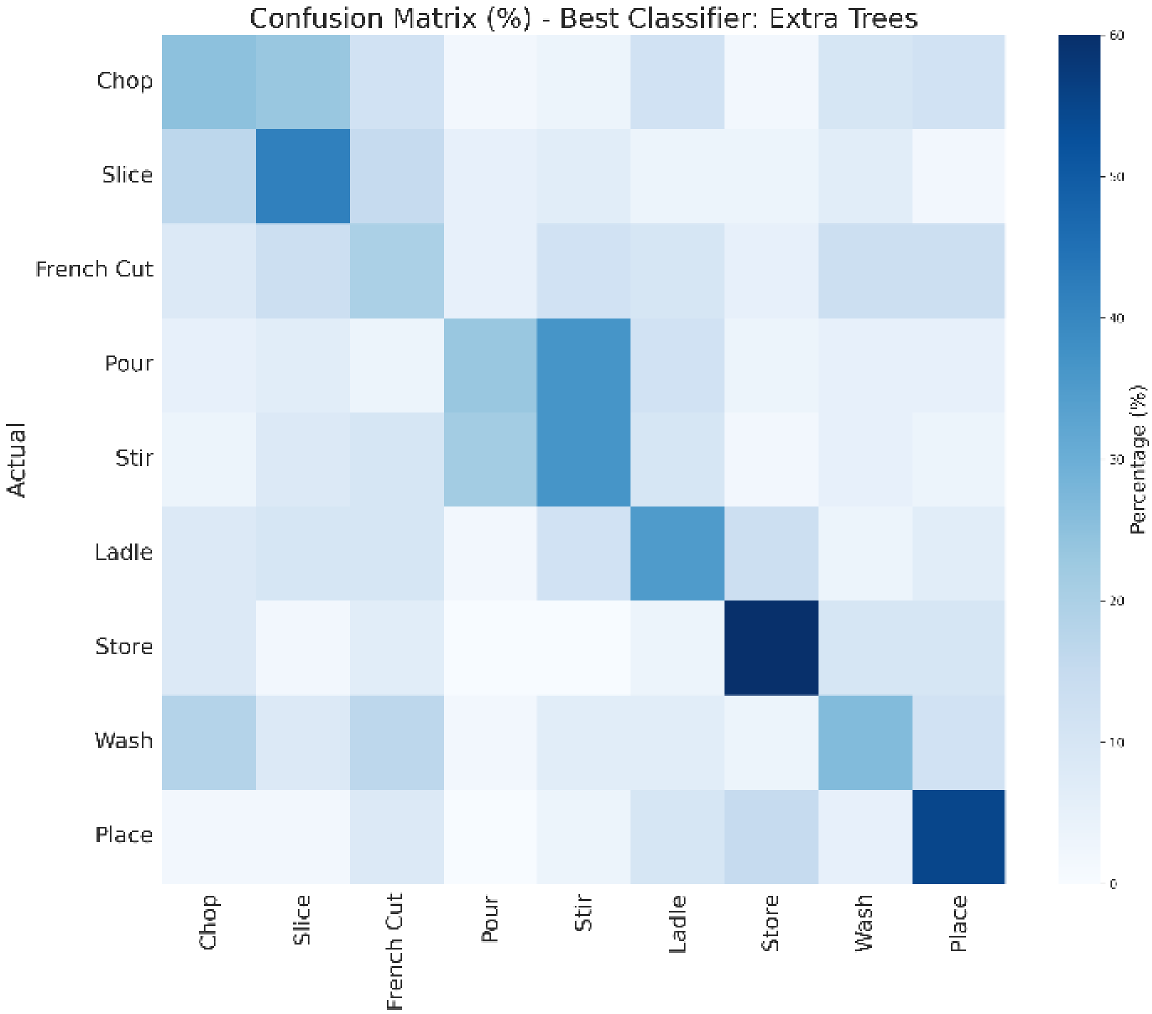

Figure 56 Confusion matrix for the Extra Trees classifier recognizing nine granular kitchen activities, evaluated using 5-fold cross-validation.

To contextualize these results, a systematic analysis of how activity granularity impacts classification performance was conducted. Beginning with the three fundamental coarse-grained activities, additional mid-grained and fine-grained activities were incrementally incorporated into the recognition task. Figure 57 demonstrates the progressive accuracy of the proposed method as the activity set is expanded, revealing how each added activity affects overall recognition performance. Furthermore, Figure 58 presents a distribution of the best-performing classifiers across all experimental configurations with different activity sets.

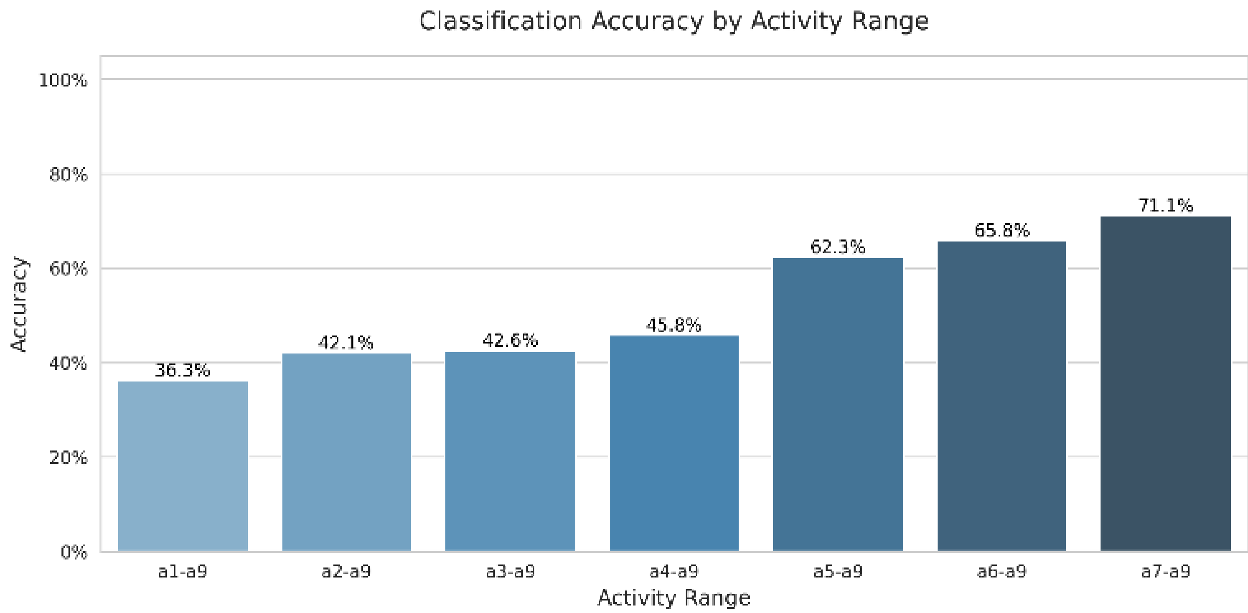

Figure 57 The relationship between activity set complexity (granularity) and mean classification accuracy achieved by the optimized model where chop (a1), slice (a2), French cut (a3), pour (a4), stir (a5), ladle (a6), store (a7), wash (a8), and place (a9).

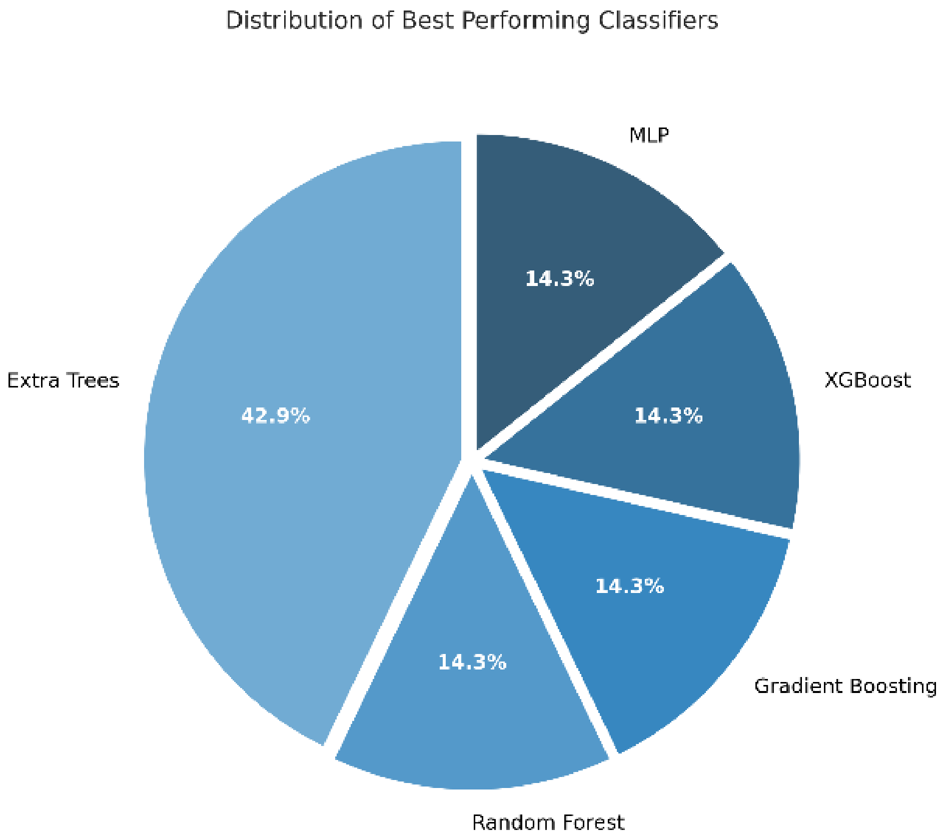

Figure 58 The distribution of the top-performing classifier across experiments involving different activity sets.

This experiment achieved its primary objective of rigorously evaluating the proposed optimized model on the significantly more challenging task of recognizing nine granular kitchen activities. The results clearly demonstrate a fundamental limitation: while the MANOVA-guided genetic algorithm successfully identified a discriminative 25-feature subset, the best achievable accuracy was 35.9%. This stark performance drop from coarse-grained activity recognition (73.33% in Experiment 1, Section 6.4.3) quantitatively validates the core challenge defined in Problem 1 (The Context-Specific Granularity Problem). The analysis revealed high confusion between kinematically similar activities like chopping, slicing, and French cutting, underscoring that subtle motion distinctions are not easily captured by the current RF signal features and classical learning paradigm. The analysis of performance degradation with increasing granularity (Figure 57) provides crucial empirical evidence for this phenomenon. Therefore, while the optimization method functioned as intended, the results conclusively highlight that recognizing fine-grained kitchen activities with high accuracy remains an open challenge, pointing to the need for more sophisticated sensing or learning approaches beyond the scope of classical methods.

6.4.5. Experiment 3 – Environment-Activity Recognition Using Two-Way MANOVA

To evaluate the ability of the proposed method to recognize human activities across different environments, the method was extended by applying a two-way MANOVA for feature selection. Two-way MANOVA is a statistical technique that examines the influence of two independent variables and their interaction on multiple dependent variables simultaneously. In this experiment, the independent factors were Environment and Activity, and the goal was to select features that best discriminate between combinations of these factors, i.e., environment-activity pairs. This approach is particularly valuable in complex device-free scenarios where both spatial and behavioral dynamics impact the observed signal patterns.

The dataset used in this experiment included recordings from three environments, where three users performed three coarse-grained kitchen activities: Store, Wash, and Place, with 10 trials per activity. Although the data was collected using two WiFi sensors, only the data from the transceiver unit was used to eliminate the effect of sensor variation.

$$\text{Number of data samples} = 3 \times 3 \times 3 \times 10 = 270$$

The full set of 43 statistical features was extracted from both the time and frequency domains, covering characteristics such as mean, standard deviation, skewness, entropy, and spectral properties. A two-way MANOVA-based genetic algorithm was then applied to identify the optimal subset of features that best distinguished between the nine possible environment-activity combinations (e.g., e1_a7, e2_a9, etc.).

Once feature selection was complete, multiple classifiers were trained and evaluated using 5-fold cross-validation, comparing their average performance. The best-performing model was Extra Trees, which achieved a mean accuracy of 68.51%. This represents a promising result, considering the complexity of simultaneously distinguishing between both activities and environments.

The classification report for the best model (Extra Trees) indicated particularly strong performance in recognizing activities in Environment 2, such as e2_a7 (F1-score: 0.85) and e2_a8 (F1-score: 0.82). However, some classes such as e3_a9 were more challenging, with lower precision and recall. These findings reflect the variability of environmental factors and their effect on WiFi signal propagation.

Figure 59 shows the confusion matrix for this experiment. In this matrix, e1, e2, and e3 represent kitchens 1, 2, and 3 respectively, while a7, a8, and a9 correspond to storing items in the fridge, washing the dishes, and placing the dishes in the cabinet. Overall, this experiment demonstrates that the proposed two-way MANOVA-guided feature selection effectively enhances recognition performance across diverse real-world environments.

While this level of accuracy may not be sufficient for a practical HAR system, the goal of this experiment is to shed light on the importance of environment-independent HAR method. For comparison, heuristic-based feature selection achieved an average accuracy of 60%, and PCA-based feature reduction resulted in 61.11% average accuracy for recognizing the same classes. The results indicate that the proposed method can be effectively used as an alternative to heuristic and PCA approaches. While heuristic methods require domain knowledge and may not be applicable in all scenarios, and PCA transforms features into abstract components, proposed approach preserves the original feature meanings, making it better suited for domain-specific analysis and interpretation.

This experiment successfully addressed its core objective of evaluating the proposed method's ability to achieve cross-environment robustness by introducing a two-way MANOVA-based feature selection approach. The results demonstrate the efficacy of this advanced technique, with the optimized model achieving a mean accuracy of 68.51% in the complex task of distinguishing environment-activity pairs. This performance represents a significant improvement over heuristic (60%) and PCA-based (61.11%) methods, validating the proposed method as a superior alternative for tackling Problem 2 (The Environmental Robustness Problem). The variability in performance across different environments (e.g., high accuracy in Environment 2 versus challenges with e3_a9) provides critical empirical evidence of how environmental factors influence RF signals. While the absolute accuracy may be below the threshold for a deployable system, this experiment successfully delivers a crucial insight: explicitly modeling the interaction between activity and environment during feature selection is a powerful strategy for improving generalization, thereby offering a concrete direction for developing environment-independent HAR systems.

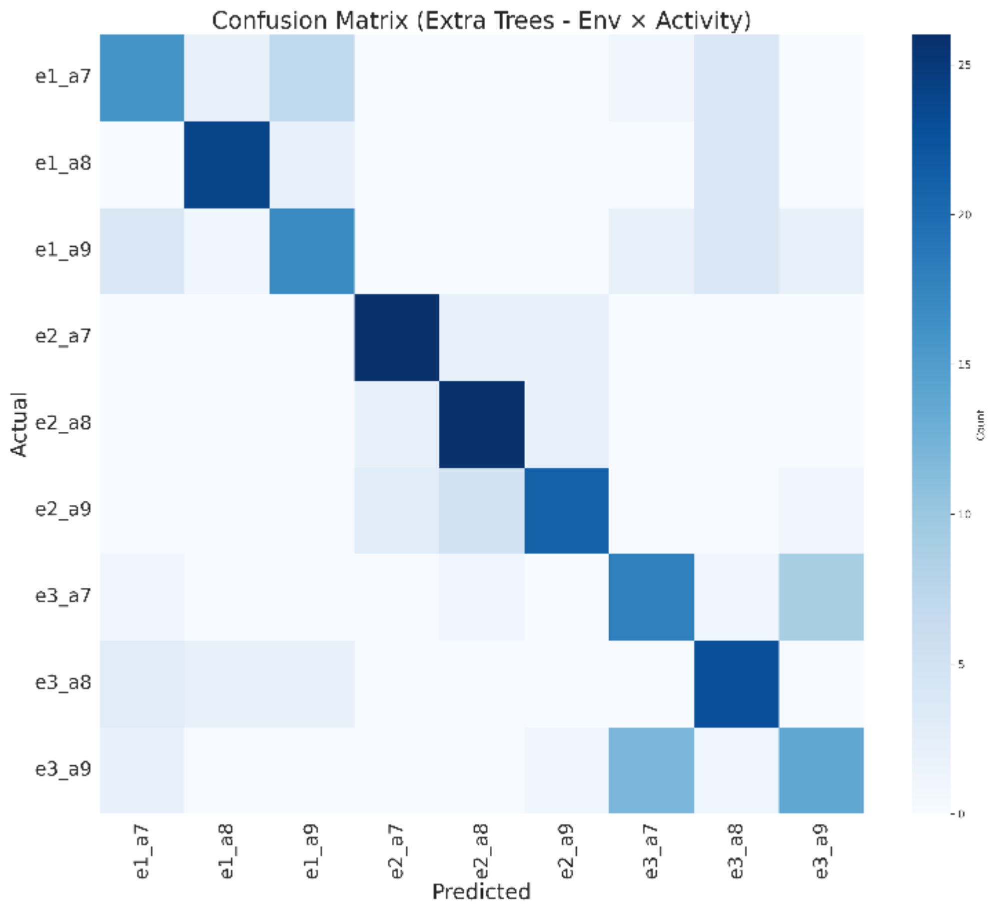

Figure 59 Confusion matrix for recognizing pairs of environment and activity using proposed optimized method

Chapter 7 Discussion

7.1 Interpretation of Results

This section synthesizes and evaluates the results of this dissertation against the research problems and questions defined in 4.3. The findings demonstrate significant progress in several key areas, while also revealing inherent challenges that define the boundaries of the proposed approach. Table 15 provides a high-level summary of how the research outcomes address the initial objectives, indicating which objectives were met and which remain open challenges.

Table 15 Summary of Research Problem Outcomes

Research Problem / Question	Outcome Status	Key Finding / Explanation
Problem 1: The Context-Specific Granularity Problem	Partially Met	The method successfully recognized coarse-grained activities (73.33% accuracy). However, performance on fine-grained activities was low (35.9%), quantitatively confirming the extreme difficulty of this task with current WiFi sensing and highlighting it as a primary challenge for future work. Section 6.4.3 and 6.4.4.
Problem 2: The Environmental Robustness Problem	Met	The two-way MANOVA approach for environment-activity recognition achieved 68.51% accuracy, significantly outperforming heuristic and PCA baselines. This demonstrates a concrete method for improving cross-environment generalization, a core aspect of robustness. Section 6.4.5.
Problem 3: The Data Scarcity and Benchmarking Problem	Met	An inclusive, multi-granularity dataset was successfully collected in three real-world kitchens with multiple participants, directly addressing the scarcity of public benchmarks for device-free kitchen HAR. Section 6.1.
Problem 4: The Optimization Problem	Met	The MANOVA-guided genetic algorithm for feature selection was validated as a superior alternative to heuristic and deep learning approaches for the given constraints, providing an optimized, systematic pipeline for classical learning. Section 6.4.
RQ1: What are the key steps in device-free HAR?	Met	Figure 60. Section 7.1.1.
RQ2/RQ3: Can WiFi sensing recognize kitchen activities?	Met	Yes, for coarse- and mid-grained activities with high accuracy. For fine-grained activities, recognition is possible but with low accuracy, establishing the current limits of the technology for this context. Section 6.4.4.
RQ4/RQ4.1: How does activity type/granularity affect performance?	Met	The system evaluation from coarse- to fine-grained activities (Figure 57) provided clear empirical evidence of a performance gap, directly quantifying the impact of granularity. Section 6.4.4.

RQ4.2: How does sensor placement affect performance?	Met	Feasibility Study 4 empirically identified that the optimal sensor placement is one which maintains both the user and the performing area within the Line-of-Sight (LoS), providing a critical design guideline. Section 5.4.
RQ5: How can ML/DL approaches be optimized?	Met	The comparative evaluation established that an optimized classical learning pipeline (feature selection + classifier ensemble) is currently the most effective approach for small-scale, context-sensitive kitchen HAR datasets. Section 6.4.

7.1.1. Results of the Systematic Literature Review

The systematic literature review [32] synthesized the common steps and components prevalent in device-free HAR research, as illustrated in Figure 60. The following paragraphs analyze these findings, highlighting established practices, common challenges, and identified gaps within the field.

The review indicates that a well-defined real-world application is a critical foundation for device-free HAR research, as it provides context and enhances the practical relevance and replicability of a proposed method. A key finding of this review is that unlike in other HAR domains, no comprehensive survey specifically investigating the applications of device-free HAR methods existed prior to this work. Based on reviewed papers, applications of device-free HAR are 1) human-computer interactions (including handwriting recognition [195], finger writing [196], finger spelling recognition [197], air writing [198], hand gestures [199] and sign language [54]), 2) Elderly care (including fall recognition [200], exercise recognition [201], moving recognition [202], sleep disorder diagnosis [203]), 3) health care (unhealthy eating habits detection [204], life monitoring [205]), 4) driver activity recognition [206], and 5) load monitoring [207].

The research process subsequently requires the definition of human activities that fulfill the application's requirements. The review identified a challenge here: the absence of a comprehensive study on device-free HAR applications makes it difficult to standardize the activities that should be recognized for each application domain. Nevertheless, the SLR [32] documents various activity categorizations found in the literature.

The next phase involves the selection of a sensory system. Section 3.1 and Table 3 detail the modalities, categorizations, and relative popularity of different sensors used in the reviewed literature. The choice is governed by a trade-off between various characteristics such as resolution, detection range, obtrusiveness, and environmental robustness, as systematically graded by Fu et al. [29]. Furthermore, the selection of an appropriate environment for data collection is a critical decision, with the review showing a predominance of controlled laboratory settings over real-world deployments, potentially limiting the practical generalizability of many studies.

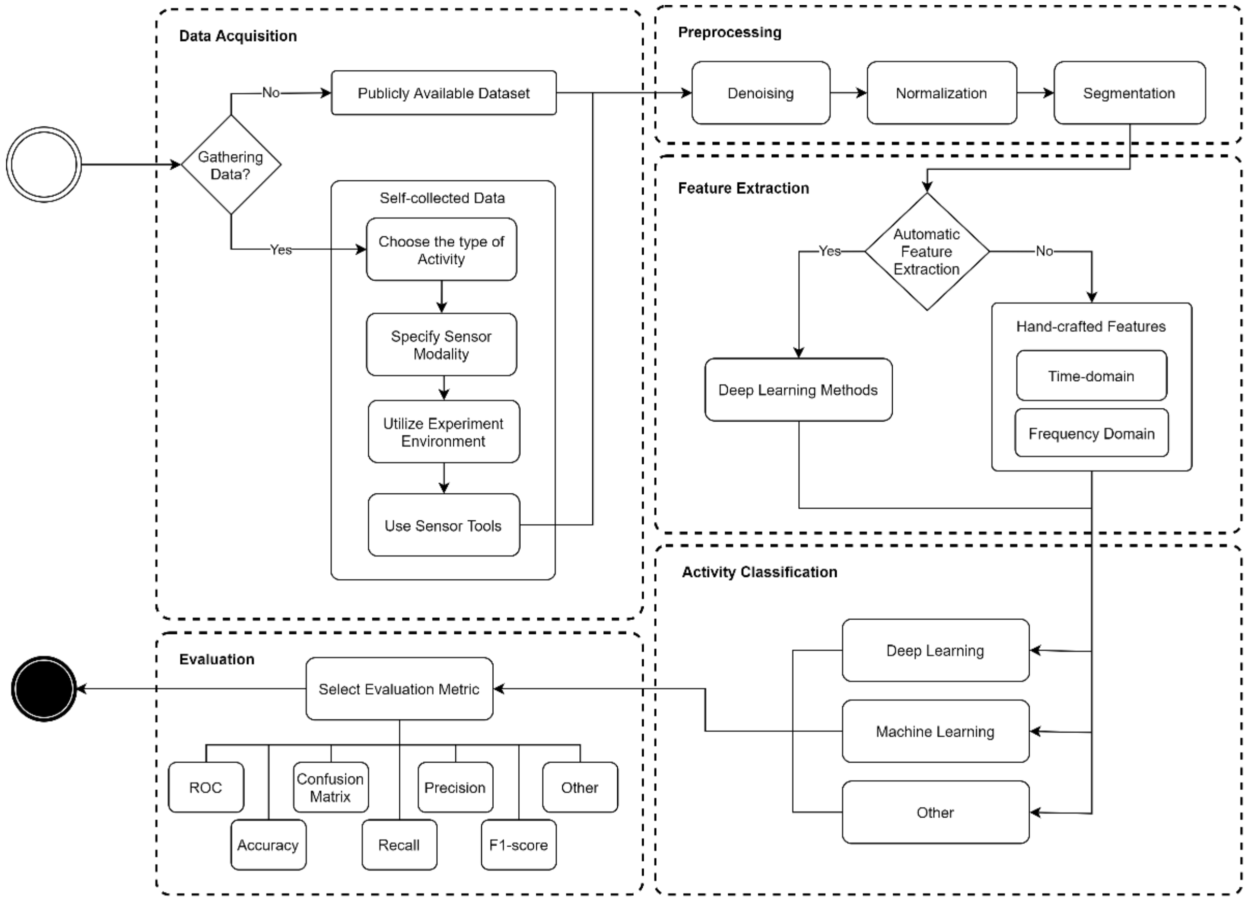

Figure 60 Steps of a device-free HAR method

Data collection from participants is a subsequent step. The review revealed a significant lack of consensus regarding appropriate cohort sizes for data collection. A notable finding was that 45 out of 242 reviewed papers (19%) evaluated their methods using data from only a single participant (see Figure 12). This practice inherently leads to user-dependent results, severely limiting the real-world applicability and generalizability of these methods.

The processing pipeline typically involves preprocessing noisy raw data, followed by feature extraction and classification. The review catalogued common preprocessing and feature extraction techniques (see relevant sections and tables). A key observation is that raw signal data is high-dimensional and necessitates the extraction of representative features, typically statistical properties from the time or frequency domain. For the classification stage, the review did not identify a single universally superior algorithm but did find that deep learning architectures with memory cells, such as LSTMs and GRUs, are particularly well-matched for modeling the temporal sequences inherent in device-free HAR data.

The final step is evaluation. The literature shows that comparing multiple algorithms is standard practice.

Performance is assessed using evaluation metrics; accuracy, precision, recall, and F1-score are the most prevalent, as detailed in the evaluation subsection. A critical conclusion from the review is that relying on a single metric provides an incomplete and potentially misleading assessment of classifier performance. A robust evaluation necessitates the use of multiple metrics to comprehensively capture different aspects of performance, such as the trade-off between precision and recall.

7.1.2. Results of the Multi-stage Evaluation

The results of Feasibility Study 3 demonstrated that placement significantly impacts HAR performance. The evaluation, conducted in an authentic kitchen using multiple transceiver-receiver configurations, revealed that performance varies considerably across different placements. Crucially, the results indicate that the highest recognition accuracy is achieved when the sensor placement ensures a clear LoS between the transceiver, the user, and the area where activities are performed. This configuration outperformed others, establishing that strategic, context-aware sensor placement is a critical factor for optimizing device-free HAR systems in real-world kitchens.

The findings from our multi-stage evaluation reveal important insights about the viability and performance of various methods in device-free kitchen activity recognition using WiFi CSI and RSSI signals. First and foremost, the deep learning approach, while promising in theory, struggled to generalize effectively in our constrained experimental setup. Despite incorporating context-aware features to augment the learning capacity of the LSTM model, the limited dataset size (270 samples) significantly restricted the model's ability to learn complex temporal dynamics. The result was poor training convergence and underfitting, aligning with earlier studies [183], [184] that stress the need for large, diverse datasets to successfully train deep neural networks for HAR.

The transfer learning module, which leveraged a pre-trained LSTM from a different activity domain (i.e., walking, sitting, standing), also underperformed. While this strategy aimed to adapt generic temporal features to the target domain, the mismatch between coarse full-body motions and fine-grained object-interactive kitchen tasks proved to be a major barrier. WiFi CSI is highly sensitive to physical layout and subtle motion differences, and thus, knowledge transfer from dissimilar domains resulted in low and stagnant accuracy (26–39%). These results underscore the critical limitations of applying naïve transfer learning to tasks with significant domain shift, especially in RF-based sensing where spatial context and object interaction matter greatly.

In contrast, the classical machine learning approach, which employed hand-crafted statistical and frequency features from CSI amplitude data, demonstrated stronger baseline performance. Even with a modest dataset, classical methods showed reasonable activity separability, achieving an average accuracy of 68.3% using Random Forest for three coarse-grained kitchen activities. This result validates the practical effectiveness of traditional classifiers when combined with well-designed, domain-informed features, particularly in small data settings.

Building on this observation, the optimized classical method introduced a novel feature selection strategy based on a MANOVA-guided genetic algorithm. By evolving combinations of features based on Wilks' lambda values, the algorithm identified subsets of features that maximized inter-class separability for both activity recognition and environment-activity pair discrimination. This hybrid statistical-learning approach outperformed heuristic baselines, reaching an accuracy of 73.33% for coarse-grained activities and 68.51% for environment-activity combinations. These results not only highlight the value of principled feature selection but also demonstrate the robustness of the optimized pipeline against environmental variability.

However, the results for fine-grained activity recognition were notably lower. Even the best model (i.e., Extra Trees with genetically selected features), achieved only 35.9% average accuracy across nine granular kitchen activities. This drop in performance reflects the inherent challenge of using RF signals to distinguish visually and kinematically similar activities (e.g., chopping vs. slicing), especially when these actions generate weak or overlapping signal patterns. Despite advanced feature selection and classifier tuning, the limited signal variability captured by CSI in such subtle movements imposes a natural ceiling on recognition accuracy. This reinforces the idea that while statistical feature optimization improves discriminative power, the physical limits of RF propagation in fine-movement contexts may necessitate complementary sensing modalities or multimodal fusion (e.g., combining CSI with audio or inertial sensors).

Finally, the environment-activity recognition experiment using two-way MANOVA illustrates a practical step toward real-world deployment of HAR systems. By addressing both spatial and behavioral variability, the proposed method showed that it is possible to build models with cross-environment generalization potential, something often missing in traditional device-free HAR studies. Compared to PCA and heuristic selection, the MANOVA-GA method preserved feature interpretability while achieving better classification accuracy, providing a meaningful balance between performance and explainability.

It is important to note a methodological consideration regarding the evaluation of the optimized classical method. The feature selection process using the MANOVA-guided genetic algorithm (GA) was performed prior to cross-validation. This means the same data was used to both select the optimal feature subset and evaluate the final classifier's performance in the subsequent k-fold cross-validation. While this provides a valid estimate of the method's performance when the feature selection rule is considered part of the learned model, it can introduce an optimistic bias compared to a fully nested cross-validation scheme, where the feature selection is repeated independently within each training fold. The reported accuracies (e.g., 73.33% for coarse-grained activities) should therefore be interpreted as the performance of the entire optimized pipeline (feature selection + classifier) on the available data, rather than as an unbiased estimate of its performance on entirely new data from the same distribution. This design was chosen for computational efficiency and to establish a strong performance baseline for the proposed feature selection heuristic. Future work employing this method should implement a nested cross-validation protocol to obtain a fully unbiased performance estimate and confirm the generalizability of the selected features.

In summary, this research affirms that while deep learning and transfer learning have theoretical advantages, classical methods with optimized feature selection currently offer a more robust and interpretable solution for small-scale, environment-sensitive, device-free HAR systems. Moreover, it highlights the limitations of current datasets and sensing modalities in capturing nuanced human behaviors, especially in complex domains like kitchens. Future work should focus on expanding dataset diversity, exploring multimodal sensing integration, and developing hybrid models that blend domain-informed feature selection with deep representation learning. These steps are essential for building HAR systems that are not only accurate but also generalizable and deployable in real-world environments.

7.2 Comparison with Related Work

The proposed method demonstrates several advantages over existing device-free HAR systems, particularly in the context of kitchen activity recognition. Below, this section compares the proposed approach with related work in terms of sensing modality, feature extraction, classification techniques, and robustness to environmental variations.

7.2.1. Sensing Modality and Hardware Efficiency

- Related Work: Many prior studies rely on specialized hardware (e.g., USRP, Intel 5300 NIC) for high-resolution CSI extraction, which is expensive and impractical for real-world deployment.

Others use RSSI-based methods, which lack granularity for fine-grained activity recognition.

- Proposed Approach: This work utilizes low-cost, commercially available ESP32 modules, thereby enhancing the system's accessibility and scalability for real-world deployment. Despite the hardware limitations inherent in such devices, the optimized feature selection and classical learning pipeline achieves competitive accuracy, demonstrating that high-end hardware is not a strict prerequisite for effective HAR.

7.2.2. Feature Extraction and Selection

- Related Work: Many existing methods rely on heuristic feature selection or deep learning without explicit feature optimization, leading to suboptimal performance in cross-domain scenarios.
- Proposed Approach: In contrast, this work introduces a hybrid feature selection strategy using a Genetic Algorithm guided by MANOVA. Unlike heuristic or PCA-based methods, this approach systematically identifies features that maximize inter-class separability while preserving interpretability. A novel extension of this is the use of two-way MANOVA for environment-activity recognition, which explicitly models and accounts for environmental variations, a factor often overlooked in prior work that assumes a fixed experimental setup.

7.2.3. Classification Performance

- Related Work: Deep learning methods (e.g., LSTMs, CNNs) struggle with small datasets and require extensive training data. Classical methods (e.g., Random Forest, SVM) often rely on manually selected features, limiting generalization.
- Proposed Approach: This work demonstrates that for small-scale datasets, an optimized classical learning pipeline (achieving 73.33% accuracy for coarse-grained activities) can significantly outperform both deep learning (28-43%) and transfer learning (26-39%) approaches. Furthermore, robustness is ensured by evaluating a comprehensive panel of classifiers (e.g., Extra Trees, Gradient Boosting, Random Forest) rather than relying on a single model.

7.2.4. Robustness to Environmental Variations

- Related Work: Most studies evaluate HAR in controlled lab environments, ignoring real-world variability.
- Proposed Approach: The proposed method explicitly addresses environmental variability. The

two-way MANOVA feature selection achieved 68.51% accuracy in distinguishing activities across different kitchens, demonstrating a degree of adaptability to unseen environments. Additionally, the selected features show resilience to minor changes in sensor placement, reducing the dependency on precise, fixed configurations.

7.2.5. Granularity of Recognized Activities

- Related Work: Many HAR systems focus only on coarse-grained activities (e.g., walking, sitting) and struggle with fine-grained motions (e.g., chopping vs. slicing).
- Proposed Approach: This work provides a systematic evaluation across a spectrum of activity granularity, from coarse-grained (73.33% accuracy) to fine-grained (35.9% accuracy). This hierarchical analysis offers concrete insights into the specific challenges associated with recognizing subtle, object-interactive motions compared to more general activities.

7.3 Limitations

The proposed method is effective in controlled settings but still has several limitations that must be acknowledged. These constraints stem from hardware, environmental factors, and inherent challenges in device-free HAR using WiFi CSI.

7.3.1. Hardware and Sensor Limitations

- This research utilized the ESP32-WROOM-32 module. The performance of the proposed method is intrinsically tied to the capabilities and limitations of this specific hardware, particularly its Wi-Fi chipset and antenna design. Different transceivers (e.g., Intel 5300 NIC, Atheros AR9580) offer varying numbers of spatial streams, subcarrier resolution, and sampling rates. Therefore, the findings on feature importance and classification performance may not be directly generalizable, and a different sensor choice could potentially yield different results, either positively or negatively impacting accuracy.

7.3.2. Sensitivity to Environmental Layout

- Fixed Sensor Placement Requirement: Even minor changes in sensor or furniture positions alter multipath propagation, degrading recognition accuracy. This limits real-world deployment where environments are dynamic.

- LoS Dependency: Activities performed outside the direct LoS between transceivers may not generate detectable signal variations. It affects the generalizability of proposed system.

7.3.3. Activity-Related Noise

- Inter-User Variability: Differences in body size, movement speed, and style (e.g., aggressive vs. gentle stirring) lead to inconsistent CSI patterns, reducing model generalizability.
- Limited Fine-Grained Discrimination: While coarse-grained activities (e.g., "Store" vs. "Wash") achieve ~73% accuracy, distinguishing fine-grained motions (e.g., "Chop" vs. "Slice") remains challenging (35.9% accuracy).

7.3.4. Ambient Noise Factors

- Wi-Fi Interference: Real-world kitchens often contain microwaves, Bluetooth devices, and neighboring Wi-Fi networks, which cause unpredictable signal fluctuations.
- Non-Human Motions: Movements of pets, ceiling fans, or curtains may affect the signal path and quality of the data.

7.3.5. Dataset Scale and Diversity

- Few Participants: Data from only three users restricts the model's ability to generalize across demographics (e.g., height, handedness).
- Single-Activity at a Time: The dataset captures isolated activities, whereas real cooking involves concurrent actions (e.g., stirring while walking), which the system cannot yet consider.

7.3.6. Computational and Practical Barriers

- Genetic Algorithm Overhead: The MANOVA-based feature selection, while effective, can be computationally intensive and impractical for real-time edge deployment.
- Lack of Real-Time Testing: The offline evaluation does not account for latency or processing delays in live scenarios.

Chapter 8 Conclusion and Future Works

This dissertation explored multiple approaches for device-free kitchen activity recognition using WiFi CSI, including deep learning, transfer learning, and classical machine learning techniques. The experimental results demonstrated that, given the limited dataset size and the complexity of subtle kitchen activities, deep learning models and transfer learning strategies face significant challenges in achieving robust performance, as they require larger, more diverse datasets to generalize effectively.

In contrast, classical machine learning methods combined with carefully engineered statistical and frequency-domain features proved more effective for recognizing coarse-grained kitchen activities. This work developed a novel MANOVA-guided genetic algorithm to optimize the feature set, which enhanced classification accuracy and robustness across varying environments. This hybrid approach successfully balanced interpretability with performance, presenting a practical solution for real-world device-free HAR.

Despite promising results for coarse-grained tasks, recognizing fine-grained kitchen activities remains challenging due to the subtle differences in RF signal patterns. This highlights the inherent limitations of using WiFi CSI alone for complex activity recognition and suggests that future research should investigate multimodal sensing approaches or more sophisticated deep learning architectures tailored for fine-grained actions. Questions remain regarding which level of granularity is both feasible to capture and most valuable for supporting recommender systems in smart homes.

Overall, the findings of this research emphasize the critical importance of dataset size, feature engineering, and environmental considerations in device-free HAR. The proposed optimized classical learning method offers a strong foundation for future work aimed at developing accurate, generalizable, and interpretable WiFi-based activity recognition systems for smart home applications.

Future research in device-free human activity recognition (HAR) should focus on several key areas to enhance system performance and applicability. Improving noise reduction during preprocessing and advancing feature extraction methods will help the system better capture subtle activity details. Classifier fusion, integrating both traditional and deep learning models, can improve classification accuracy and versatility. Additionally, reducing dependency on specific environments and users, as well as achieving real-time responsiveness, are essential for deploying HAR systems in diverse, real-world contexts. Addressing these challenges will lead to more robust, adaptable, and accurate HAR systems suitable for applications across various fields.

8.1 Underexplored Sensor Modalities

Future work should explore beyond WiFi CSI to include multi-modal sensor fusion. Techniques that combine WiFi with other modalities, such as cameras (e.g., AutoDLAR [208]) or acoustic sensors, could overcome the inherent limitations of a single sensing technology. Furthermore, emerging sensing paradigms, such as those leveraging 5G/6G infrastructure and Reconfigurable Intelligent Surfaces (RIS) [209], represent a promising but underexplored frontier for device-free HAR.

8.2 Advanced Preprocessing Techniques

Future work must develop more advanced preprocessing techniques to isolate activity-related signals from environmental and RF noise. Investigating adaptive filtering, blind source separation, or advanced noise-canceling algorithms tailored for CSI data could significantly improve signal clarity before feature extraction.

8.3 Enhanced Feature Engineering

Future research should prioritize feature engineering methods that capture the most informative, context-independent characteristics of human activity. This includes developing environment-agnostic, user-agnostic, and activity-agnostic features that generalize across diverse settings and subjects. This will be essential for HAR systems that need to perform accurately and reliably across various application domains without requiring extensive recalibration.

8.4 Improved Classification via Fusion Techniques

Future work should investigate hybrid classification models that fuse the strengths of classical and deep learning approaches. Ensemble methods, stacking, or other fusion techniques could be employed to create more robust and accurate classifiers that leverage the interpretability of classical models and the representational power of deep networks.

8.5 Fine-Grained Activity Recognition

A critical direction for future work is a dedicated focus on fine-grained activity recognition. Overcoming the current accuracy ceiling for subtle motions requires novel approaches, potentially involving higher-resolution sensing, better temporal modeling, or the multimodal fusion mentioned in Section 8.1. Success in this area would enable applications in health monitoring, precision sports training, and personalized services.

8.6 Low Computational Cost, High Accuracy Approaches

Future work should explore computationally efficient algorithms that maintain high accuracy while minimizing processing overhead. Techniques such as lightweight feature extraction, model compression, pruning, or the development of novel algorithms designed for edge deployment are essential for making HAR systems scalable and practical for resource-constrained devices in real-world settings.

8.7 Data-Efficient Deep Learning with Aggressive Augmentation

Due to the sensitivity of deep learning models to limited dataset size, the future researchers may explore simpler, more data-efficient neural architectures trained from scratch, rather than relying on transfer learning from dissimilar domains. Specifically, a promising future direction is to combine the data augmentation strategies with a lightweight model such as a 2-layer CNN trained directly on the augmented kitchen activity data. This approach is fundamentally different from the transfer learning method evaluated in Chapter 6, as it does not depend on pre-trained weights from an irrelevant source domain. Instead, a shallow CNN may avoid the overfitting that plagued the LSTM while still capturing the informative spatial patterns in CSI data when trained on a sufficiently expanded dataset via augmentation. Investigating this architecture would help determine the minimum viable model complexity for device-free HAR and establish a more robust baseline for deep learning in data-constrained scenarios.

References

[1] K. Polikanov and S. Sotnik, “Smart home with house module: overview of automation technologies”, Digital innovation & sustainable development Conference Proceedings, pp. 20-21, 2024.

[2] Z. Kakavand, A. A. Nazari Shirehjini, and S. Shirmohammadi, “Exploring Privacy in Smart Homes: A Cognitive Walkthrough of Automation Levels for Data Monetization,” in HCI International 2024 Posters, vol. 2120, pp. 265–275, 2024, doi: 10.1007/978-3-031-62110-9_28.

[3] S. Kalimuthu, T. Perumal, R. Yaakob, E. Marlisah, and S. Raghavan, “Multiple human activity recognition using iot sensors and machine learning in device-free environment: Feature extraction, classification, and challenges: A comprehensive review,” in AIP Conference Proceedings, Banda Aceh, Indonesia: AIP Publishing, p. 120003. 2024, doi: 10.1063/5.0179747.

[4] A. Gupta, K. Gupta, K. Gupta, and K. Gupta, “A Survey on Human Activity Recognition and Classification,” in 2020 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, pp. 0915–0919, July 2020, doi: 10.1109/ICCSP48568.2020.9182416.

[5] H. Raeis, M. Kazemi, and S. Shirmohammadi, “Human Activity Recognition with Device-Free Sensors for Well-Being Assessment in Smart Homes,” IEEE Instrumentation Measurement Magazine, vol. 24, no. 6, pp. 46–57, 2021, doi: 10.1109/MIM.2021.9513637.

[6] M. H. Siddiqi, N. Almashfi, A. Ali, M. Alruwaili, Y. Alhwaiti, S. Alanazi, and M. M. Kamruzzaman, “A Unified Approach for Patient Activity Recognition in Healthcare Using Depth Camera,” IEEE Access, vol. 9, pp. 92300–92317, 2021, doi: 10.1109/ACCESS.2021.3092403.

[7] P. Sok, T. Xiao, Y. Azeze, A. Jayaraman, and M. V. Albert, “Activity Recognition for Incomplete Spinal Cord Injury Subjects Using Hidden Markov Models,” IEEE Sensors J., vol. 18, no. 15, pp. 6369–6374, Aug. 2018, doi: 10.1109/JSEN.2018.2845749.

[8] M. A. Gul, M. H. Yousaf, S. Nawaz, Z. Ur Rehman, and H. Kim, “Patient Monitoring by Abnormal Human Activity Recognition Based on CNN Architecture,” Electronics, vol. 9, no. 12, p. 1993, Nov. 2020, doi: 10.3390/electronics9121993.

[9] E. Fridriksdottir and A. G. Bonomi, “Accelerometer-Based Human Activity Recognition for Patient Monitoring Using a Deep Neural Network,” Sensors, vol. 20, no. 22, pp. 6424, Nov. 2020, doi: 10.3390/s20226424.

[10] S. Zaidi, B. Jagadeesh, K. V. Sudheesh, and A. A. Audre, “Video Anomaly Detection and Classification for Human Activity Recognition,” in 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC), pp. 544–548, Sept. 2017, doi: 10.1109/CTCEEC.2017.8455012.

[11] D. G. Shreyas, S. Raksha, and B. G. Prasad, “Implementation of an Anomalous Human Activity Recognition System,” SN COMPUT. SCI., vol. 1, no. 3, p. 168, May 2020, doi: 10.1007/s42979-020-00169-0.

[12] D. Wu, J. Chen, W. Deng, Y. Wei, H. Luo, and Y. Wei, “The Recognition of Teacher Behavior Based on Multimodal Information Fusion,” Mathematical Problems in Engineering, vol. 2020, pp. 1–8, Oct. 2020, doi: 10.1155/2020/8269683.

[13] N. Darapaneni, D. Singh, S. Chandra, A. R. Paduri, N. Kilhore, S. Chopra, A. Srivastav, R. Tomer, and S. S. Deshmukh, “Activity & Emotion Detection of Recognized kids in CCTV Video for Day Care Using SlowFast & CNN,” in IEEE World AI IoT Congress (AIIoT), Seattle, WA, USA, pp. 0268–0274, May 2021, doi: 10.1109/AIIoT52608.2021.9454221.

[14] Z. Hussain, Q. Z. Sheng, and W. E. Zhang, “A review and categorization of techniques on device-free human activity recognition,” Journal of Network and Computer Applications, vol. 167, pp. 102738, Oct. 2020, doi: 10.1016/j.jnca.2020.102738.

[15] H. Duan, M. Huang, Y. Yang, J. Hao, and L. Chen, “Ambient Light-Based Hand Gesture Recognition Enabled by Recurrent Neural Network,” IEEE Access, vol. 8, pp. 7303–7312, 2020, doi: 10.1109/ACCESS.2019.2963440.

[16] M. H. Kabir, M. H. Rahman, and W. Shin, “CSI-IANet: An Inception Attention Network for Human-Human Interaction Recognition Based on CSI Signal,” IEEE Access, vol. 9, pp. 166624–166638, 2021, doi: 10.1109/ACCESS.2021.3134794.

[17] H. Cheng and W. Lou, “Push the Limit of Device-Free Acoustic Sensing on Commercial Mobile Devices,” in IEEE INFOCOM 2021 - IEEE Conference on Computer Communications, pp. 1–10, May 2021, doi: 10.1109/INFOCOM42981.2021.9488703.

[18] V. Nguyen, M. Ibrahim, S. Rupavatharam, M. Jawahar, M. Gruteser, and R. Howard, “Eyelight: Light-and-Shadow-Based Occupancy Estimation and Room Activity Recognition,” in IEEE INFOCOM 2018 - IEEE Conference on Computer Communications, pp. 351–359, Apr. 2018, doi: 10.1109/INFOCOM.2018.8485867.

[19] J. Liu, H. Liu, Y. Chen, Y. Wang, and C. Wang, “Wireless Sensing for Human Activity: A Survey,” IEEE Communications Surveys & Tutorials, vol. 22, no. 3, pp. 1629–1645, 2020, doi: 10.1109/COMST.2019.2934489.

[20] M. Raja and S. Sigg, “Applicability of RF-based methods for emotion recognition: A survey,” in 2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops), pp. 1–6, Mar. 2016, doi: 10.1109/PERCOMW.2016.7457119.

[21] S. Shukri and L. M. Kamarudin, “Device free localization technology for human detection and counting with RF sensor networks: A review,” Journal of Network and Computer Applications, vol. 97, pp. 157–174, 2017, doi: 10.1016/j.jnca.2017.08.014.

[22] S. Wang and G. Zhou, “A review on radio-based activity recognition,” Digital Communications and Networks, vol. 1, no. 1, pp. 20–29, Feb. 2015, doi: 10.1016/j.dcan.2015.02.006.

[23] J. Wang, Q. Gao, X. Ma, Y. Zhao, and Y. Fang, “Learning to Sense: Deep Learning for Wireless Sensing with Less Training Efforts,” IEEE Wireless Communications, vol. 27, no. 3, pp. 156–162, June 2020, doi: 10.1109/MWC.001.1900409.

[24] C. Wang, S. Chen, Y. Yang, F. Hu, F. Liu, and J. Wu, “Literature review on wireless sensing-Wi-Fi signal-based recognition of human activities,” Tsinghua Science and Technology, vol. 23, no. 2, pp. 203–222, Apr. 2018, doi: 10.26599/TST.2018.9010080.

[25] M. A. A. Al-qaness, M. Abd Elaziz, S. Kim, A. A. Ewees, A. A. Abbasi, Y. A. Alhaj, and A. Hawbani, “Channel State Information from Pure Communication to Sense and Track Human Motion: A Survey,” Sensors, vol. 19, no. 15, p. 3329, July 2019, doi: 10.3390/s19153329.

[26] Z. Wang, K. Jiang, Y. Hou, Z. Huang, W. Dou, C. Zhang, and Y. Guo , “A Survey on CSI-Based Human Behavior Recognition in Through-the-Wall Scenario,” IEEE Access, vol. 7, pp. 78772–78793, 2019, doi: 10.1109/ACCESS.2019.2922244.

[27] Z. Wang, K. Jiang, Y. Hou, W. Dou, C. Zhang, Z. Huang, and Y. Guo, “A Survey on Human Behavior Recognition Using Channel State Information,” IEEE Access, vol. 7, pp. 155986–156024, 2019, doi: 10.1109/ACCESS.2019.2949123.

[28] H. F. Thariq Ahmed and H. Ahmad, “Device free human gesture recognition using Wi-Fi CSI: A survey,” Engineering Applications of Artificial Intelligence, vol. 87, 2020, doi: 10.1016/j.engappai.2019.103281.

[29] B. Fu, N. Damer, F. Kirchbuchner, and A. Kuijper, “Sensing technology for human activity

recognition: A comprehensive survey," IEEE Access, vol. 8, pp. 83791–83820, 2020, doi: 10.1109/ACCESS.2020.2991891.

[30] A. Lentzas and D. Vrakas, "Non-intrusive human activity recognition and abnormal behavior detection on elderly people: a review," Artificial Intelligence Review, vol. 53, no. 3, pp. 1975–2021, Mar. 2020, doi: 10.1007/s10462-019-09724-5.

[31] N. Gupta, S. K. Gupta, R. K. Pathak, V. Jain, P. Rashidi, and J. S. Suri, "Human activity recognition in artificial intelligence framework: a narrative review," Artificial Intelligence Review, vol. 55, no. 6, pp. 4755–4808, Aug. 2022, doi: 10.1007/s10462-021-10116-x.

[32] M. G. Moghaddam, A. A. N. Shirehjini, and S. Shirmohammadi, "Device-Free Human Activity Recognition: A Systematic Literature Review," IEEE Open Journal of Instrumentation and Measurement, vol. 4, pp. 1–34, 2025, doi: 10.1109/OJIM.2024.3502885.

[33] F. Miao, Y. Huang, Z. Lu, T. Ohtsuki, G. Gui, and H. Sari, "Wi-Fi Sensing Techniques for Human Activity Recognition: Brief Survey, Potential Challenges, and Research Directions," ACM Comput. Surv., vol. 57, no. 5, pp. 1–30, May 2025, doi: 10.1145/3705893.

[34] S. Kalimuthu, T. Perumal, E. Marlisah, R. Yaakob, V. BS, and N. H. Ismail, "Human Activity Recognition Based on Device-Free Wi-Fi Sensing: A Comprehensive Review," Malaysian Journal of Computer Science, vol. 37, no. 3, pp. 252–269, July 2024.

[35] J. Xiao, H. Li, M. Wu, H. Jin, M. J. Deen, and J. Cao, "A Survey on Wireless Device-free Human Sensing: Application Scenarios, Current Solutions, and Open Issues," ACM Comput. Surv., vol. 55, no. 5, pp. 1–35, May 2023, doi: 10.1145/3530682.

[36] S. Bian, M. Liu, B. Zhou, and P. Lukowicz, "The State-of-the-Art Sensing Techniques in Human Activity Recognition: A Survey," Sensors, vol. 22, no. 12, p. 4596, June 2022, doi: 10.3390/s22124596.

[37] J. Wang, Q. Gao, M. Pan, and Y. Fang, "Device-Free Wireless Sensing: Challenges, Opportunities, and Applications," IEEE Network, vol. 32, no. 2, pp. 132–137, Mar. 2018, doi: 10.1109/MNET.2017.1700133.

[38] T. Raj, T. Nisar, M. Abbas, R. Priyadarshini, S. Naz, and U. Tiwari, "Qualitative Analysis of Techniques for Device-Free Human Activity Recognition," RIA, vol. 37, no. 3, pp. 639–353, June 2023, doi: 10.18280/ria.370313.

[39] L. Bibbò et al., "MEMS and IoT in HAR: Effective Monitoring for the Health of Older People," Applied Sciences, vol. 15, no. 8, Apr. 2025, doi: 10.3390/app15084306.

[40] Z. Chen, L. Teng, L. Xu, J. Yu, and J. Liang, "MP-HAR: A Novel Motion-Powered Real-Time Human Activity Recognition System," IEEE Internet of Things Journal, vol. 11, no. 5, pp. 7652–7663, Mar. 2024, doi: 10.1109/JIOT.2023.3315789.

[41] J. Yang, H. Zou, and L. Xie, "SecureSense: Defending Adversarial Attack for Secure Device-Free Human Activity Recognition," IEEE Transactions on Mobile Computing, vol. 23, no. 1, pp. 823–834, Jan. 2024, doi: 10.1109/TMC.2022.3226742.

[42] M. Pang, X. Yang, J. Liu, P. Li, F. Yan, and P. Chen, "Device-Free Activity Recognition: A Survey," Communications in Computer and Information Science, vol. 1321, pp. 223–243, 2020, doi: 10.1007/978-981-33-4214-9_17.

[43] D. Tse and P. Viswanath, "MIMO I: spatial multiplexing," Fundamentals of Wireless Communication, pp. 341, 2004.

[44] X. Wang, K. Niu, J. Xiong, B. Qian, Z. Yao, T. Lou, and D. Zhang, "Placement matters: Understanding the effects of device placement for WiFi sensing," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 6, pp. 1–25, 2022.

[45] T. Li, C. Shi, P. Li, and P. Chen, "A Novel Gesture Recognition System Based on CSI Extracted from a Smartphone with Nexmon Firmware," Sensors, vol. 21, no. 1, p. 222, Dec. 2020, doi: 10.3390/s21010222.

[46] Y. Ma, G. Zhou, and S. Wang, "WiFi Sensing with Channel State Information," ACM Computing Surveys, vol. 52, no. 3, pp. 1–36, May 2020, doi: 10.1145/3310194.

[47] M. G. Moghaddam, A. A. N. Shirehjini, and S. Shirmohammadi, "A WiFi-based Method for Recognizing Fine-grained Multiple-Subject Human Activities," IEEE Transactions on Instrumentation and Measurement, pp. 1, 2023, doi: 10.1109/TIM.2023.3289547.

[48] C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning," Electronic Markets, vol. 31, no. 3, pp. 685–695, Sept. 2021, doi: 10.1007/s12525-021-00475-2.

[49] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[50] N. Yehia, "Understanding Long Short-Term Memory (LSTM) Networks," Machine Learning Archive. [Online]. Available: https://mlarchive.com/deep-learning/understanding-long-short-term-memory-networks/

[51] J. Zhang, F. Wu, B. Wei, Q. Zhang, H. Huang, S. W. Shah, and J. Cheng, "Data Augmentation and

Dense-LSTM for Human Activity Recognition Using WiFi Signal," IEEE Internet of Things Journal, vol. 8, no. 6, pp. 4628–4641, Mar. 2021, doi: 10.1109/JIOT.2020.3026732.

[52] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A Comprehensive Survey on Transfer Learning," Proceedings of the IEEE, vol. 109, no. 1, pp. 43–76, Jan. 2021, doi: 10.1109/JPROC.2020.3004555.

[53] Q. Bu, X. Ming, J. Hu, T. Zhang, J. Feng, and J. Zhang, "TransferSense: towards environment independent and one-shot wifi sensing," Personal and Ubiquitous Computing, vol. 26, no. 3, pp. 555–573, June 2022, doi: 10.1007/s00779-020-01480-6.

[54] W. Lin, L. Yu, and J. Nan, "WiCLR: A Sign Language Recognition System Framework Based on Wireless Sensing," Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST, vol. 303 LNICST, pp. 102–117, 2019, doi: 10.1007/978-3-030-36442-7_7.

[55] F. Xiao, J. Chen, Z. Li, H. Huang, and L. Sun, "Improved LDA Dimension Reduction Based Behavior Learning with Commodity WiFi for Cyber-Physical Systems," ACM Trans. Cyber-Phys. Syst., vol. 3, no. 4, Sept. 2019, doi: 10.1145/3342219.

[56] N. Yu, W. Wang, A. X. Liu, and L. Kong, "QGesture: Quantifying Gesture Distance and Direction with WiFi Signals," Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., vol. 2, no. 1, Mar. 2018, doi: 10.1145/3191783.

[57] M. Atif, B. Yoo, and H. Ko, "Device Free Wireless Sensing based Human Activity Recognition Using Commercial Off-the-Shelf IoT Single-Board Computers," in 2024 IEEE International Conference on Consumer Electronics (ICCE), pp. 1–6, Jan. 2024, doi: 10.1109/ICCE59016.2024.10444143.

[58] A. Natarajan, V. Krishnasamy, and M. Singh, "Device-Free Human Activity Recognition in Through-the-Wall Scenarios Using Single-Link Wi-Fi Channel Measurements," IEEE Sensors Journal, vol. 25, no. 14, pp. 27556–27565, July 2025, doi: 10.1109/JSEN.2025.3576244.

[59] Z. Hao, Y. Duan, X. Dang, Y. Liu, and D. Zhang, "Wi-sl: Contactless fine-grained gesture recognition uses channel state information," Sensors (Switzerland), vol. 20, no. 14, pp. 1–26, 2020, doi: 10.3390/s20144025.

[60] Y. Zhao, R. Gao, S. Liu, L. Xie, J. Wu, H. Tu, and B. Chen, "Device-Free Secure Interaction With Hand Gestures in WiFi-Enabled IoT Environment," IEEE Internet of Things Journal, vol. 8, no. 7, pp. 5619–5631, 2021, doi: 10.1109/JIOT.2020.3032623.

[61] X. Huang and M. Dai, “Indoor Device-Free Activity Recognition Based on Radio Signal,” vol. 66, no. 6, pp. 5316–5329, 2017.

[62] J. Platt, “Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines,” Microsoft, MSR-TR-98-14, Apr. 1998. [Online]. Available: https://www.microsoft.com/en-us/research/publication/sequential-minimal-optimization-a-fast-algorithm-for-training-support-vector-machines/

[63] J. Yan, L. Wan, W. Wei, X. Wu, W. P. Zhu, and D. P. K. Lun, “Device-Free Activity Detection and Wireless Localization Based on CNN Using Channel State Information Measurement,” IEEE Sensors Journal, vol. 21, no. 21, pp. 24482–24494, Nov. 2021, doi: 10.1109/JSEN.2021.3114206.

[64] L. E. Peterson, “K-nearest neighbor,” Scholarpedia, vol. 4, no. 2, p. 1883, 2009.

[65] J. Yang, H. Zou, H. Jiang, and L. Xie, “Device-Free Occupant Activity Sensing Using WiFi-Enabled IoT Devices for Smart Homes,” IEEE Internet of Things Journal, vol. 5, no. 5, pp. 3991–4002, Oct. 2018, doi: 10.1109/JIOT.2018.2849655.

[66] B. Sheng, Y. Fang, F. Xiao, and L. Sun, “An Accurate Device-Free Action Recognition System Using Two-Stream Network,” IEEE Transactions on Vehicular Technology, vol. 69, no. 7, pp. 7930–7939, July 2020, doi: 10.1109/TVT.2020.2993901.

[67] J. Chen, F. Li, H. Chen, S. Yang, and Y. Wang, “Dynamic Gesture Recognition Using Wireless Signals with Less Disturbance,” Personal Ubiquitous Comput., vol. 23, no. 1, pp. 17–27, Feb. 2019, doi: 10.1007/s00779-1182-x.

[68] M. Belgiu and L. Drăguţ, “Random Forest in remote sensing: A review of applications and future directions,” ISPRS Journal of Photogrammetry and Remote Sensing, vol. 114, pp. 24–31, Apr. 2016, doi: 10.1016/j.isprsjprs.2016.01.011.

[69] Y. Wang and Y. Zheng, “Modeling RFID Signal Reflection for Contact-Free Activity Recognition,” Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., vol. 2, no. 4, Dec. 2018, doi: 10.1145/3287071.

[70] S. Sharma, H. Mohammadmoradi, M. Heydariaan, and O. Gnawali, “Device-Free Activity Recognition Using Ultra-Wideband Radios,” in 2019 International Conference on Computing, Networking and Communications (ICNC), pp. 1029–1033, Feb. 2019, doi: 10.1109/ICCNC.2019.8685504.

[71] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," IEEE Transactions on Systems, Man, and Cybernetics, vol. 21, no. 3, pp. 660-674, 1991, doi: 10.1109/21.97458.

[72] Y. Zeng, P. H. Pathak, and P. Mohapatra, "WiWho: WiFi-Based Person Identification in Smart Spaces," in 2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), Apr. 2016, pp. 1-12, doi: 10.1109/IPSN.2016.7460727.

[73] M. G. Moghaddam, A. A. Nazari Shirehjini, and S. Shirmohammadi, "The Effect of Sensor Placement in a Cooking Activity Recognition System," in 2024 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), pp. 1-6, May 2024, doi: 10.1109/I2MTC60896.2024.10560619.

[74] P. Lukowicz, J. A. Ward, H. Junker, M. Stäger, G. Tröster, A. Atrash, and T. Starner, "Recognizing workshop activity using body worn microphones and accelerometers," in International conference on pervasive computing, Berlin, Heidelberg, pp. 18-32, April 2004.

[75] N. Ravi, N. Dandekar, P. Mysore, and M. Littman, "Activity Recognition from Accelerometer Data," in AAAI, Jan. 2005, pp. 1541-1546.

[76] P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea, "Machine Recognition of Human Activities: A Survey," IEEE Transactions on Circuits and Systems for Video Technology, vol. 18, no. 11, pp. 1473-1488, Nov. 2008, doi: 10.1109/TCSVT.2008.2005594.

[77] S. Patil and A. Kobsa, "Privacy Considerations in Awareness Systems: Designing with Privacy in Mind," in Awareness systems: advances in theory, methodology and design, pp. 187-206, 2009, doi: 10.1007/978-1-84882-477-5_8.

[78] B. J. Rhodes, "The wearable remembrance agent: A system for augmented memory," Personal Technologies, vol. 1, no. 4, pp. 218-224, Dec. 1997, doi: 10.1007/BF01682024.

[79] M. Buettner, R. Prasad, M. Philipose, and D. Wetherall, "Recognizing daily activities with RFID-based sensors," in Proceedings of the 11th international conference on Ubiquitous computing, pp. 51-60, 2009, doi: 10.1145/1620545.1620553.

[80] S. Sigg, M. Scholz, S. Shi, Y. Ji, and M. Beigl, "RF-Sensing of Activities from Non-Cooperative Subjects in Device-Free Recognition Systems Using Ambient and Local Signals," IEEE Transactions on Mobile Computing, vol. 13, no. 4, pp. 907-920, Apr. 2014, doi: 10.1109/TMC.2013.28.

[81] M. Scholz, S. Sigg, H. R. Schmidtke, and M. Beigl, "Challenges for device-free radio-based activity recognition," in Workshop on Context Systems, Design, Evaluation and Optimization, 2011.

[82] P. Cassara, F. Potortı, P. Barsocchi, and P. Nepa, "Single-channel versus multi-channel scanning in device-free indoor radio localization," in Proceedings of the 2013 International Conference on Indoor

Positioning and Indoor Navigation (IPIN), 2013.

[83] D. Stødle, "Device-Free Interaction and Cross-Platform Pixel Based Output to Display Walls," Ph.D. dissertation, Dept. Comput. Sci., Univ. Tromsø, Tromsø, Norway, 2009. [Online]. Available: https://www.cs.uit.no/~daniels/stoedle-dissertation.pdf

[84] A. E. Kosba, A. Saeed, and M. Youssef, "RASID: A robust WLAN device-free passive motion detection system," in 2012 IEEE International Conference on Pervasive Computing and Communications, pp. 180-189, Mar. 2012, doi: 10.1109/PerCom.2012.6199865.

[85] M. Vrigkas, C. Nikou, and I. A. Kakadiaris, "A review of human activity recognition methods," Frontiers in Robotics and AI, vol. 2, p. 28, 2015.

[86] Q. Zhou, J. Xing, and Q. Yang, "Device-free occupant activity recognition in smart offices using intrinsic Wi-Fi components," Building and Environment, vol. 172, p. 106737, Apr. 2020, doi: 10.1016/j.buildenv.2020.106737.

[87] F. Wang, J. Liu, and W. Gong, "Multi-Adversarial In-Car Activity Recognition Using RFIDs," IEEE Transactions on Mobile Computing, vol. 20, no. 6, pp. 2224-2237, June 2021, doi: 10.1109/TMC.2020.2977902.

[88] K. Muthukrishnan, M. Lijding, N. Meratnia, and P. Havinga, "Sensing Motion Using Spectral and Spatial Analysis of WLAN RSSI," in Smart Sensing and Context, Berlin, Heidelberg, vol. 4793, pp. 62-76, 2007, doi: 10.1007/978-3-540-75696-5_4.

[89] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu, "E-eyes: device-free location-oriented activity identification using fine-grained WiFi signatures," in Proceedings of the 20th annual international conference on Mobile computing and networking, in MobiCom '14, New York, NY, USA, pp. 617-628, Sept. 2014, doi: 10.1145/2639108.2639143.

[90] Z. Yang, Z. Zhou, and Y. Liu, "From RSSI to CSI: Indoor localization via channel response," ACM Computing Surveys, vol. 46, no. 2, pp. 1-32, 2013.

[91] G. Wang, Y. Zou, Z. Zhou, K. Wu, and L. M. Ni, "We can hear you with Wi-Fi!," in Proceedings of the 20th annual international conference on Mobile computing and networking, in MobiCom '14, New York, NY, USA, pp. 593-604, Sept. 2014, doi: 10.1145/2639108.2639112.

[92] S. Yao, S. Hu, Y. Zhao, A. Zhang, and T. Abdelzaher, "DeepSense: A Unified Deep Learning Framework for Time-Series Mobile Sensing Data Processing," in Proceedings of the 26th International Conference on World Wide Web, pp. 351-360, Apr. 2017, doi: 10.1145/3038912.3052577.

[93] X. Liu, J. Cao, S. Tang, J. Wen, and P. Guo, "Contactless Respiration Monitoring Via Off-the-Shelf WiFi Devices," IEEE Transactions on Mobile Computing, vol. 15, no. 10, pp. 2466-2479, Oct. 2016, doi: 10.1109/TMC.2015.2504935.

[94] H. Wang, D. Zhang, Y. Wang, J. Ma, Y. Wang, and S. Li, "RT-Fall: A Real-Time and Contactless Fall Detection System with Commodity WiFi Devices," IEEE Transactions on Mobile Computing, vol. 16, no. 2, pp. 511-526, Feb. 2017, doi: 10.1109/TMC.2016.2557795.

[95] N. Paulino, M. Oliveira, F. Ribeiro, L. Outeiro, P. A. Lopes, F. Vilarinho, S. Inácio, and L. M. Pessoa, "Human Activity Recognition with a 6.5 GHz Reconfigurable Intelligent Surface for Wi-Fi 6E," Jan. 24, 2025, arXiv:2501.14423, doi: 10.48550/arXiv.2501.14423.

[96] Y. Hu, F. Zhang, C. Wu, B. Wang, and K. J. R. Liu, "DeFall: Environment-Independent Passive Fall Detection Using WiFi," IEEE Internet of Things Journal, vol. 9, no. 11, pp. 8515-8530, June 2022, doi: 10.1109/JIOT.2021.3116136.

[97] B. Yu, Y. Wang, K. Niu, Y. Zeng, T. Gu, L. Wang, C. Guan, and D. Zhang, "WiFi-Sleep: Sleep Stage Monitoring Using Commodity Wi-Fi Devices," IEEE Internet of Things Journal, vol. 8, no. 18, pp. 13900-13913, Sept. 2021, doi: 10.1109/JIOT.2021.3068798.

[98] W. Liang, R. Tang, S. Jiang, R. Wang, Y. Zhao, C. Z. Xu, X. Long, Z. Chen, and X. Li, "LiWi-HAR: Lightweight WiFi-Based Human Activity Recognition Using Distributed AIoT," IEEE Internet of Things Journal, vol. 11, no. 1, pp. 597-611, Jan. 2024, doi: 10.1109/JIOT.2023.3286455.

[99] B. Li, W. Cui, L. Zhang, Q. Yang, M. Wu, and J. T. Zhou, "Democratizing Federated WiFi-Based Human Activity Recognition Using Hypothesis Transfer," IEEE Transactions on Mobile Computing, vol. 23, no. 12, pp. 15132-15148, Dec. 2024, doi: 10.1109/TMC.2024.3457788.

[100] Y. Pan, Z. Zhou, W. Gong, and Y. Fang, "SAT: A Selective Adversarial Training Approach for WiFi-Based Human Activity Recognition," IEEE Transactions on Mobile Computing, vol. 23, no. 12, pp. 12706-12716, 2024.

[101] B. N. Schilit and M. M. Theimer, "Disseminating active map information to mobile hosts," IEEE Network, vol. 8, no. 5, pp. 22-32, Sept. 1994, doi: 10.1109/65.313011.

[102] Y. Asim, M. A. Azam, M. Ehatisham-ul-Haq, U. Naeem, and A. Khalid, "Context-Aware Human Activity Recognition (CAHAR) in-the-Wild Using Smartphone Accelerometer," IEEE Sensors Journal, vol. 20, no. 8, pp. 4361-4371, Apr. 2020, doi: 10.1109/JSEN.2020.2964277.

[103] B. Xu, A. Bikakis, D. Onah, A. Vlachidis, and L. Dickens, "Context Helps: Integrating Context

Information with Videos in a Graph-Based HAR Framework," in International Conference on Neural-Symbolic Learning and Reasoning, pp. 3-28, Sept. 2024, doi: 10.1007/978-3-031-71167-1_1.

[104] S. Zhang, Y. Zhong, H. Jia, X. Ding, and T. Jiang, "Cross-Environment Device-Free Human Action Recognition via Wi-Fi Signals," Electronics, vol. 14, no. 11, June 2025, doi: 10.3390/electronics14112299.

[105] W. Min, S. Jiang, L. Liu, Y. Rui, and R. Jain, "A survey on food computing," ACM Computing Surveys (CSUR), vol. 52, no. 5, pp. 1-36, 2019.

[106] M. G. Moghaddam, A. A. Nazari Shirehjini, and S. Shirmohammadi, "Device-Free Fine-Grained Dining Activity Sensing," in 2023 IEEE Sensors Applications Symposium (SAS), pp. 1-6, July 2023, doi: 10.1109/SAS58821.2023.10254010.

[107] M. G. Moghaddam, A. A. N. Shirehjini, and S. Shirmohammadi, "The Effect of Activity Granularity on A Kitchen Activity Recognition System," in 2024 IEEE Sensors Applications Symposium (SAS), pp. 1-5, July 2024, doi: 10.1109/SAS60918.2024.10636508.

[108] M. A. Samir, Z. A. Mohamed, M. A. A. Hussein, and A. Atia, "Applying Deep Learning to Track Food Consumption and Human Activity for Non-intrusive Blood Glucose Monitoring," in 2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, NY, USA, pp. 0319-0324, Dec. 2021, doi: 10.1109/UEMCON53757.2021.9666662.

[109] D. Bhattacharya, D. Sharma, W. Kim, M. F. Ijaz, and P. K. Singh, "Ensem-HAR: An Ensemble Deep Learning Model for Smartphone Sensor-Based Human Activity Recognition for Measurement of Elderly Health Monitoring," Biosensors, vol. 12, no. 6, p. 393, June 2022, doi: 10.3390/bios12060393.

[110] A. Snoun, T. Bouchrika, and O. Jemai, "Deep-learning-based human activity recognition for Alzheimer's patients' daily life activities assistance," Neural Comput & Applic, vol. 35, no. 2, pp. 1777-1802, Jan. 2023, doi: 10.1007/s00521-022-07883-1.

[111] C. Pham and P. Olivier, "Slice&dice: Recognizing food preparation activities using embedded accelerometers," in European Conference on Ambient Intelligence, Berlin, Heidelberg, pp. 34-43, November 2009.

[112] Y. Mohammad, K. Matsumoto, and K. Hoashi, "A dataset for activity recognition in an unmodified kitchen using smart-watch accelerometers," in Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia, pp. 63-68, Nov. 2017.

[113] J. Monteiro, R. Granada, R. C. Barros, and F. Meneguzzi, "Deep neural networks for kitchen activity

recognition," in 2017 International Joint Conference on Neural Networks (IJCNN), May 2017, pp. 2048-2055, doi: 10.1109/IJCNN.2017.7966102.

[114] S. Bansal, S. Khandelwal, S. Gupta, and D. Goyal, "Kitchen activity recognition based on scene context," in 2013 IEEE International Conference on Image Processing, pp. 3461-3465, Sep. 2013.

[115] S. Zolfaghari, T. Stoev, and K. Yordanova, "Enhancing Kitchen Activity Recognition: A Benchmark Study of the Rostock KTA Dataset," IEEE Access, vol. 12, pp. 14364-14384, 2024, doi: 10.1109/ACCESS.2024.3356352.

[116] M. Liu, S. Suh, J. F. Vargas, B. Zhou, A. Grünerbl, and P. Lukowicz, "A Wearable Multi-modal Edge-Computing System for Real-Time Kitchen Activity Recognition," in International Joint Conference on Artificial Intelligence, Singapore, pp. 132-145, Nov. 2024, doi: 10.1007/978-981-97-9003-6_9.

[117] Y. H. Wu, Y. Chen, S. Shirmohammadi, and C. H. Hsu, "AI-Assisted Food Intake Activity Recognition Using 3D mmWave Radars," in Proceedings of the 7th International Workshop on Multimedia Assisted Dietary Management, pp. 81-89, 2022.

[118] Y. H. Yeh, D. P. Wong, C. T. Lee, and P. H. Chou, "Deep Learning-Based Real-Time Activity Recognition with Multiple Inertial Sensors," in 4th International Conference on Image, Video and Signal Processing, pp. 92-99, 2022.

[119] L. Arrotta, G. Civitarese, and C. Bettini, "DeXAR: Deep Explainable Sensor-Based Activity Recognition in Smart-Home Environments," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 6, pp. 1-30, 2022.

[120] S. Mekruksavanich, P. Jantawong, N. Hnoohom, and A. Jitpattanakul, "Deep Learning Networks for Eating and Drinking Recognition based on Smartwatch Sensors," in Research, Invention, and Innovation Congress: Innovative Electricals and Electronics (RI2C), Bangkok, Thailand, pp. 106-111, 2022.

[121] L. Arrotta, C. Bettini, and G. Civitarese, "MICAR: multi-inhabitant context-aware activity recognition in home environments," in Distrib Parallel Databases, vol. 31, no. 4, pp. 571-602, 2022.

[122] J. Maitre, K. Bouchard, C. Bertuglia, and S. Gaboury, "Recognizing activities of daily living from UWB radars and deep learning," Expert Systems with Applications, vol. 164, p. 113994, 2021.

[123] L. G. Fahad and S. F. Tahir, "Activity recognition and anomaly detection in smart homes," Neurocomputing, vol. 423, pp. 362-372, Jan. 2021.

[124] S. Stankoski, M. Jordan, H. Gjoreski, and M. Luštrek, "Smartwatch-Based Eating Detection: Data

Selection for Machine Learning from Imbalanced Data with Imperfect Labels," Sensors, vol. 21, no. 5, 2021.

[125] L. G. Fahad and S. F. Tahir, "Activity recognition in a smart home using local feature weighting and variants of nearest-neighbors classifiers," Journal of Ambient Intelligence and Humanized Computing, vol. 12, no. 2, pp. 2355-2364, 2021.

[126] C. M. Ranieri, S. Macleod, M. Dragone, P. A. Vargas, and R. A. F. Romero, "Activity Recognition for Ambient Assisted Living with Videos, Inertial Units and Ambient Sensors," Sensors, vol. 21, no. 3, 2021.

[127] A. Gorji, T. Gielen, M. Bauduin, H. Sahli, and A. Bourdoux, "A Multi-radar Architecture for Human Activity Recognition in Indoor Kitchen Environments," in 2021 IEEE Radar Conference (RadarConf21), Atlanta, GA, USA, pp. 1-6, May 2021, doi: 10.1109/RadarConf2147009.2021.9455238.

[128] M. Garcia-Constantino, A. Konios, M. A. Mustafa, C. Nugent, and G. Morrison, "Ambient and Wearable Sensor Fusion for Abnormal Behaviour Detection in Activities of Daily Living," in IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), pp. 1-6, 2020.

[129] F. A. Machot, M. R. Elkobaisi, and K. Kyamakya, "Zero-shot human activity recognition using non-visual sensors," Sensors (Switzerland), vol. 20, no. 3, pp. 1-17, 2020, doi: 10.3390/s20030825.

[130] E. Gambi, G. Temperini, R. Galassi, L. Senigagliesi, and A. D. Santis, "ADL Recognition Through Machine Learning Algorithms on IoT Air Quality Sensor Dataset," IEEE Sensors Journal, vol. 20, no. 22, pp. 13562-13570, 2020.

[131] W. Zhang, T. Zhu, C. Yang, J. Xiao, and H. Ning, "Sensors-based Human Activity Recognition with Convolutional Neural Network and Attention Mechanism," in IEEE 11th International Conference on Software Engineering and Service Science (ICSESS), pp. 158-162, 2020.

[132] L. Arrotta, C. Bettini, G. Civitarese, and R. Presotto, "Context-Aware Data Association for Multi-Inhabitant Sensor-Based Activity Recognition," in 21st IEEE International Conference on Mobile Data Management (MDM), pp. 125-130, 2020.

[133] K. Viard, M. P. Fanti, G. Faraut, and J. J. Lesage, "Human activity discovery and recognition using probabilistic finite-state automata," IEEE Transactions on Automation Science and Engineering, vol. 17, no. 4, pp. 2085-2096, 2020.

[134] G. B. Brahim, W. El-Hajj, C. El-Hayek, and H. Hajj, "A New Semantic-based Multi-Level

Classification Approach for Activity Recognition Using Smartphones," International Journal of Software Engineering and Knowledge Engineering, vol. 30, no. 8, pp. 1051-1078, 2020.

[135] D. Liciotti, M. Bernardini, L. Romeo, and E. Frontoni, "A sequential deep learning application for recognising human activities in smart homes," Neurocomputing, vol. 396, pp. 501-513, July 2020, doi: 10.1016/j.neucom.2018.10.104.

[136] B. X. Yu, Y. Liu, and K. C. Chan, "Skeleton Focused Human Activity Recognition in RGB Video", arXiv preprint arXiv:2004.13979, 2020.

[137] Ö. F. İnce, I. F. Ince, M. E. Yıldırım, J. S. Park, J. K. Song, and B. W. Yoon, "Human activity recognition with analysis of angles between skeletal joints using an RGB-depth sensor," ETRI Journal, vol. 42, no. 1, pp. 78-89, Feb. 2020, doi: 10.4218/etrij.2018-0577.

[138] P. Asghari, E. Soleimani, and E. Nazerfard, "Online human activity recognition employing hierarchical hidden Markov models," Journal of Ambient Intelligence and Humanized Computing, vol. 11, no. 3, pp. 1141-1152, 2020.

[139] S. Sen, V. Subbaraju, A. Misra, R. Balan, and Y. Lee, "Annapurna: An automated smartwatch-based eating detection and food journaling system," Pervasive and Mobile Computing, vol. 68, 2020.

[140] S. Stankoski, N. Resçiç, G. Mezic, and M. Lustrek, "Real-time Eating Detection Using a Smartwatch," in EWSN, pp. 247-252, 2020.

[141] T. Hossain, M. S. Islam, M. A. R. Ahad, and S. Inoue, "Human activity recognition using wearable devices," in Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers, pp. 81-84, 2019.

[142] M. R. Alam, M. B. I. Reaz, and M. A. M. Ali, "A Review of Smart Homes—Past, Present, and Future," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 42, no. 6, pp. 1190-1203, Nov. 2012, doi: 10.1109/TSMCC.2012.2189204.

[143] O. Taiwo and A. E. Ezugwu, "Internet of things-based intelligent smart home control system," Security and Communication Networks, no. 1, p. 9928254, 2021.

[144] M. Li and Y. Wu, "Intelligent control system of smart home for context awareness," International Journal of Distributed Sensor Networks, vol. 18, no. 3, p. 15501329221082030, 2022.

[145] D. B. Kaber and M. R. Endsley, "The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task," Theoretical Issues in

Ergonomics Science, vol. 5, no. 2, pp. 113-153, Mar. 2004, doi: 10.1080/1463922021000054335.

[146] E. Tural, D. Lu, and D. A. Cole, "Safely and actively aging in place: Older adults' attitudes and intentions toward smart home technologies," Gerontology and geriatric medicine, vol. 7, p. 23337214211017340, 2021.

[147] J. Wang, N. Spicher, J. M. Warnecke, M. Haghi, J. Schwartze, and T. M. Deserno, "Unobtrusive Health Monitoring in Private Spaces: The Smart Home," Sensors, vol. 21, no. 3, Jan. 2021, doi: 10.3390/s21030864.

[148] N. Y. Philip, J. J. P. C. Rodrigues, H. Wang, S. J. Fong, and J. Chen, "Internet of Things for In-Home Health Monitoring Systems: Current Advances, Challenges and Future Directions," IEEE Journal on Selected Areas in Communications, vol. 39, no. 2, pp. 300-310, Feb. 2021, doi: 10.1109/JSAC.2020.3042421.

[149] L. Babangida, T. Perumal, N. Mustapha, and R. Yaakob, "Internet of Things (IoT) Based Activity Recognition Strategies in Smart Homes: A Review," IEEE sensors journal, vol. 22, no. 9, pp. 8327-8336, 2022.

[150] P. Kulurkar, C. kumar Dixit, V. C. Bharathi, A. Monikavishnuvarthini, A. Dhakne, and P. Preethi, "AI based elderly fall prediction system using wearable sensors: A smart home-care technology with IOT," Measurement: Sensors, vol. 25, p. 100614, Feb. 2023, doi: 10.1016/j.measen.2022.100614.

[151] J. Liu, C. Xiao, K. Cui, J. Han, X. Xu, and K. Ren, “Behavior Privacy Preserving in RF Sensing,” IEEE Transactions on Dependable and Secure Computing, vol. 20, no. 1, pp. 784–796, Jan. 2023, doi: 10.1109/TDSC.2022.3143880.

[152] J. Liu, J. Wang, Q. Gao, X. Li, M. Pan, and Y. Fang, “Diversity-Enhanced Robust Device-Free Vital Signs Monitoring Using mmWave Signals,” IEEE Transactions on Mobile Computing, vol. 23, no. 12, pp. 12922–12937, Dec. 2024, doi: 10.1109/TMC.2024.3419005.

[153] A. A. N. Shirehjini, “Methods, arrangements and system related to exchanging user-related data,” U.S. Patent Application US20200104862A1, Apr. 02, 2020.

[154] S. Suherman, “WiFi-Friendly Building to Enable WiFi Signal Indoor,” Bulletin of Electrical Engineering and Informatics, vol. 7, no. 2, pp. 264–271, June 2018, doi: 10.11591/eei.v7i2.871.

[155] D. Zhang, J. Wang, J. Jang, J. Zhang, and S. Kumar, “On the Feasibility of Wi-Fi Based Material Sensing,” in The 25th Annual International Conference on Mobile Computing and Networking, New York, NY, USA, pp. 1–16, Oct. 2019, doi: 10.1145/3300061.3345442.

[156] M. Cimdins, S. O. Schmidt, F. John, M. Constapel, and H. Hellbrück, “MA-RTI: Design and Evaluation of a Real-World Multipath-Assisted Device-Free Localization System,” Sensors, vol. 23, no. 4, p. 2199, Feb. 2023, doi: 10.3390/s23042199.

[157] P. Neupane, G. Liu, H. C. Wu, S. Y. Chang, and J. Ye, “Novel optimal multisensor placement for indoor rectilinear line-of-sight coverage,” IEEE Sensors Journal, vol. 21, no. 20, pp. 23435–23451, 2021.

[158] Espressif Systems, “ESP32 Series of SoCs,” Espressif Systems (Shanghai) Co., Ltd., 2022. [Online]. Available: https://www.espressif.com/en/products/socs/esp32

[159] H. F. T. Ahmed, H. Ahmad, K. Narasingamurthi, H. Harkat, and S. K. Phang, “DF-WiSLR: Device-Free Wi-Fi-based Sign Language Recognition,” Pervasive and Mobile Computing, vol. 69, 2020, doi: 10.1016/j.pmcj.2020.101289.

[160] J. Wang, X. Zhang, Q. Gao, X. Ma, X. Feng, and H. Wang, “Device-Free Simultaneous Wireless Localization and Activity Recognition With Wavelet Feature,” IEEE Transactions on Vehicular Technology, vol. 66, no. 2, pp. 1659–1669, Feb. 2017, doi: 10.1109/TVT.2016.2555985.

[161] H. Ding and W. Zhu, “Activity classification of the elderly based on lightweight convolutional neural networks,” ACM International Conference Proceeding Series, pp. 186–190, 2022, doi: 10.1145/3581807.3581834.

[162] Z. Chen, C. Cai, T. Zheng, J. Luo, J. Xiong, and X. Wang, “RF-Based Human Activity Recognition Using Signal Adapted Convolutional Neural Network,” IEEE Transactions on Mobile Computing, vol. 22, no. 1, pp. 487-499, 2021, doi: 10.1109/TMC.2021.3073969.

[163] J. Zhang, B. Wei, and J. Cheng, “HARaaS: HAR as a Service Using Wifi Signal in IoT-Enabled Edge Computing: Poster Abstract,” in Proceedings of the 18th Conference on Embedded Networked Sensor Systems, New York, NY, USA, pp. 681–682, 2020.

[164] Y. Liu, W. Huang, S. Jiang, B. Zhao, S. Wang, S. Wang, and Y. Zhang, “TransTM: A device-free method based on time-streaming multiscale transformer for human activity recognition,” Defence Technology, vol. 32, pp.619-628, 2023, doi: https://doi.org/10.1016/j.dt.2023.02.021.

[165] Y. Zheng, Y. Zhang, K. Qian, G. Zhang, Y. Liu, C. Wu, and Z. Yang, “Zero-Effort Cross-Domain Gesture Recognition with Wi-Fi,” in Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services, New York, NY, USA, pp. 313–325, June 2019, doi: 10.1145/3307334.3326081.

[166] T. Huang, S. Wang, and S. Li, “Low-Cost and User Independent Action Recognition Using Wifi

Signals," SSRN Electronic Journal, pp. 1–30, 2022, doi: 10.2139/ssrn.4160593.

[167] J. S. Kim, W. S. Jeon, and D. G. Jeong, "WiFi-Based Low-Complexity Gesture Recognition Using Categorization," in 2022 IEEE 95th Vehicular Technology Conference: (VTC2022-Spring), pp. 1–7, 2022, doi: 10.1109/VTC2022-Spring54318.2022.9860660.

[168] H. Raeis, M. Kazemi, and S. Shirmohammadi, "InARMS: Individual Activity Recognition of Multiple Subjects with FMCW radar," in 2022 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), IEEE, pp. 1–5, May 2022, doi: 10.1109/I2MTC48687.2022.9806694.

[169] L. Babangida, T. Perumal, N. Mustapha, and R. Yaakob, "Internet of things ({IoT}) based activity recognition strategies in smart homes: A review," IEEE Sensors Journal, vol. 22, no. 9, pp. 8327–8336, 2022.

[170] C. Jin, X. Meng, X. Li, J. Wang, M. Pan, and Y. Fang, "Rodar: Robust Gesture Recognition Based on mmWave Radar Under Human Activity Interference," IEEE Transactions on Mobile Computing, pp. 1–14, 2024, doi: 10.1109/TMC.2024.3402356.

[171] P. S. Santhalingam, A. A. Hosain, D. Zhang, P. Pathak, H. Rangwala, and R. Kushalnagar, "MmASL: Environment-Independent ASL Gesture Recognition Using 60 GHz Millimeter-Wave Signals," Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., vol. 4, no. 1, Mar. 2020, doi: 10.1145/3381010.

[172] J. Wu, C. Harrison, J. P. Bigham, and G. Laput, "Automated Class Discovery and One-Shot Interactions for Acoustic Activity Recognition," in Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, New York, NY, USA, pp. 1–14, 2020.

[173] R. Alazrai, A. Awad, A. Baha'A, M. Hababeh, and M. I. Daoud, "A dataset for wi-fi-based human-to-human interaction recognition," Data in Brief, Elsevier, vol. 31, 2020, doi: 10.1016/j.dib.2020.105668.

[174] S. Sigg, U. Blanke, and G. Tröster, "The telepathic phone: Frictionless activity recognition from wifi-rssi," in IEEE international conference on pervasive computing and communications (PerCom), pp. 148-155, 2014.

[175] M. G. Moghaddam, A. A. N. Shirehjini, and S. Shirmohammadi, "Wi-Fi (CSI and RSSI) data of six basic knife activities for cooking (chopping, cubing, French cutting, julienning, mincing, and slicing)." Zenodo, Apr. 18, 2023. doi: 10.5281/ZENODO.7843704.

[176] S. M. Hernandez, D. Erdag, and E. Bulut, "Towards dense and scalable soil sensing through low-cost WiFi sensing networks," in 2021 IEEE 46th Conference on Local Computer Networks (LCN), pp.

549–556, 2021.

[177] M. G. Moghaddam, A. A. N. Shirehjini and S. Shirmohammadi, "Device-free Kitchen Activity Recognition via AI-assisted Wireless Sensing," in IEEE Transactions on Instrumentation and Measurement, Submitted.

[178] M. G. Moghaddam, A. A. N. Shirehjini, and S. Shirmohammadi, “A WiFi-based System for Recognizing Fine-grained Multiple-Subject Human Activities,” in IEEE International Instrumentation and Measurement Technology Conference (I2MTC), pp. 1–6, May 2022.

[179] S. Pandurangan, M. Papandrea, and M. Gelsomini, “Fine-Grained Human Activity Recognition - A new paradigm,” pp. 1–8, 2022, doi: 10.1145/3558884.3558893.

[180] Espressif Systems, “ESP32-DevKitC VIE Development Board,” Espressif Systems (Shanghai) Co., Ltd., 2023. [Online]. Available: https://www.espressif.com/en/products/devkits/esp32-devkitc.

[181] Espressif Systems, “Get Started with ESP-IDF,” ESP-IDF Programming Guide, 2023. [Online]. Available: https://docs.espressif.com/projects/esp-idf/en/stable/get-started/index.html.

[182] D. Lyon, “The Discrete Fourier Transform, Part 4: Spectral Leakage,” Journal of Object Technology, vol. 8, no. 7, Nov. 2009, doi: 10.5381/jot.2009.8.7.c2.

[183] S. G. Dhekane and T. Ploetz, “Transfer Learning in Human Activity Recognition: A Survey,” Jan. 18, 2024, arXiv: arXiv:2401.10185. doi: 10.48550/arXiv.2401.10185.

[184] F. Bartoli, G. Lisanti, L. Ballan, and A. Del Bimbo, “Context-Aware Trajectory Prediction,” in 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, pp. 1941–1946, Aug. 2018, doi: 10.1109/ICPR.2018.8545447.

[185] A. Esmaeilzehi, E. Khazaei, K. Wang, N. K. Kalsi, P. C. Ng, H. Liu, Y. Yu, D. Hatzinakos, and K. Plataniotis, “HARWE: A multi-modal large-scale dataset for context-aware human activity recognition in smart working environments,” Pattern Recognition Letters, vol. 184, pp. 126–132, Aug. 2024, doi: 10.1016/j.patrec.2024.06.017.

[186] J. Yang, Y. Xu, H. Cao, H. Zou, and L. Xie, “Deep learning and transfer learning for device-free human activity recognition: A survey,” Journal of Automation and Intelligence, vol. 1, no. 1, p. 100007, Dec. 2022, doi: 10.1016/j.jai.2022.100007.

[187] J. Schäfer, B. R. Barrsiwal, M. Kokhkharova, H. Adil, and J. Liebehenschel, “Human activity recognition using csi information with nexmon,” Applied Sciences, Switzerland, vol. 11, no. 19, 2021, doi: 10.3390/app11198860.

[188] C. P. Dautov and M. S. Ozerdem, “Wavelet transform and signal denoising using Wavelet method,” in 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, Turkey, pp. 1–4, May 2018, doi: 10.1109/SIU.2018.8404418.

[189] M. G. Moghaddam, A. R. Nowdeh, N. Warnakulasuriya, G. S. S. P. Pala, A. A. N. Shirehjini, and S. Shirmohammadi, “Wi-GRAK,” Kaggle, 2025, doi: 10.34740/KAGGLE/DSV/12123800.

[190] M. G. Moghaddam, A. R. Nowdeh, N. Warnakulasuriya, G. S. S. P. Pala, A. A. N. Shirehjini, and S. Shirmohammadi, “Descriptor: WiFi-Granular Activities in Kitchen Dataset (Wi-GRAK),” IEEE Data Descriptions, 2025.

[191] R. G. O’Brien and M. K. Kaiser, “MANOVA method for analyzing repeated measures designs: An extensive primer,” Psychological Bulletin, vol. 97, no. 2, pp. 316–333, 1985, doi: 10.1037/0033-2909.97.2.316.

[192] T. K. Nguyen, V. D. Ly, and S. O. Hwang, “Effective feature selection based on MANOVA,” International Journal of Internet Technology and Secured Transactions, vol. 10, no. 4, Apr. 2020.

[193] M. Bejani, D. Gharavian, and N. M. Charkari, “Audiovisual emotion recognition using ANOVA feature selection method and multi-classifier neural networks,” Neural Comput & Applic, vol. 24, no. 2, pp. 399–412, Feb. 2014, doi: 10.1007/s00521-012-1228-3.

[194] T. Haider, M. H. Khan, and M. S. Farid, “An Optimal Feature Selection Method for Human Activity Recognition Using Multimodal Sensory Data,” Information, vol. 15, no. 10, Oct. 2024, doi: 10.3390/info15100593.

[195] X. Cao, B. Chen, and Y. Zhao, “Wi-Wri: Fine-Grained Writing Recognition Using Wi-Fi Signals,” in 2016 IEEE Trustcom/BigDataSE/ISPA, pp. 1366–1373, Aug. 2016, doi: 10.1109/TrustCom.2016.0216.

[196] C. Wang, J. Liu, Y. Chen, H. Liu, L. Xie, W. Wang, B. He, and S. Lu, “Multi - Touch in the Air: Device-Free Finger Tracking and Gesture Recognition via COTS RFID,” In IEEE Conference on Computer Communications (INFOCOM), pp. 1691–1699, 2018, doi: 10.1109/INFOCOM.2018.8486346.

[197] Q. Zhou, J. Xing, J. Li, and Q. Yang, “A Device-Free Number Gesture Recognition Approach Based on Deep Learning,” in 12th International Conference on Computational Intelligence and Security (CIS), Dec. 2016, pp. 57–63. doi: 10.1109/CIS.2016.0022.

[198] D. Wu, R. Gao, Y. Zeng, J. Liu, L. Wang, T. Gu, and D. Zhang, “FingerDraw: Sub-Wavelength Level Finger Motion Tracking with WiFi Signals,” Proc. ACM Interact. Mob. Wearable Ubiquitous

Technol., vol. 4, no. 1, Mar. 2020, doi: 10.1145/3380981.

[199] Y. Zhong, Y. Huang, and T. Jiang, “Device-Free Sensing for Gesture Recognition by Wi-Fi Communication Signal Based on Auto-encoder/decoder Neural Network,” Lecture Notes in Electrical Engineering (LNEE), vol. 571, pp. 887–894, 2020, doi: 10.1007/978-981-13-9409-6_103.

[200] F. Wang, W. Gong, and J. Liu, “On Spatial Diversity in WiFi-Based Human Activity Recognition: A Deep Learning-Based Approach,” IEEE Internet of Things Journal, vol. 6, no. 2, pp. 2035–2047, Apr. 2019, doi: 10.1109/JIOT.2018.2871445.

[201] S. Tan, L. Zhang, Z. Wang, and J. Yang, “MultiTrack: Multi-User Tracking and Activity Recognition Using Commodity WiFi,” in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, in CHI ’19. New York, NY, USA, pp. 1–12, 2019, doi: 10.1145/3290605.3300766.

[202] S. M. Hernandez and E. Bulut, “Lightweight and Standalone IoT Based WiFi Sensing for Active Repositioning and Mobility,” in 2020 IEEE 21st International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM), pp. 277–286, Aug. 2020, doi: 10.1109/WoWMoM49955.2020.00056.

[203] C. Liu, J. Xiong, L. Cai, L. Feng, X. Chen, and D. Fang, “Beyond Respiration: Contactless Sleep Sound-Activity Recognition Using RF Signals,” Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., vol. 3, no. 3, Sept. 2019, doi: 10.1145/3351254.

[204] Z. Lin, Y. Xie, X. Guo, Y. Ren, Y. Chen, and C. Wang, “WiEat: Fine-grained Device-free Eating Monitoring Leveraging Wi-Fi Signals,” in 29th International Conference on Computer Communications and Networks (ICCCN), pp. 1–9, Aug. 2020, doi: 10.1109/ICCCN49398.2020.9209628.

[205] Q. Chen, B. Tan, K. Chetty, and K. Woodbridge, “Activity recognition based on micro-Doppler signature with in-home Wi-Fi,” in IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom), pp. 1–6, 2016, doi: 10.1109/HealthCom.2016.7749457.

[206] Q. Li, M. Li, X. Ke, W. Kong, L. Yang, Z. Tang, X. Chen, and D. Fang, “WisDriver:A WiFi and Smartphone Sensing System for Safely Driving,” in IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM), pp. 1–6, Aug. 2019, doi: 10.1109/PACRIM47961.2019.8985091.

[207] W. Sun, T. Chen, J. Zheng, Z. Lei, L. Wang, B. Steeper, P. He, M. Dressa, F. Tian, and C. Zhang, “VibroSense: Recognizing Home Activities by Deep Learning Subtle Vibrations on an Interior Surface of a House from a Single Point Using Laser Doppler Vibrometry,” Proc. ACM Interact. Mob. Wearable

Ubiquitous Technol., vol. 4, no. 3, Sept. 2020, doi: 10.1145/3411828.

[208] L. W. Lu, C. Lin, X. Fan, B. Han, X. Han, and Z. Qin, "Autodlar: a semi-supervised cross-modal contact-free human activity recognition system," ACM Transactions on Sensor Networks, vol. 20, no. 4, pp. 1-20, 2024.

[209] H. R. Hashima and H. Yamaguchi, "The Future of Beyond 5G Sensing: Transforming Activity Recognition with Reconfigurable Intelligent Surfaces," in International Conference on Activity and Behavior Computing (ABC), pp. 1-10, 2024.

[210] X. Yang, J. Sayono, and Y. Zhang, "CubeSense++: Smart environment sensing with interaction-powered corner reflector mechanisms," In Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, pp. 1-12, 2023

[211] H. F. T. Ahmed, H. Ahmad, S. K. Phang, C. A. Vaithilingam, H. Harkat, and K. Narasingamurthi, "Higher order feature extraction and selection for robust human gesture recognition using CSI of COTS Wi-Fi devices," Sensors, vol. 19, no. 13, pp. 2959, 2019.

[212] X. Li, J. Hu, and J. Luo, "Efficient Beamforming Feedback Information-Based Wi-Fi Sensing by Feature Selection," IEEE Wireless Communications Letters, vol. 13, no. 9, pp. 2347-2351, 2024.