Data in Brief 57 (2024) 111101

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Radio frequency-based human activity dataset collected using ESP32 microcontroller in line-of-sight and non-line-of-sight indoor experiment setups

Zhe-Yu Lim, Lee-Yeng Ong*, Meng-Chew Leow

Faculty of Information Science and Technology, Multimedia University Melaka Campus, Jalan Ayer Keroh Lama, 75450 Melaka, Malaysia

A R T I C L E   I N F O

Article history:
Received 17 July 2024
Revised 2 October 2024
Accepted 28 October 2024
Available online 3 November 2024

Dataset link: Radio Frequency-based Human Activity Dataset (Original data)

Keywords:
Radio frequency
Received signal strength indicator (RSSI)
Channel state information (CSI)
Human activity detection
Line-of-sight (LOS)
Non-line-of-sight (NLOS)

A B S T R A C T

This study presents the "ESP32 Dataset," a dataset of radio frequency (RF) data intended for human activity detection. This dataset comprises 10 activities carried out by 8 volunteers in three different indoor floor plan experiment setups. Line-of-sight (LOS) scenarios are represented by the first two experiment setups, and non-line-of-sight (NLOS) scenarios are simulated in the third experiment setup. For every activity, the volunteers performed 20 trials, hence there were 1,600 recorded trials overall per experiment setup in the sample (8 people $\times$ 10 activities $\times$ 20 trials) . In order to obtain the Received Signal Strength Indicator (RSSI) and Channel State Information (CSI) values from the recorded transmissions, the D-Link AX3000 router and ESP32 microcontroller were used as the transmitter (Tx) and receiver (Rx) in the data collection process. This collection is an invaluable resource for academics and practitioners in the field of human activity detection since it offers rich and diversified RF data across a wide range of experiment setups and activities. In contrast to other datasets with different hardware configurations, this dataset records one RSSI value and fifty-two CSI subcarriers using the ESP-CSI Tool RF data capture tool. The number of RSSI and CSI signals, specific to the ESP32 hardware, allows for the exploration of resource-efficient activity

* Corresponding author.
E-mail address: lyong@mmu.edu.my (L.-Y. Ong).

https://doi.org/10.1016/j.dib.2024.111101

detection algorithms, which is crucial for Internet of Things (IoT) applications where low-power and cost-effective solutions are required. This dataset is particularly valuable because it reflects the constraints and capabilities of the widely used ESP32 microcontrollers, making it highly relevant for developing and testing new algorithms tailored to IoT environments. The availability of this dataset enables the development and evaluation of activity detection algorithms and methodologies, enhancing the potential for improved experimental setups in IoT applications.

Specifications Table

Subject	*Computer Science Applications*
Specific subject area	*Human activity detection using radio frequency data collected via Wi-Fi technology.*
Type of data	*Raw dataset, table*
Data collection	The Wi-Fi CSI-based application, ESP-CSI Tool [1], was utilized to gather the dataset by capturing the transmitted and received Wi-Fi packets between a Tx and a Rx. The hardware setup included a D-Link AX3000 router as the Tx and an ESP32 microcontroller as the Rx. Both the D-Link AX3000 router and ESP32 microcontroller were configured to operate at Wi-Fi Standard IEEE 802.11n (Wi-Fi 4) . The ESP32 microcontroller was connected to the ESP-CSI Tool via a laptop, serving as the intermediary for data collection. Additionally, the router was connected to a power supply to ensure continuous operation during data collection sessions. The ESP32 microcontroller was further connected to the router to facilitate the exchange of Wi-Fi packets for data capture. Data were collected from 8 volunteers performing 10 activities each, across three indoor experiment setups (LOS and NLOS scenarios) . Each volunteer completed 20 trials per activity, resulting in 1,600 trials per experiment setup. The collected data are stored in .csv file format.
Data source location	Institution: Information Science Lab, Faculty of Information Science and Technology, Multimedia University City/Town/Region: 75,450 Bukit Beruang, Melaka Country: Malaysia Latitude and longitude (GPS coordinates) for collected samples/data: 2.2497454310576592, 102.27613505300872
Data accessibility	Repository name: Mendeley Data. Data identification number: 10.17632/x4x5xttvwt.1 Direct URL to data: https://data.mendeley.com/datasets/x4×5xttvwt/1
Related research article	*None*

1. Value of the Data

- The dataset captures both LOS and NLOS scenarios, offering insights into the impact of environmental conditions on radio frequency-based activity detection.
- The dataset enables benchmarking of existing algorithms, validation of new techniques, and comparison across studies in the field of activity recognition.
- The data was collected using consistent hardware and software setup, ensuring reliability and reproducibility for researchers.
- With the proliferation of Internet of Things (IoT) devices, the dataset provides valuable insights into leveraging RF data for enhancing IoT applications by enabling more accurate and context-aware activity detection algorithms.

2. Background

In recent years, radio frequency data-based human activity detection has become increasingly popular in sectors such as security, smart homes, and healthcare [2]. The widespread usage of commercial Wi-Fi equipment indoors makes radio frequency data an affordable and effective option for identifying and detecting human activities [3]. ESP32 microcontrollers are extensively used in IoT applications due to their low cost, versatility, and ease of use [4]. The strengths of ESP32 microcontroller include integrated Wi-Fi capabilities, low power consumption, significant processing power, cost-effectiveness, and flexibility in supporting multiple development environments and programming languages [5]. These features make the ESP32 particularly suited for human activity detection, as they can efficiently capture RSSI and CSI data, comparable to other hardware options. This dataset leverages the capabilities of ESP32 microcontrollers to provide a unique resource for researchers, capturing a variety of activity data in diverse indoor experiment setups. This contributes to the development of robust and generalizable solutions for human activity detection using radio frequency data, underscoring the importance of the ESP32 in this field.

3. Data Description

The collected raw data were organized into a main directory with three subdirectories, corresponding to the three indoor experiment setups. The information gathered from 8 different volunteers can be found in each of the subdirectories. There was a total of 1,600 recorded trials per setting (8 volunteers × 10 activities × 20 trials), as each volunteer performed 20 trials for each activity. Due to this, the subdirectory of each experiment setup has 1,600 files, each of which is a comma-separated values (*.csv)* file that represents a distinct trial. Table 1 provides the list of activities.

The format for each data file is "Ex_Sy_Az_Ti.*csv*," and Table 2 describes the specifications for data file naming. For example, data collected in the first experiment setup for the eighth volunteer doing the fifth activity (walking from the center towards the Tx) is recorded in the data file "E1_S8_A05_T20.*csv*," where the trial number is 20.

Table 1
List of activities of the ESP32 dataset.

Activity Indicator	Activity
A01	Jumping jack
A02	Squatting
A03	Hand swing (front and back)
A04	Walking from the centre towards the Rx
A05	Walking from the centre towards the Tx
A06	Bouncing basketball
A07	Jogging in place
A08	Forward bend 90°
A09	Sitting down on chair
A10	Standing

Table 2
Data files naming convention of the ESP32 dataset.

Symbol	Abbreviation for	Range
E	Experiment for Scenario	{1, 2, 3}
S	Subject (Volunteer)	{1, 2, 3, ..., 8}
A	Activity	{1, 2, 3, ..., 10}
T	Trial	{1, 2, 3, ..., 20}

In the course of an activity trial executed by a volunteer, a sequence of m packets is logged. Each packet is individually saved in a distinct row within the *.csv* file pertaining to the activity trial. The attributes of each entry in a row are comprehensively outlined in Table 3.

Table 3
Description of fields of each row entry within the *.csv* file.

Field	Description
Seq	The order of packets captured or transmitted.
Timestamp	The exact time when the packets was captured or received.
target_seq	The sequence number associated with the activity.
Target	The activity performed by the volunteer.
Mac	The MAC address of the Tx.
Rssi	RSSI represents the strength of the signal captured by the Rx.
Rate	The speed of communication link between devices, typically measured in Mbps (Megabits per second).
sig_mode	The modulation and encoding scheme used for transmitting data.
Mcs	The Modulation and Coding Scheme (MCS) specifies the combination of modulation and error correction coding used in wireless communication.
Cwb	The channel bandwidth or the range of frequencies allocated for communication.
smoothing	This field indicates whether smoothing is applied during transmission.
not_sounding	This field indicates whether a sounding frame is being used for channel estimation and feedback.
aggregation	This field indicates whether multiple data frames are combined into a single transmission unit.
stbc	This field indicates whether Space-Time Block Coding (STBC) is employed to improve the reliability of wireless communication by transmitting redundant data across multiple antennas.
fec_coding	The type of Forward Error Correction (FEC) coding used, if any.
sgi	This field indicates whether Short Guard Interval (SGI) is enabled to reduce the guard interval between symbols in wireless communication.
noise_floor	The level of background noise in the communication channel
ampdu_cnt	The number of A-MPDU (Aggregate MAC Protocol Data Unit) transmitted or received.
channel_primary	The specific frequency band used for transmission
channel_secondary	The secondary channel used to increase data transmission rate, if applicable.
local_timestamp	The timestamp relative to the local time of the device capturing the packet.
ant	The antenna used for transmitting or receiving the packet.
sig_len	The size of the signal or frame in bytes.
rx_state	the operational state of the Rx.
len	The length of the packet in bytes.
first_word	The information about the first word or header of the packet.
data	The actual payload or content of the packet. In this context, data refers to the CSI subcarriers captured.

4. Experimental Design, Materials and Methods

4.1. Environment setup

The data collection took place at the Information Science Lab, Multimedia University (Melaka Campus), where three distinct environmental setups were employed. The dimensions of the Information Science Lab are 8.3 m × 5.6 m.

This dataset provides a comprehensive representation of indoor experiment setups with two LOS scenarios and one NLOS scenario. With this configuration, researchers can capture the variability within LOS conditions and explore the possibility of proximity influencing the strength of radio frequency data. In the two LOS scenarios, the distance between the Tx and Rx is 7 m and 5.2 m for the first experiment setup and second experiment setup, respectively. The purpose of this intentional distance difference is to explore if radio frequency data is influenced by

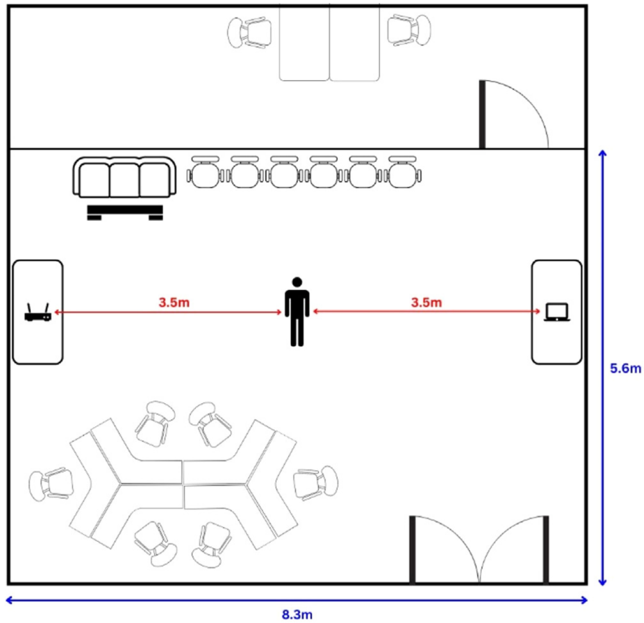

Fig. 1. The Floor Plan of the LOS Scenario 1.

proximity. Additionally, the distinct NLOS scenario provides insights into how diverse environmental barriers impact RF-based activity detection. Hence, this study observes how RF signals are blocked or reflected in NLOS scenario by having a wooden wall as a barrier. This setup enabled the analysis of physical obstructions on signal strength and reliability, providing insights into RF behavior in NLOS scenarios. Through the simulation of these diverse environmental conditions, the dataset helps researchers develop and evaluate robust activity detection algorithms that work in a variety of scenarios, improving the algorithms' dependability and practicality.

In the LOS Scenario 1, the distance between the Tx and Rx is 7m. Volunteers were instructed to perform the activities at the midpoint between the Tx and Rx. Fig. 1 illustrates the floor plan of the LOS Scenario 1 at the common area of Information Science Lab.

In the LOS Scenario 2, the Tx and Rx were placed 5.2 m apart from each other. Similarly, the volunteers were instructed to perform the activities at the midpoint between the Tx and Rx. Fig. 2 illustrates the floor plan of the LOS Scenario 2.

The experiment setup of NLOS Scenario is different from the setups in LOS scenarios. In the NLOS Scenario, there is a wooden wall barrier between the volunteer and the devices (Tx and Rx) . The Tx was placed at the common area of the Information Science Lab while the Rx was placed inside the inner room of Information Science Lab. Fig. 3 illustrates the floor plan of the NLOS scenario.

4.2. Software and equipment

In order to transmit and receive the Wi-Fi packets, a transmitter of D-Link AX3000 router and a receiver of ESP32 microcontroller were utilized. RF signals were transmitted via the antennas of the D-Link AX3000 router (Tx), and signals were received by the ESP32 microcontroller (Rx) over Wi-Fi. The RF signals were monitored through Wi-Fi packets to evaluate any effects of human activity on RF transmission. The transmitted packets were captured and processed using

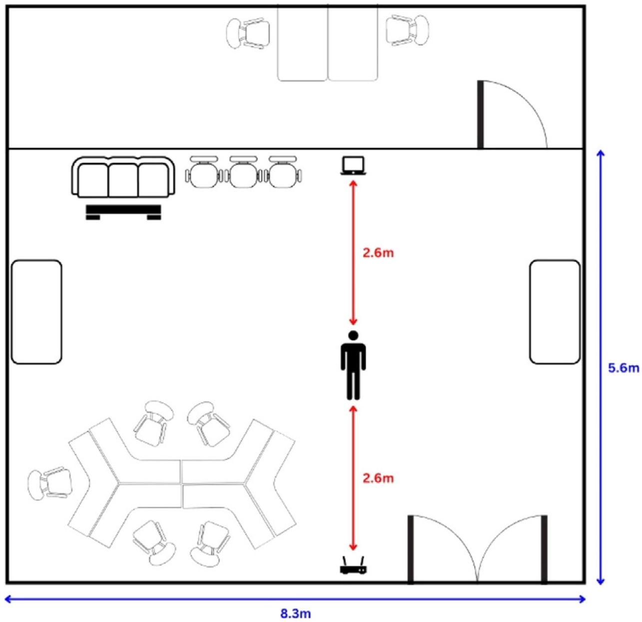

Fig. 2. The Floor Plan of the LOS Scenario 2.

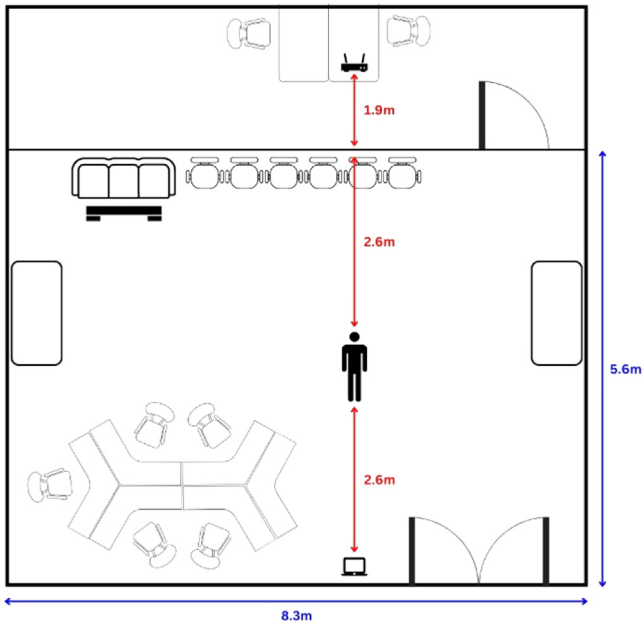

Fig. 3. The Floor Plan of the NLOS Scenario.

the ESP-CSI Tool [1], which is available as open-source software on GitHub. It is necessary to configure the ESP-CSI Tool to be compatible with the baud rate of the microcontroller.

In this study, the CPU frequency of the ESP32 microcontroller is 240 MHz and the baud rate is configured to 115,200. The ESP32 microcontroller operated at the eleventh channel with a channel bandwidth of 40 MHz throughout the data collection process. Both the D-Link AX3000 router and ESP32 microcontroller were configured to operate at Wi-Fi Standard IEEE 802.11n (Wi-Fi 4) . Since RF signals typically propagates in a straight line, referred to as LOS, the setup with a transmitter of D-Link AX3000 router and a receiver of ESP32 microcontroller is sufficient to analyze the impact of human interference on RF transmission without the need for additional devices or sensors. Fig. 4 shows the transmitter and receiver.

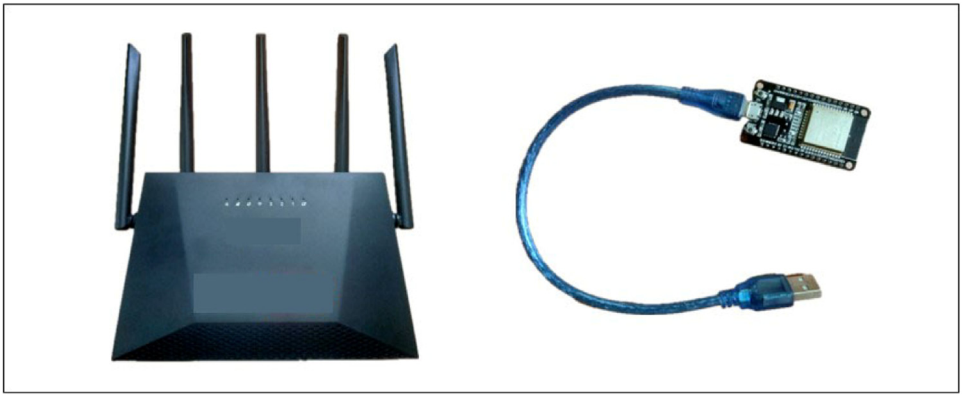

Fig. 4. Transmitter D-link AX3000 router and receiver ESP32 microcontroller.

4.3. Experimental procedure

A timing diagram was designed to ensure accurate performance of the data collection process. The timing diagram outlined the beginning and ending times of each activity. Volunteers were notified of these timings using a whistle to indicate the beginning and ending times of each activity. Fig. 5 shows the timing diagram, where a sound icon is used to represent whistle sound.

The volunteers for this study are chosen based on specific criteria to ensure the reliability and relevance of the data collected. The age range of the volunteers is between 22 and 25 years and all volunteers are in normal health status. The group of volunteers consists of three females and five males. This selection was made because the activities involved are generally suitable for adults and not intended for senior citizens or children.

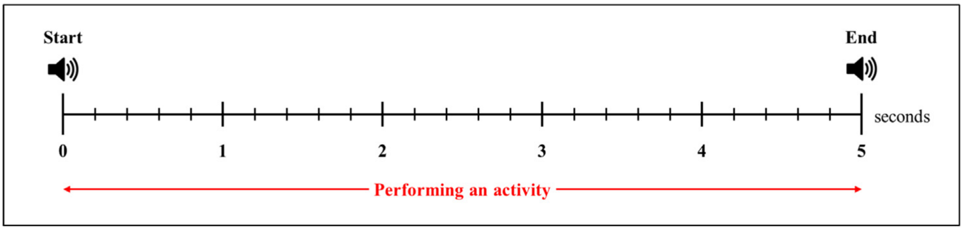

Fig. 5. Timing diagram of the data collection process.

Every volunteer was tasked to complete 10 distinct types of activities. Before commencing the data collection, several steps were explained to the volunteers to ensure that the data collection progressed smoothly. The participating volunteers were instructed to repeat each task for 20 trials in order to gather multiple instances of the same activity. The volunteers were specifically instructed to:

- Perform each activity at the midpoint between the Tx and the Rx.
- Start and stop performing the activity upon hearing the whistle sound.
- Engage seriously in performing the activities. Any form of idleness or negligence, such as laziness or slacking off, are prohibited during the data collection process.

The volunteers were instructed to perform each activity trial midway between the Tx and Rx to ensure that the signal strength is balanced and consistent. At this midpoint, the strength of the signal from the Tx is equal to the strength of the signal received, minimizing variables that could affect the data collected [6]. This consistency is crucial for obtaining reliable and comparable data across all trials.

Limitations

The dataset's reliance on only 8 volunteers may introduce limitations in terms of diversity and representation. With a small number of participants, there's a risk of overlooking individual variability in behavior, which could impact the dataset's ability to capture the full spectrum of human activities. Moreover, while each volunteer conducted 20 trials for each activity, totaling 1,600 recorded trials per experiment setup, this dataset can be used as a preliminary investigation of the human activities in these settings. A larger and more diverse pool of volunteers, coupled with an increased number of trials, would provide a more comprehensive dataset, potentially yielding more robust insights and findings. Additionally, the presence of furniture within the indoor experiment setups can introduce multipath propagation, a phenomenon where signals reflect off surfaces, creating multiple signal paths between Txs and Rxs. This can lead to signal distortion, interference, and variations in signal strength, affecting the accuracy of data collected in indoor settings. Furthermore, while the dataset provides detailed descriptions of the activities performed by volunteers, it does not offer visualization tools for activity signals. This might limit the usability and utility of the dataset for researchers in the field of human activity detection.

Ethics Statement

This study was approved by the Ethics Committees of Multimedia University on 03.07.2024 with the approval number: EA0222024.

Credit Author Statement

Zhe-Yu Lim: Writing-Original Draft, Lab Setup, Data Collection, Data Curation;
Lee-Yeng Ong: Supervision, Writing-Reviewing and Editing;
Meng-Chew Leow: Funding Acquisition;

Data Availability

Radio Frequency-based Human Activity Dataset (Original data) (Mendeley Data)

Acknowledgments

This work is funded by Telekom Malaysia Research and Development under Grant RDTC/221073 (MMUE/230002). The authors would like to thank the volunteers who have participated in the data collection process.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Z.Z. Cheng, F. Hui, Y.M. Tang, J.G. Wu, GitHub - espressif/esp-csi: applications based on Wi-Fi, CSI (Channel State information), 2020.
[2] Z. Chen, C. Cai, T. Zheng, J. Luo, J. Xiong, X. Wang, RF-based human activity recognition using signal adapted convolutional neural network, IEEE Trans. Mob. Comput. 22 (1) (2023) 487–499, doi:10.1109/TMC.2021.3073969.
[3] J. Yang, Y. Liu, Z. Liu, Y. Wu, T. Li, Y. Yang, A framework for human activity recognition based on WiFi CSI signal enhancement, Int. J. Antennas Propag. 2021 (2021), doi:10.1155/2021/6654752.
[4] M. Babiuch, P. Foltynek, P. Smutny, Using the ESP32 microcontroller for data processing, in: 2019 20th International Carpathian Control Conference (ICCC), IEEE, 2019, pp. 1–6, doi:10.1109/CarpathianCC.2019.8765944.
[5] A. Maier, A. Sharp, Y. Vagapov, Comparative analysis and practical implementation of the ESP32 microcontroller module for the internet of things, in: 2017 Internet Technologies and Applications (ITA), IEEE, 2017, pp. 143–148, doi:10.1109/ITECHA.2017.8101926.
[6] C. Brzozek, B.M. Zeleke, M.J. Abramson, K.K. Benke, G. Benke, Radiofrequency electromagnetic field exposure assessment: a pilot study on mobile phone signal strength and transmitted power levels, J. Expo Sci. Environ. Epidemiol. 31 (1) (2021) 62–69, doi:10.1038/s41370-019-0178-6.