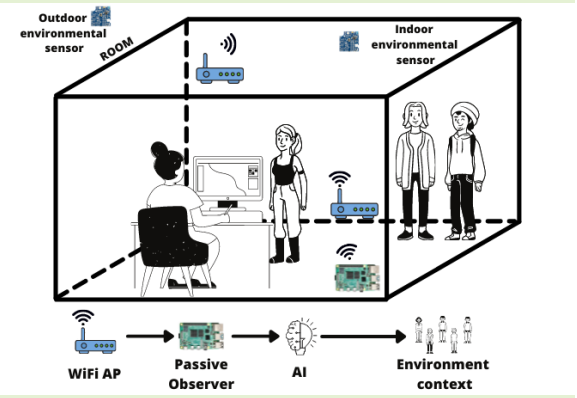
A Dataset on CSI-based Activity Recognition in Real-World Environments

Florenc Demrozi *Member, IEEE*, Cristian Turetta *Member, IEEE*, Alejandro Masrur *Member, IEEE*, Martin Schmidhammer *Member, IEEE*, Christian Gentner, Samarjit Chakraborty *Fellow, IEEE*, Graziano Pravadelli *Senior, IEEE*, and Philipp Kindt

***Abstract*— Activity recognition, e.g., identifying individuals, recognizing their physical activities, or estimating their number in a room, based on WiFi's Channel State Information (CSI) has been studied intensively in the last decade. While most existing works consider analyzing CSI data from a single person in a rather constrained environment, almost none of them has been successful in generalizing these results to unconstrained, real-world environments, in particular, when multiple individuals are present. In this paper, to address this problem, we introduce a fully annotated dataset ($\approx$ 70 GB of data) containing CSI and environmental data collected from two real-world offices over multiple days of continuous monitoring. To the best of our knowledge, this is the first freely available dataset of its kind. On the one hand, our dataset evidences that vastly disregarded *implicit changes* in the environment – due to small objects being repositioned, added or removed – are the main reason for the lack of generalizability by existing approaches. On the other hand, we expect it to promote further research work in this area and, thereby, to facilitate general solutions for CSI-based activity recognition in real-world environments.**



***Index Terms*— Channel State Information, Activity Recognition, Wireless Communication, Pattern Recognition**

I. Introduction

While originally designed for communication purposes, IEEE 802.11 networks (aka WiFi) have also become popular for sensing. In particular, when a WiFi signal travels from a sender to a receiver, it is altered by the environment through absorption, reflection, and scattering [1]–[5]. The receiver can detect these changes and thereby reconstruct activities or events of interest taking place in the environment [6]–[10]. For example, the following human-related contexts can be reconstructed: performed activities [11], [12], the number of persons being present in a room [13]–[16], spoken words [17], [18], people's identity [19]–[21], heart rate [22]–[24], respiratory rate [23], [25]–[28], body temperature [29], sleep quality [30]–[33], emotional status [34], [35], localization [8], [36]–[39], gait analysis [40], [41], smoking [42], eating [43], pose estimation [44]–[46], or gesture recognition [47]–[51].

WiFi-based sensing is interesting for activity recognition, healthcare, and environmental monitoring, in particular, when device-free systems are required, which do not require users to wear or carry any devices on their bodies [9], [52]–[55]. Moreover, since WiFi networks are verily ubiquitous, the only additional hardware required for sensing purposes is a receiver or passive observer as more commonly refer to it [56], [57]. This can be a commercial off-the-shelf (COTS) radio, resulting in a low-cost solution for the mentioned applications.

The passive observer extracts the so-called *Channel State Information* (CSI), i.e., it keeps track of variations in the amplitude and phase of WiFi signals, which are caused by changes in the environment. This data is typically processed by pattern recognition (i.e., machine and deep learning) algorithms that infer what changes in the environment might have led to the received CSI amplitude and phase variations [58]–[60]. To this end, a classification or regression model is first trained using labeled CSI data and then used to classify or estimate previously unseen data [7], [56], [57], [61].

However, a system trained on data obtained in a specific environment and involving a certain set of persons does not perform well in a different environment, even if the same persons are considered. Such *explicit* or evident changes in

"This work was supported in part by the U.S. Department of Commerce under Grant BS123456."

F.D is with the Department of Electrical Engineering and Computer Science, University of Stavanger, Norway, (name.surname@uis.no) (Corresponding Author)

C.T and G.P are with the Department of Computer Science, University of Verona, Italy, (name.surname@univr.it)

A.M and P.K are with Faculty of Computer Science, TU Chemnitz, Germany, (name.surname@informatik.tu-chemnitz.de)

M.S and C.G are with the German Aerospace Center (DLR), (name.surname@dlr.de)

S.C is with Department of Computer Science, University of North Carolina at Chapel Hill (UNC), USA, (name.surname@cs.unc.edu)

the environment are already known to be a major problem for CSI-based activity recognition [19], [52], [62]. In other words, CSI data is often collected in a typically constrained/controlled environment (e.g., a fixed room/office with a fixed distance between WiFi receiver and transmitter, etc.). However, classification models – either based on classic machine learning (CML) or deep learning (DL) – mainly generalize to the collected data [7], [58], [60], [61]. Thus, using the model in a different environment, in which no training data has been collected, unavoidably leads to unsatisfactory results.

Moreover, as discussed later in detail, this is aggravated by the fact that, even within the same environment, *implicit* or subtle changes may also severely affect classification results. For example, in almost any real-world environment, the positions of objects (e.g., a monitor, a chair, etc.) change at least marginally over time, or simply humidity and temperature gradually vary, thereby, possibly affecting the recognition performance [63]–[65]. As a result, existing solutions lack generality and do not easily extend to real-world environments.

Our Contributions: In this paper, we are concerned with the above limitations and make the following, basically two-fold contributions.

First and foremost, to investigate the reasons for this lack of generality, we collected and carefully labeled CSI data in two different real-world offices, involving multiple subjects performing unconstrained daily activities, for 4 consecutive days. To the best of our knowledge, this is the first dataset ($\approx$ 70 GB) of its kind, which we make freely available to the community for further analysis.

Second, we conducted a preliminary analysis of our own on the collected data using signal processing and machine learning methods. We unequivocally show that, besides explicit/evident changes, also implicit/subtle changes in the environment are indeed the cause of a declining accuracy in CSI-based activity recognition. In particular, this is revealed when comparing CSI data recorded during different nights. The data from each night exhibits significantly different CSI measurements, even though having been recorded in the same office without any human activity. The only difference is that some objects (e.g., chairs, monitors, etc.) were moved during the day and left at different positions.

Further, we observed that CSI data also varied within the same night, although there were no changes in the environment, neither explicit nor implicit. We observed a correlation with changing environmental conditions, such as temperature and humidity, and were even able to train different regression models capable of estimating temperature and humidity based on CSI data with a surprisingly low error as detailed later.

Our dataset vs. existing datasets: Table I presents an overview of the existing freely available datasets similar to the one accompanying this paper. In particular, such datasets are collected in study setups/environments where:

- participants only perform a very restricted range of activities (e.g., walking, specific gesture, jumping, etc.);
- participants are trained to perform the activities in a specific manner;
- participants perform activities within a specific area in the environment (most commonly, only between the receivers and the emitters);
- only one participant is present at a time in the environment;
- the used instrumentation is costly (i.e., SDR, mmWave, or an arrangement of multiple WiFi receivers/emitters).

As a result, such datasets can only be used for the originally intended purpose and do not extend well to other case studies nor do they include information about environmental conditions such as temperature or humidity, which we show to have an influence on results.

Instead, our dataset, along with the CSI data, contains detailed environmental information, such as temperature, humidity, CO_2, and air pressure. In addition, it is fully annotated with the activities of all persons (also simultaneously inside the environment) at every point in time, which – we hope – enables further research in this area. Having fully annotated CSI data from real-world settings, together with detailed environmental information, allows researchers to focus on developing, comparing and improving robust CSI-based activity recognition systems.

Paper Structure: The rest of the paper is organized as follows. Section II introduces some related concepts and preliminaries. Section III presents our data collection setup and its associated processing steps. Finally, Section IV introduces our dataset and discusses some results substantiating our above assertions, while Section V concludes the paper.

II. Concepts and Preliminaries

This section describes some necessary concepts, i.e., the already mentioned Channel State Information (CSI) and the Received Signal Strength Indicator (RSSI), and discusses preliminaries such as the used devices and software.

A. Channel State Information (CSI)

The signal sent on a wireless channel is altered in the environment by additive noise, interference, signal attenuation and multipath propagation. Multipath propagation is experienced when the transmitted signal arrives at the receiver via different propagation paths. These propagation paths involve reflections, diffractions and scattering of the wireless signal, leading to different delays. Hence, the received signal consists of a superposition of multiple altered instances of the transmitted signal, which are called *multipath components*.

Variations in the environment, e.g. moving persons, moved objects, etc., change the pattern of superposed signals, and can hence be detected by the receiver. In the frequency domain, the signal $X(f)$ is emitted by a WiFi device. Similarly, a receiver receives a signal $Y(f)$, for which the following holds:

$$Y(f) = X(f) \cdot H(f) + N(f), \tag{1}$$

where $H(f)$ is the channel transfer function, $N(f)$ represents the noise on the channel and f the frequency. WiFi

TABLE I
OVERVIEW OF EXISTING FREELY AVAILABLE DATASETS.

Ref. (Year)	# of Participants	# of Activities	Defined Activities	# of Environments	Constrained Yes/No	Dataset Size	# of Emitters	# of Receiver	Env. Size
[66] (2019)	10	16	horizontal arm waving, two hand waving, tossing paper, drawing tick, phone, drawing X clap hands, high arm waving, drinking water throwing objects upward, kicking forward kicking to the side, squatting, sitting down bowing, walking	3*	Yes	≈ 100 Mb	1	3	6x8 m 6x10 m 6x8 m
[67] (2019)	9	6	clapping, walking, waving, jumping, sitting, falling	1*	Yes	≈ 550 Mb	1	1	8.6x3.4 m
[68] (2021)	1	8	empty, lying, sitting, standing, sitting down, standing up, walking, falling	3*	Yes	≈ 830 Mb	1	2	4x4.5 m 3.5x4.5 m 4.5x5
[69] (2022)	3	5-9**	empty room, sitting, walking, running, jumping, standing, sitting down, standing up, jumping, moving arms	3*	Yes	≈ 20 GB	1	1-4**	5x6 m 5x6 m 7.5x3.7 m
Our [70]	6	5	entering the office, walking, standing, sitting, leaving the office, empty room	2	No	≈ 70 GB	2**	1	12x6 m 6x4 m

* Participants perform their activities in a specific area of the environment. ** Different tests in different configurations.

simultaneously transmits data on different frequencies, called *subcarriers*. When receiving a WiFi signal, the receiver can obtain an estimate on $H(f)$. It is given in the form of a vector of complex values $\mathbf{H}$.

Each element of $\mathbf{H}$ defines the CSI estimate of a certain subcarrier. Hence, $\mathbf{H}$ is an estimate how the WiFi signal is propagated from the transmitter to the receiver. Thus, $\mathbf{H}$ also contains information on, e.g., human activities in the signal propagation path [4], [7], [57], [61].

B. Received Signal Strength Indicator (RSSI)

The RSSI provides an indication of the power level at which data frames are received. The rationale is that a higher RSSI value implies a stronger signal power and, hence, a closer distance between sender and receiver. In this work, besides CSI, we use RSSI measurements to implement various filtering techniques as explained later in detail.

C. Devices used

To collect and store CSI data, we used an experimental setting consisting of the following devices:

- Two wireless routers (Fritz!Box 7530, $ 190[1]) were used to create two different WiFi networks at 2.4 GHz for transmitting data from a sender (i.e., a Raspberry Pi) to a receiver (i.e., a PC or laptop).
- Two Raspberry Pi 4 Model B ($ 35, 85 x 56 x 17 mm) containing a WiFi radio of the BCM43 device family from Broadcom. They were used to collect CSI data from the WiFi networks created by the two routers.
- Two Nordic Thingy 52 devices ($ 36, 50 x 55 x 15 mm), used to collect environmental data such as temperature (T), humidity (H), air pressure (P), CO_2 level, and light intensity (L) [71].
- A Laptop (Dell Inspiron 7559) handling the CSI data storage and used for labeling.
- An Android smartphone to record videos of the environment for context annotation.

[1]We indicate the estimative cost of the used equipment for information purposes. This, however, may vary depending on different factors, which are beyond relevance.

D. CSI extraction software

To extract CSI data from the WiFi radios, we use the Nexmon firmware patching framework [72], [73], which works on multiple Broadcom radios – as the one featured on the Raspberry Pi boards we use – and allows accessing the CSI data from the host computer. For data preprocessing, visualization and recording, we use the WiFiEye software framework [74].

III. DATA COLLECTION

In this section, we describe our data collection setup and the designed data processing pipeline we employed for this dataset analysis.

A. Real-world environments

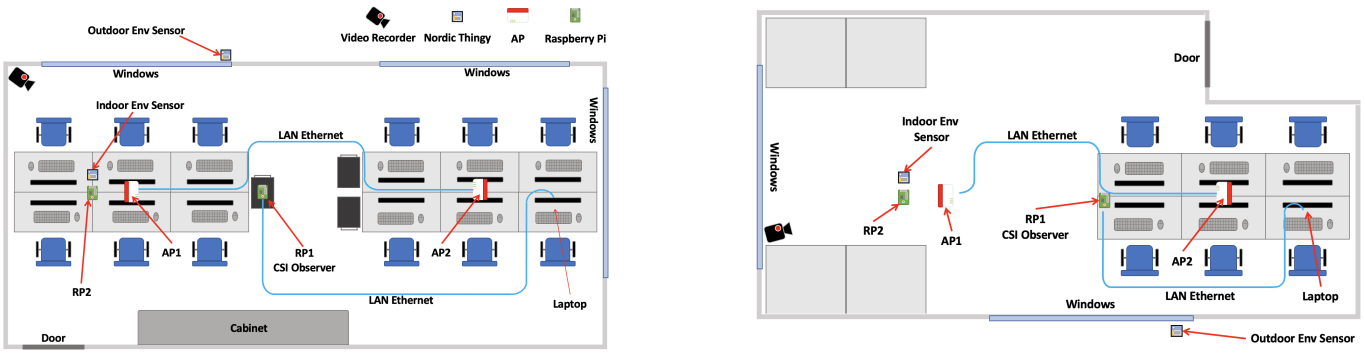
The most significant limitations of commonly used methods for CSI data collection based on COTS radios are as follows.[2] First, data is collected under well-defined constraints. For example, exactly one participant has to be in the environment at a specific timestamp. This participant is instructed to perform specific activities at specific positions, e.g., in the line of sight between sender and receiver. Second, the dimensions of the environment are reduced to, e.g., only a few meters between sender and receiver.

In contrast to this, we collected our data in two different real-world offices. The first office, shown in Figure 1 (a), contains twelve work spaces divided in two blocks, with one monitor and one desktop PC each.

The size of this office is $12 \times 6 \times 3$ meters, with one entrance door and three (2×1.8 meters) windows. Internal walls, i.e., those without windows) are made of plasterboard with a thickness of 12 centimeters, whereas external walls are of reinforced concrete with a thickness of 55 centimeters.

The second office, shown in Figure 1 (b), presents a size of $6 \times 4 \times 2.75$ meters, containing six work spaces, one entrance

[2]Note that using more complex hardware such as software-defined radios (SDRs) might ease these restrictions; however, it also increases costs considerably and hinders applicability. In this paper, we hence focus on COTS radios.



(a) Office 1 (b) Office 2

Fig. 1. Overview of the two real-world offices used for CSI data collection in this paper

door, and two windows of 2.5×3.5 meters. Internal and external walls have the same composition as the previously described office.

B. Device arrangement

As shown in Figure 1 (a) and (b), we used six different devices (i.e., two APs, two Raspberry Pi boards, and two Nordic Thingy devices) to collect CSI and environmental data (i.e., humidity, temperature, etc.). In addition, to annotate the recorded data, i.e., with labels corresponding to the activities of the subjects in the environment, we used an Android smartphone, which continuously recorded videos during the experiments. A Dell Inspiron 7559 laptop was used to store the collected CSI and environmental data.

The APs were placed at the center of the two blocks of work spaces, 5 meters apart from each other at the height of 140 cm. A LAN cable connects the APs into a mesh network. The first Nordic Thingy device is placed inside the room near RP2 (i.e., Raspberry Pi 2) and transmits its environmental context (i.e., temperature, humidity, CO_2 level, pressure, and light intensity) to the latter upon each status variation. The second Nordic Thingy is placed outside the window, but otherwise behaves the same as the first one.

The Raspberry RP1, patched with the Nexmon framework, collects and sends the CSI data to the Dell laptop. RP1 is approximately placed in the middle of the office.

People in our experiments could stand/walk in between RP1 and AP2, but not between RP1 and AP1. The same applies for RP2 (i.e., Raspberry Pi 2), however, this latter generates traffic towards AP1 instead. A similar setup was replicated for the second office shown in Figure 1 (b).

C. Communication between devices

As sketched in Figure 1 (a) and (b), we used the following communication technologies between devices:

- WiFi (i.e., 802.11ax) enables the communication between RP2[3] and AP1, AP2 and the Dell laptop.
- A LAN is used to connect AP1 to AP2, RP1 to the Dell laptop.
- Bluetooth Low Energy (BLE) communicates the indoor and outdoor Nordic Thingy devices with RP2.

WiFi's utilization, i.e., the amount of traffic sent over the network, is of central importance when extracting CSI information. In particular, if the network is not being used (i.e., there are no data packets transmitted), then the CSI extractor (i.e., RP1) will extract CSI data at a frequency of 10 Hz (i.e., the frequency with which an AP transmits advertising beacons), potentially reducing the discerning capabilities of the CSI-based activity recognition. In our setup, to prevent this from happening, RP2 generates traffic at around 120 Hz towards AP1. AP1 transmits such data towards AP2 through LAN and then from AP2 transmits to the Dell laptop again via WiFi. Finally, RP1 extracts the CSI data from the WiFi network constituted by the above device arrangement around AP1 and AP2.

D. Collected data

As mentioned above, we collected the following data based on the described setup consisting of two real-world offices:

- CSI data, collected by RP1 in the corresponding offices;
- environmental data from within and outside the offices, collected by the Nordic Thingy devices;
- annotation data, manually extracted from the recorded videos of the environment during the experiments.

The CSI and environment data is synchronized based on timestamps (at the Dell laptop and RP1). However, to synchronize this data with the video recording, one participant is instructed to shake the indoor Nordic Thingy device clearly in front of the video recording smartphone at the beginning of the data collection. This causees a glitch in the environmental data with can be associated with a timestamp and, hence, sincronized to the video recording. The collected (i.e., CSI and environmental) data and the generated annotations are stored in *.csv* files. In the following, this section discusses each single type of collected data.

CSI data: Once that CSI data is extracted by RP1 throughout the Nexmon patch, it is immediately transmitted to the Dell laptop creating a *.csv* file for each monitored AP. Table II presents the format with which this data is stored.

In particular, each captured CSI sample presents the reception timestamp of the corresponding frame, CSI

[3]Note that RP1 (i.e., which extracts CSI data) behaves as a sniffer without being connected to any WiFi network.

TABLE II
CSI DATA FORMAT

	CSI Amplitude			CSI Phase			
Timestamp	S_0	...	S_{63}	S_0	...	S_{63}	RSSI
15:38:45.550	0.027	...	1	1388.14	...	7563.08	-48
15:38:45.600	0.027	...	1	1388.14	...	7563.08	-46
15:38:45.650	0.027	...	1	1388.14	...	7461.06	-50
...	...	...	...	...	...	...	...
15:38:45.700	0.027	...	1	1388.14	...	7563.16	-55

amplitude and phase for each of the 64 subcarriers together with the corresponding RSSI value.

Environmental data: Table III presents an overview of the data collected by the *indoor* and *outdoor* Nordic Thingy devices. In particular, each of such devices collects the temperature (T), humidity (H), CO_2 level (C), air pressure (P), and light intensity or luminosity (L). A new sample is stored every time the environmental indoor or outdoor conditions vary.

TABLE III
ENVIRONMENTAL DATA FORMAT

	Indoor					Outdoor				
Timestamp	T $(°C)$	H $(\%)$	C (ppm)	P (hPa)	L (lum)	T $(°C)$	H $(\%)$	C (ppm)	P (hPa)	L (lum)
15:38:45.284	21.97	43	400	1005.29	920	7.23	91	340	1004.87	672
15:38:45.425	21.82	43	400	1005.29	920	7.23	91	340	1004.87	672
15:40:12.680	22.32	40	412	1005.29	920	7.23	91	340	1004.87	672
...	...	...	...	...	...	...	...	...	...	...
15:45:35.701	22.66	36	452	1005.29	1022	7.35	92	341	1004.90	624

Annotation data: Table IV presents an overview of the annotation data. In particular, observing the video recording, a new row is added to the annotations file at each context change in the environment. A context change identifies a change of the participant's activity, such as the transition from sitting to standing (and vice versa). Columns P_1 to P_n identify the participant's ID and contain the activity they start to perform at the particular timestamp. The observed activities are: 1) entering the office, 2) walking, 3) standing, 4) sitting, 5) leaving the office, and 6) out of the office. Column occupancy detection (OD) equals 0 if the office is empty and 1, if someone is in the environment. Finally, the column occupancy counting (OC) is an integer value representing the number occupants in the environment.

TABLE IV
ANNOTATION DATA FORMAT

Timestamp	P_1	P_2	P_3	..	P_n	OD	OC
15:38:45.000	Sitting	Entering	Sitting	...	Walking	1	4
15:38:56.500	Sitting	Sitting	Sitting	...	Leaving	1	4
15:38:58.000	Sitting	Standing	Sitting	...	Out	1	3
...	...	...	...	...	...	...	...
17:38:50.050	Out	Out	Out	...	Out	0	0

Combinations and variations of such annotations are also possible, but not further considered in this paper. For example, it is possible to identify time instants during which all participants are sitting or standing, or to include information about the identity of participants currently present in the environment, etc.

E. Data processing

To mitigate the issue related to the sampling frequency of the CSI data, to adapt the data to the input format required by CSI-based recognition models, and to evaluate such models, we performed the processing steps shown in Figure 2. The designed processing flow is composed of three blocks:

- Dataset generation: This block generates the dataset that accompanies this paper. Note that this processing step does not clean, modify, or create any synthetic data. All data is obtained by measurements.
- Preprocessing: This block removes noise and extract features from the CSI data.
- CSI-based recognition models evaluation: This block is concerned with dimensionality reduction, training and testing of different pattern recognition models as well as with evaluating their performance.

In the following, this section discusses each single block and sub-block shown in Figure 2.

Data under-sampling: Starting from the collected data, formatted as shown in Tables II and III, the first processing step concerns under-sampling the data from the existing dynamic sampling frequency to a constant sampling frequency. In particular, we under-sampled AP1 data to 100 Hz and AP2 data to 20 Hz to consider different sampling frequencies in the dataset and later be able to compare results based on different such frequencies.

This step is performed by segmenting the AP1/AP2 raw data for each subcarrier in time windows of one second and extracting 100/20 equidistant samples from the segments. This process homogenizes the sampling frequency, since this raw data present a dynamic sampling frequency greater than 100Hz and 20 Hz, respectively, for traffic generated by AP1 and AP2.

The same approach is performed on the environmental data, thus, resampling such data at 20 Hz. Greater sampling frequencies were not used, since the maximal sampling frequency observed on the environment data was 23 Hz. Again, note that no synthetic data was generated, but only measured data is used.

Data synchronization: As already mentioned, the CSI and the environmental data are synchronized based on timestamps at the Dell laptop and at RP1 correspondingly. Video data required for labeling needs to be manually synchronized instead. To this end, we need to identify the timestamp of the first video frame in the environmental data and, thereby, also in the CSI data. This is achieved again by instructing one participant to shake the indoor Nordic Thingy device in front of the camera before experiments start. This then produces a clearly distinguishable glitch in the environmental data and, hence, the corresponding video frame can be anchored to the a timestamp, which is then used as a starting point for the next processing step.

Video-based data annotation: The data annotation procedure uses Python Video Annotator [75] on the video recorded at ten frames per second. At each variation of the environment

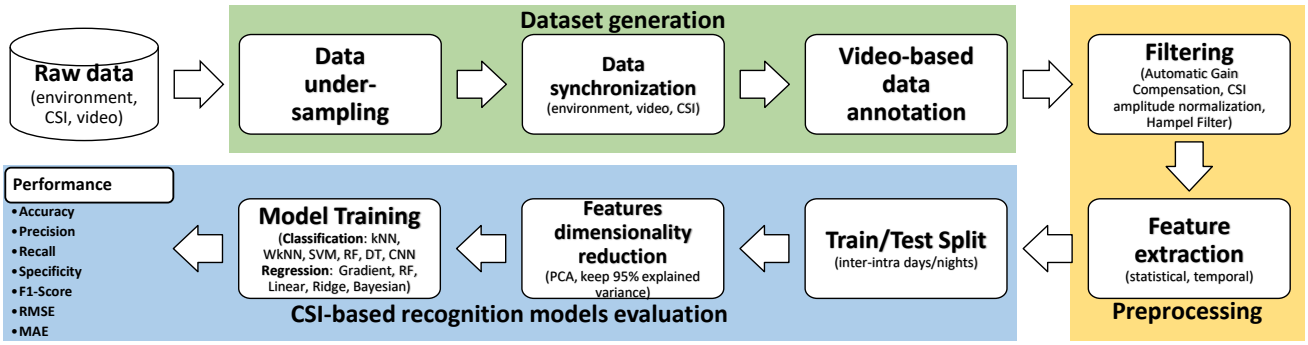


Fig. 2. CSI data processing workflow

status, the annotator inserts a new sample at the annotation file (as described in Table IV). Further annotations (e.g., position of the users in the room and their activities/status) were added manually in a following step and based on carefully observing the video recording.

Filtering: The propagation pattern of radio signals is affected by the environmental changes (e.g., occupants moving or objects moving), HW noise, and multi-path propagation. Moreover, the amplitude values are affected by the Automatic Gain Control (AGC) function that adjust the amplitude of the received signal, distorting CSI data. To prevent this from happening, we calibrate the measured CSI amplitudes by removing AGC, which can be done using RSSI measurements. This is because RSSI is obtained before AGC takes effect, while CSI is obtained thereafter [4], [76]. The calibration follows the equation:

$$CSI_i = CSI_i \times \sqrt{\frac{10^{\frac{RSSI}{10}}}{\sum CSI_i^2}}, \tag{2}$$

introduced in [5]. With regard to Table II, RSSI identifies the measured RSSI values at each timestamp, and CSI_i identifies the amplitude of the $i-th$ subcarrier received at the same timestamp.

Once that AGC is compensated, the two CSI datasets – obtained from AP1 and AP2 respectively – are merged by reducing the AP1's sampling frequency from 100 Hz to 20 Hz and removing the phase information. This results in a (merged) dataset of N rows and 129 columns (i.e., 64/64 CSI amplitude streams from AP1/AP2 and the timestamp stream). Subsequently, we removed all pilot subcarriers (i.e., 0, 1, 2, 3, 32, 61, and 63), which generally present a standard deviation equal to 0, i.e., which contain no usable CSI information. The remaining amplitude streams are normalized at each timestamp as described next:

$$\overline{CSI_i} = \frac{CSI_i - max(CSI)}{max(CSI)}, \tag{3}$$

where CSI_i identifies the amplitude of the $i-th$ subcarrier (i.e., the remaining subcarriers after excluding 0, 1, 2, 3, 32, 61, and 63) and $max(CSI)$ denotes the maximum amplitude value received among all subcarriers at that a given timestamp.

Finally, we apply a Hampel filter [52], [65], [77] using a 2-second time window, since this is the shortest time interval, in which a participant's context change can be reliably detected. This step thus returns a dataset of N rows and 115 columns (57/57 CSI amplitude streams from AP1/AP2 and the timestamp stream).

Feature extraction: Since raw data is not always interpretable [78], in this processing step, the CSI amplitude data of each subcarrier is segmented into 1-second time windows and subsequently used as input to the feature extraction module (i.e., Time Series Feature Extraction Library (TSFEL) library [79]). In particular, the feature extraction module extracts a total of 54 features from each 1-second time window/segment (e.g., interquartile range, kurtosis, min, max, mean, mean absolute deviation, median, median absolute deviation, etc., which correspond to features in the temporal and statistical domain). The complete list of features is available in [80].

We did not consider frequency domain features in order to reduce the dimension of the data to treat in the subsequent processing steps. Another reason for excluding features from the frequency domain is that they require a considerable time and computational effort to be extracted compared to features from the temporal and statistical domains.

This step returns a dataset of $N/20$ rows (i.e., N initial rows / 20 Hz) and 6156 columns (114 amplitude streams x 54 features). Moreover, the number of columns is further reduced by eliminating all features presenting a standard deviation of 0 (i.e., eliminating 1680 features, thus maintaining 4476 features). Note that, since we do not have any overlap between consecutive time windows/segments, there is no interdependence between the features extracted from them.

Train/test split: This process is usually performed using the following two techniques, a) k-fold and b) leave-out [81], [82]. The former divides the dataset into "k" folds and iteratively uses "k-1" folds to train and "one" fold to test the model. Each fold will randomly contain samples from all the classes, preserving the original dataset's class distribution.

The latter splits the dataset based on specific criteria. For example, iteratively use data from specific days, subjects, or hours to train the model and the data from other days, subjects, or hours to test it. For both techniques, the accuracy is represented by the average accuracy of all interactions.

Feature dimensionality reduction: Due to the big dimension of the training and testing datasets (more than 250 thousand samples or rows) and 4476 features or columns), it is of central importance to reduce their size. We use the well-known Principal Component Analysis (PCA) algorithm to reduce the size of the input dataset by keeping a sufficient number of principal components such that the variance of the initial dataset is preserved by at least 95 % [83].

Model training: We use five classification CML models (i.e., k-Nearest-Neighbor (k-NN), Weighted k-Nearest-Neighbor (Wk-NN), Support Vector Machine (SVM), Random Forest (RF), and Decision Tree (DT)), one DL model (i.e., Convolution Neural Networks (CNN)), and five regression models (i.e., Gradient boosting, RF, Linear regressor, Ridge regressor, and Bayesian regressor). To increase the performance of such models, we made use of the grid search hyper tuning in the training process [84], [85]. The "best" resulting model's performance was measured using the metrics of Precision, Recall, Specificity, F1-Score, and Accuracy for classification models, while Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) were used for regression models [86].

In conclusion, for each observed environment (i.e., each of the real-world offices), we collected and processed the following:

a) two CSI datasets (AP1 sampled at 100Hz and AP2 sampled at 20 Hz, formatted as shown in Table II),
b) the indoor and outdoor environmental data (sampled at 20 Hz and formatted as shown in Table III), and
c) the annotation data (sampled at 20 Hz and formatted as shown in Table IV).

As mentioned before, we make all this data freely available for further analysis, which can be downloaded from [70].

IV. Dataset and Analysis

This section presents an overview of the dataset collected in the first office (shown in Figure 1 (a)) using the setup and the processing steps described before. This dataset provides, for example, the possibility to perform, in increasing order of difficulty, the following recognition tasks: a) occupancy detection (i.e., whether the environment is empty or not), b) occupancy estimation (i.e., the number of persons/occupants in the environment), c) identity recognition (i.e., who is in the environment at a given time), and d) occupants' activity recognition (i.e., walking, sitting, running, etc.). Clearly, the complexity of recognizing such tasks increases when performed over data representing more occupants simultaneously. We will discuss the limitations of CSI-based recognition methodologies, concentrating on studying the environment's occupancy detection and on how implicit/subtle changes affect the models' accuracy.

A. Dataset characteristics

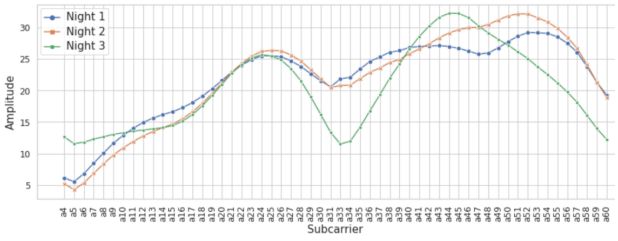


Fig. 3. The CSI patterns for the same empty environment are different in the different nights.

The data collection started on *January 04, 2022, at 15:08:40*[4], and concluded on *January 07, 2022, at 17:38:40*, for a total of *74 hours* (i.e., 268117 seconds) or 32174040×129 samples and a final size of around 70 GB.

The sampling frequencies of AP1 and AP2 were respectively in the range 112 Hz to 375 Hz and 20 Hz to 35 Hz. Such data represents the input *raw data* shown in Figure 2. At the end of the *dataset generation* step, four different *.csv* files are generated: 1) CSI data of AP1 sampled at 100 Hz, 2) CSI data of AP2 sampled at 20 Hz, 3) environment data sampled at 20 Hz, and 4) annotation data sampled at 20 Hz, with the formats shown in Tables II, III and IV. Again, all this data is made available under [70].

Six different persons (two females and four males) participated in our experiments and were informed of the data monitoring system and instructed to perform their daily routine without constraints. Table V provides an overview of the distribution of activities (i.e., one-second segments) per participant in our experiments.

Finally, Table VI shows the occupancy distribution in this environment/office. It becomes evident that our dataset presents an imbalance between the different occupancy states, increasing the difficulty of designing an accurate CSI-based activity recognition with this goal.

B. Dataset evaluation

Starting from the dataset obtained after the *dataset generation* process, we performed the filtering and feature extraction

[4]Timestamp of the first video frame, with which environmental and CSI data was synchronize as explained above.

TABLE V
ACTIVITY DISTRIBUTION PER PARTICIPANT

P_{ID}	Sitting	Standing	Walking	Entering	Leaving	Out
P_1	39111	1947	899	26	28	226106
P_2	8664	304	197	4	6	258942
P_3	79738	2280	1022	56	58	184963
P_4	5163	33	28	2	2	262889
P_5	35310	2788	618	16	16	229369
P_6	0	20	231	2	2	267862

TABLE VI
ONE-SECOND SEGMENTS OCCUPANCY DISTRIBUTION

Occupancy	Empty = 0	Occcupied = 1			
Occupants	Zero	One	Two	Three	Four
Total	169492	49309	28474	16622	4220
268117	169492 (62.94%)	98625 (37.06%)			

TABLE VII
CSI-BASED OCCUPANCY DETECTION (1)-(4) AND NIGHT ID RECOGNITION (5)

	Test (1) k-fold on all the dataset					Test (2) k-fold on all days					Test (3) leave-out inter days					Test (4) leave-out inter three-hours slots					Test (5) k-fold intra nights				
Model	Pre	Rec	Spe	F1	Acc	Pre	Rec	Spe	F1	Acc	Pre	Rec	Spe	F1	Acc	Pre	Rec	Spe	F1	Acc	Pre	Rec	Spe	F1	Acc
SVM	98.8	99.2	99.3	99.0	99.3	98.8	98.3	98.3	98.6	98.7	78.5	54.4	54.4	64.3	69.2	40.9	50.0	52.3	45.0	48.8	100	100	100	100	100
k-NN	99.0	97.0	99.4	98.0	98.5	97.9	98.0	98.0	98.0	98.2	76.4	62.3	62.3	68.6	73.6	54.6	49.6	56.0	52.0	49.8	100	100	100	100	100
Wk-NN	99.0	97.1	99.4	98.0	98.5	98.0	98.0	98.0	98.0	98.2	76.3	62.3	62.3	68.6	73.6	66.9	49.5	55.9	56.9	49.8	100	100	100	100	100
RF	98.1	97.7	98.9	97.9	98.4	96.8	95.0	95.0	95.9	96.4	66.7	53.5	53.5	59.4	68.8	50.6	63.6	38.7	56.4	54.0	100	100	100	100	100
DT	90.4	89.1	94.4	89.7	92.4	86.9	86.1	86.1	86.5	88.1	67.0	57.0	57.0	61.6	68.9	67.6	54.0	48.5	60.1	50.8	99.7	99.7	99.8	99.7	99.7
CNN	99.3	97.4	99.4	98.3	98.6	99.2	98.8	98.8	99.0	98.9	77.2	68.5	54.1	72.6	73.5	55.5	67.6	41.2	61.0	58.2	100	100	100	100	100

Precision (Per), Recall (Rec), Specificity (Spe), F1-Score (F1), Accuracy (Acc)

steps described in Section III. As a result, the returned dataset is represented by 268117 one-second segments, each represented by 4476 temporal and statistical features. Such a dataset is used to evaluate different CSI-based occupancy recognition (i.e., is the environment empty or is there at least one person in it?) models differentiating on the train/test splitting method (i.e., k-fold or leave-out). The dimensionality reduction and the training step are, however, the same for each method.

Next, we discuss the results obtained by training and testing the different models according to the workflow presented in Figure 2. We consider 9 different contexts/scenarios, for which results are shown in Tables VII, VIII, and IX.

Tests 1-4: The results of tests (1) and (2) show that CML and DL models, trained respectively over the whole dataset using the k-fold technique and the data obtained from the days of monitoring, achieve high F1-Score (99 %) on occupancy recognition. This result is in accordance with the existing works from the literature and suggests that models tested on data that has been collected at time instants relatively close to the ones used for training independently from their type, achieve high generalization capabilities. This is because no variations occur in the environment or, in other words, the data used for training already contains samples of all variations that are to be detected by the system.

However, emulating a real-world environment, test (3) (i.e., iteratively use 1 day to train and 2 days to test) and test (4) (i.e., iteratively use one 3-hour slot to train and the other slots to test) suggest that CML and DL models cannot easily generalize over changes in the environment from one day to another or even in the same day.

Such accuracy reduction is related to the mentioned explicit/evident and implicit/subtle changes in the environment. As discussed previously, to prevent the impact of explicit changes, the existing CSI-based activity recognition approaches exclusively work in specific areas of an environment, where users perform constrained activities, typically between the WiFi sender and receiver, and following well-defined motion patterns. This way, it can be guaranteed that samples of all CSI variations to be expected are present in the training data. However, as our results demonstrate, implicit/subtle changes – like small objects being moved, etc. – already suffice to introduce variations that confuse the system and cause a bad performance.

Tests 5-9: The results of tests (5)-(9) emphasize that implicit/subtle changes represent a consistent source of "noise" in real-world environments such as those described by this dataset. In particular, in tests (5)-(9), we concentrate on CSI data capture over night (i.e., 8:00 p.m. - 7:00 a.m.) where the environment is empty, so no explicit changes take place and the multipath propagation remains clearly stable during the same night.

As shown in test (5), we can precisely recognize the night ID (F1-Score 100 %), meaning that the CSI data perceived during the three different nights is very dissimilar from each other. This dissimilarity is mainly related to the implicit changes in the environment. Two possible types of implicit changes can induce such accuracy, a) objects in the environment left at a different position from the previous workday, and b) variations of the environmental conditions (i.e., humidity, temperature, etc.). The effect of implicit changes of type (a), as shown also in Figure 3, represents the main reason why the CSI-based models manage to distinguish between the different nights with such a high accuracy.

On the other hand, to the best of our knowledge, the effect of implicit changes of type (b) has not been significantly investigated. Test (6) is one first attempt towards this and presents the average recognition capabilities of the tested models by using data from every single night to recognize the specific hour (i.e., 11 possible classes, 8:00 p.m, 9:00 p.m, ..., 6:00 a.m). As it can be observed, the models achieve good results (F1-Score 86.8 %) in this 11-class problem. This means that within nights, where radio signal patterns are not affected by changes in the position of object, such relatively good accuracy can also be related to changes in environmental parameters such as temperature and humidity.

Test (7) presents the results obtained iteratively using two nights to train the model and one night to test it. These results show a complete inability (F1-Score 9.7 %) of the models to disambiguate among the different hours on nights not seen in the training test. This suggests that the impact of implicit changes of type (a) is much more consistent than that of type (b), which also matches our intuition in this regard.

TABLE VIII
CSI-BASED HOUR ID RECOGNITION

	Test (6): k-fold intra nights					Test (7): leave-out inter nights				
Model	Pre	Rec	Spe	F1	Acc	Pre	Rec	Spe	F1	Acc
SVM	86.9	86.8	98.7	86.8	86.8	7.7	9.1	90.9	8.4	9.1
k-NN	79.2	78.6	97.9	78.9	78.6	10.7	8.9	90.9	9.7	8.9
Wk-NN	79.3	78.8	97.9	79.0	78.8	9.9	9.3	90.9	9.6	9.3
RF	80.7	80.5	98.1	80.6	80.5	8.9	10.6	91.1	9.7	10.6
DT	53.0	52.7	95.3	52.9	52.8	9.3	9.7	91.0	9.5	9.7
CNN	86.4	86.5	98.1	86.4	86.5	12.3	10.3	91.0	7.6	10.3

Finally, in tests (8) and (9), we present the capabilities of the compared models to estimate temperature and humidity.

The temperature ranged from 18.38 to 31.60 $^{\circ}C$ during the three nights, and humidity ranged from 20 % to 48 %. In test (8), using each night separately, we achieved an average RMSE of 0.4 $^{\circ}C$ and 1.75 % in estimating temperature and humidity, respectively. Instead, in test (9), using iteratively two nights to train and one night to test the models, we achieved an average RMSE of 1.08 $^{\circ}C$ and 5.2 %. Test (8) shows that our CSI-based regression algorithm presents a low-error correlation with temperature and humidity in the absence of explicit and implicit changes of type (a). In test (9), the error triplicates, where instead the environment is affected foremost by implicit changes of type (a). This is consistent with the results of tests (5), (6), and (7), which we discussed before. Note, however, that the results do not allow any conclusions about whether the low-error correlation (w.r.t., test (8)) is related to implicit changes of type (b) such as temperature, humidity, or other undetermined changes such as could be the device overheating, in particular, at the Raspberry Pi extracting CSI data.

TABLE IX
CSI-BASED TEMPERATURE AND HUMIDITY ESTIMATION

	Test (8): k-fold intra nights				Test (9): leave-out inter nights			
	Temperature		Humidity		Temperature		Humidity	
Model	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Gradient	0.63	0.40	1.83	1.43	1.14	1.01	5.20	5.95
RF	0.60	0.44	1.81	1.40	1.08	0.85	5.86	5.52
Linear	0.40	0.30	1.75	1.35	1.17	0.95	6.07	5.77

Ridge, Bayesian and Linear regression algorithms present equal results.

V. CONCLUSIONS AND TAKEAWAYS

CSI-based activity recognition systems are becoming popular motivated by their low-cost and ubiquitousness, among others. In this paper, however, we noted difficulties when extending such systems to real-world settings, i.e., considering multiple participants at the same time, unrestricted activities, larger distances between WiFi transmitter and receiver, etc.

We hence investigated possible causes by collecting a large amount of CSI data over several days, including environmental data (such as temperature, humidity, etc.). After a first analysis of the collected data, we remark the following challenges of CSI-based activity recognition:

i) typical classification models (based on machine learning) lack generalization, when trained and tested on data from different time windows (i.e., train data collected on one day, test data on another day),
ii) implicit or subtle changes in a real-world environment (e.g., objects such as chairs being moved, new objects such as coffee mugs being introduced, etc.) considerably impact how WiFi signals propagate and, hence, alter CSI data,
iii) the impact of temperature and humidity could not be totally excluded as a possible cause of variation in the distribution pattern of WiFi signals/CSI data, and
iv) the existence of other possible implicit changes (i.e., overheating at WiFi transmitter or receiver, etc.) should be further investigated.

To the best of our knowledge, the presented is first, fully annotated dataset ($\approx$ 70 GB) of its kind, which we now make freely available to the entire community. It contains CSI and environmental data, annotated with the activities performed by six different participants during four days of unrestricted office work, i.e., participants performed their activities without any constraints/instructions as it is often the case with other datasets from the literature.

We hope that the provided dataset motivates further research work in this area and helps identify potential solutions to the above challenges. Moreover, our dataset provides the capabilities to study the following CSI-based scenarios in increasing order of difficulty in unconstrained environments (contemplating a single person as well as multiple persons in the environment): i) occupancy detection, ii) occupancy estimation (i.e., counting persons), iii) identity recognition (i.e., who is a particular environment), and iv) human activity recognition (i.e., what activities are being performed by individuals). Finally, our dataset can be used as a starting point for CSI-based temperature and humidity estimation in indoor environments, which again requires further attention and work.

REFERENCES

[1] F. Xiao, J. Chen, X. Xie, L. Gui, L. Sun, and R. Wang, "Seare: A system for exercise activity recognition and quality evaluation based on green sensing," *IEEE Transactions on Emerging Topics in Computing*, vol. 8, no. 3, pp. 752–761, 2018.

[2] Z. Zhou, F. Wang, J. Yu, J. Ren, Z. Wang, and W. Gong, "Target-oriented semi-supervised domain adaptation for WiFi-based HAR," in *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*. IEEE, 2022, pp. 420–429.

[3] Z. Wang, B. Guo, Z. Yu, and X. Zhou, "Wi-fi CSI-based behavior recognition: From signals and actions to activities," *IEEE Communications Magazine*, vol. 56, no. 5, pp. 109–115, 2018.

[4] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11 n traces with channel state information," *ACM SIGCOMM computer communication review*, vol. 41, no. 1, pp. 53–53, 2011.

[5] Z. Gao, Y. Gao, S. Wang, D. Li, and Y. Xu, "Crisloc: Reconstructable CSI fingerprinting for indoor smartphone localization," *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3422–3437, 2020.

[6] W. Taylor, Q. H. Abbasi, K. Dashtipour, S. Ansari, S. A. Shah, A. Khalid, and M. A. Imran, "A review of the state of the art in non-contact sensing for covid-19," *Sensors*, vol. 20, no. 19, p. 5665, 2020.

[7] 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. 155 986–156 024, 2019.

[8] L. Chen, J. Xiong, X. Chen, S. I. Lee, K. Chen, D. Han, D. Fang, Z. Tang, and Z. Wang, "Widesee: Towards wide-area contactless wireless sensing," in *Proceedings of the 17th Conference on Embedded Networked Sensor Systems*, 2019, pp. 258–270.

[9] M. Muaaz, A. Chelli, A. A. Abdelgawwad, A. C. Mallofré, and M. Pätzold, "WiWeHAR: Multimodal human activity recognition using wi-fi and wearable sensing modalities," *IEEE Access*, vol. 8, pp. 164 453–164 470, 2020.

[10] C. Yang, X. Wang, and S. Mao, "Demo abstract: Technology-agnostic approach to rf based human activity recognition," in *IEEE INFOCOM 2022-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE, 2022, pp. 1–2.

[11] 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*, pp. 1–54, 2022.

[12] F. Zhang, K. Niu, J. Xiong, B. Jin, T. Gu, Y. Jiang, and D. Zhang, "Towards a diffraction-based sensing approach on human activity recognition," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 1, pp. 1–25, 2019.

[13] W. Xi, J. Zhao, X.-Y. Li, K. Zhao, S. Tang, X. Liu, and Z. Jiang, "Electronic frog eye: Counting crowd using WiFi," in *IEEE INFOCOM 2014-IEEE Conference on Computer Communications*. IEEE, 2014, pp. 361–369.

[14] F. Demrozi, C. Turetta, F. Chiarani, P. H. Kindt, and G. Pravadelli, "Estimating indoor occupancy through low-cost ble devices," *IEEE Sensors Journal*, vol. 21, no. 15, pp. 17 053–17 063, 2021.

[15] H. Zou, Y. Zhou, J. Yang, and C. J. Spanos, "Device-free occupancy detection and crowd counting in smart buildings with WiFi-enabled iot," *Energy and Buildings*, vol. 174, pp. 309–322, 2018.

[16] R. Zhang, X. Jing, S. Wu, C. Jiang, J. Mu, and F. R. Yu, "Device-free wireless sensing for human detection: the deep learning perspective," *IEEE Internet of Things Journal*, vol. 8, no. 4, pp. 2517–2539, 2020.

[17] G. Wang, Y. Zou, Z. Zhou, K. Wu, and L. M. Ni, "We can hear you with wi-fi!" *IEEE Transactions on Mobile Computing*, vol. 15, no. 11, pp. 2907–2920, 2016.

[18] C. Du, X. Yuan, W. Lou, and Y. T. Hou, "Context-free fine-grained motion sensing using WiFi," in *2018 15th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. IEEE, 2018, pp. 1–9.

[19] 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)*. IEEE, 2016, pp. 1–12.

[20] C. Shi, J. Liu, H. Liu, and Y. Chen, "Smart user authentication through actuation of daily activities leveraging WiFi-enabled iot," in *Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, 2017, pp. 1–10.

[21] P. Connor and A. Ross, "Biometric recognition by gait: A survey of modalities and features," *Computer vision and image understanding*, vol. 167, pp. 1–27, 2018.

[22] J. Liu, Y. Chen, Y. Wang, X. Chen, J. Cheng, and J. Yang, "Monitoring vital signs and postures during sleep using WiFi signals," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 2071–2084, 2018.

[23] Y. Gu, X. Zhang, Z. Liu, and F. Ren, "WiFi-based real-time breathing and heart rate monitoring during sleep," in *2019 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2019, pp. 1–6.

[24] X. Wang, C. Yang, and S. Mao, "Phasebeat: Exploiting CSI phase data for vital sign monitoring with commodity WiFi devices," in *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2017, pp. 1230–1239.

[25] Y. Zeng, D. Wu, J. Xiong, E. Yi, R. Gao, and D. Zhang, "Farsense: Pushing the range limit of WiFi-based respiration sensing with CSI ratio of two antennas," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 3, pp. 1–26, 2019.

[26] Y. Zeng, D. Wu, J. Xiong, J. Liu, Z. Liu, and D. Zhang, "Multisense: Enabling multi-person respiration sensing with commodity WiFi," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 3, pp. 1–29, 2020.

[27] X. Wang, C. Yang, and S. Mao, "Resilient respiration rate monitoring with realtime bimodal CSI data," *IEEE Sensors Journal*, vol. 20, no. 17, pp. 10 187–10 198, 2020.

[28] C. Chen, Y. Han, Y. Chen, H.-Q. Lai, F. Zhang, B. Wang, and K. R. Liu, "Tr-breath: Time-reversal breathing rate estimation and detection," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 3, pp. 489–501, 2017.

[29] V. Ha and A. Nahapetian, "Received WiFi signal strength monitoring for contactless body temperature classification," in *EAI International Conference on Body Area Networks*. Springer, 2021, pp. 112–125.

[30] S. Yue, Y. Yang, H. Wang, H. Rahul, and D. Katabi, "Bodycompass: Monitoring sleep posture with wireless signals," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 2, pp. 1–25, 2020.

[31] X. Liu, J. Cao, S. Tang, and J. Wen, "Wi-sleep: Contactless sleep monitoring via WiFi signals," in *2014 IEEE Real-Time Systems Symposium*. IEEE, 2014, pp. 346–355.

[32] 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, 2015.

[33] M. Zhao, S. Yue, D. Katabi, T. S. Jaakkola, and M. T. Bianchi, "Learning sleep stages from radio signals: A conditional adversarial architecture," in *International Conference on Machine Learning*. PMLR, 2017, pp. 4100–4109.

[34] M. Zhao, F. Adib, and D. Katabi, "Emotion recognition using wireless signals," in *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking (MobiCom)*, 2016, pp. 95–108.

[35] Y. Gu, T. Liu, J. Li, F. Ren, Z. Liu, X. Wang, and P. Li, "Emosense: Data-driven emotion sensing via off-the-shelf WiFi devices," in *2018 IEEE International Conference on Communications (ICC)*. IEEE, 2018, pp. 1–6.

[36] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "Spotfi: Decimeter level localization using WiFi," in *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication*, 2015, pp. 269–282.

[37] S. Kumar, S. Gil, D. Katabi, and D. Rus, "Accurate indoor localization with zero start-up cost," in *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking (MobiCom)*, 2014, pp. 483–494.

[38] K. Ohara, T. Maekawa, and Y. Matsushita, "Detecting state changes of indoor everyday objects using wi-fi channel state information," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 3, sep 2017. [Online]. Available: https://doi.org/10.1145/3131898

[39] S. Sen, J. Lee, K.-H. Kim, and P. Congdon, "Avoiding multipath to revive inbuilding WiFi localization," in *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*, 2013, pp. 249–262.

[40] W. Wang, A. X. Liu, and M. Shahzad, "Gait recognition using WiFi signals," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2016, pp. 363–373.

[41] L. Zhang, C. Wang, M. Ma, and D. Zhang, "Widigr: Direction-independent gait recognition system using commercial wi-fi devices," *IEEE Internet of Things Journal*, vol. 7, no. 2, pp. 1178–1191, 2019.

[42] X. Zheng, J. Wang, L. Shangguan, Z. Zhou, and Y. Liu, "Smokey: Ubiquitous smoking detection with commercial WiFi infrastructures," in *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*. IEEE, 2016, pp. 1–9.

[43] 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 *2020 29th International Conference on Computer Communications and Networks (ICCCN)*. IEEE, 2020, pp. 1–9.

[44] M. Zhao, T. Li, M. Abu Alsheikh, Y. Tian, H. Zhao, A. Torralba, and D. Katabi, "Through-wall human pose estimation using radio signals," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7356–7365.

[45] M. Zhao, Y. Tian, H. Zhao, M. A. Alsheikh, T. Li, R. Hristov, Z. Kabelac, D. Katabi, and A. Torralba, "Rf-based 3d skeletons," in *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*, 2018, pp. 267–281.

[46] W. Jiang, H. Xue, C. Miao, S. Wang, S. Lin, C. Tian, S. Murali, H. Hu, Z. Sun, and L. Su, "Towards 3d human pose construction using WiFi," in *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking (MobiCom)*, 2020, pp. 1–14.

[47] A. Virmani and M. Shahzad, "Position and orientation agnostic gesture recognition using WiFi," in *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, 2017, pp. 252–264.

[48] 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*, 2019, pp. 313–325.

[49] H. Abdelnasser, M. Youssef, and K. A. Harras, "Wigest: A ubiquitous WiFi-based gesture recognition system," in *2015 IEEE conference on computer communications (INFOCOM)*. IEEE, 2015, pp. 1472–1480.

[50] C. Li, M. Liu, and Z. Cao, "Wihf: Enable user identified gesture recognition with WiFi," in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*. IEEE, 2020, pp. 586–595.

[51] R. H. Venkatnarayan, G. Page, and M. Shahzad, "Multi-user gesture recognition using WiFi," in *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, 2018, pp. 401–413.

[52] W. Jiang, C. Miao, F. Ma, S. Yao, Y. Wang, Y. Yuan, H. Xue, C. Song, X. Ma, D. Koutsonikolas *et al.*, "Towards environment independent device free human activity recognition," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking (MobiCom)*, 2018, pp. 289–304.

[53] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of WiFi signal based human activity recognition," in *Proceedings of the 21st annual International Conference on Mobile Computing and Networking (MobiCom)*, 2015, pp. 65–76.

[54] ——, "Understanding and modeling of WiFi signal based human activity recognition," in *International Conference on Mobile Computing and Networking (Mobicom)*, 2015, pp. 65–76.

[55] ——, "Device-free human activity recognition using commercial WiFi devices," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 5, pp. 1118–1131, 2017.

[56] C. Li, Z. Cao, and Y. Liu, "Deep ai enabled ubiquitous wireless sensing: A survey," *ACM Computing Surveys (CSUR)*, vol. 54, no. 2, pp. 1–35, 2021.

[57] J. M. Fernandes, J. S. Silva, A. Rodrigues, and F. Boavida, "A survey of approaches to unobtrusive sensing of humans," *ACM Computing Surveys (CSUR)*, vol. 55, no. 2, pp. 1–28, 2022.

[58] D. Wang, J. Yang, W. Cui, L. Xie, and S. Sun, "Multimodal CSI-based human activity recognition using gans," *IEEE Internet of Things Journal*, vol. 8, no. 24, pp. 17 345–17 355, 2021.

[59] Z. Chen, L. Zhang, C. Jiang, Z. Cao, and W. Cui, "WiFi CSI based passive human activity recognition using attention based blstm," *IEEE Transactions on Mobile Computing*, vol. 18, no. 11, pp. 2714–2724, 2018.

[60] P. Fard Moshiri, R. Shahbazian, M. Nabati, and S. A. Ghorashi, "A CSI-based human activity recognition using deep learning," *Sensors*, vol. 21, no. 21, p. 7225, 2021.

[61] Y. Ma, G. Zhou, and S. Wang, "WiFi sensing with channel state information: A survey," *ACM Computing Surveys (CSUR)*, vol. 52, no. 3, pp. 1–36, 2019.

[62] J. Zhang, Z. Tang, M. Li, D. Fang, P. Nurmi, and Z. Wang, "Crosssense: Towards cross-site and large-scale WiFi sensing," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking (MobiCom)*, 2018, pp. 305–320.

[63] N. C. Y. Lim, L. Yong, H. T. Su, A. Y. H. Chai, C. K. Vithanawasam, Y. L. Then, and F. S. Tay, "Review of temperature and humidity impacts on rf signals," in *2020 13th International UNIMAS Engineering Conference (EnCon)*. IEEE, 2020, pp. 1–8.

[64] J. Li, A. Sharma, D. Mishra, and A. Seneviratne, "Thermal profiling by WiFi sensing in iot networks," in *2021 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2021, pp. 1–6.

[65] ——, "Fire detection using commodity WiFi devices," in *2021 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2021, pp. 1–6.

[66] L. Guo, L. Wang, C. Lin, J. Liu, B. Lu, J. Fang, Z. Liu, Z. Shan, J. Yang, and S. Guo, "Wiar: A public dataset for WiFi-based activity recognition," *IEEE Access*, vol. 7, pp. 154 935–154 945, 2019.

[67] J. K. Brinke and N. Meratnia, "Dataset: Channel state information for different activities, participants and days," in *Proceedings of the 2nd Workshop on Data Acquisition To Analysis*, 2019, pp. 61–64.

[68] J. Schäfer, "CSI human activity," 2021. [Online]. Available: https://dx.doi.org/https://doi.org/10.3390/app11198860

[69] F. Meneghello, D. Garlisi, N. Dal Fabbro, I. Tinnirello, and M. Rossi, "SHARP: Environment and person independent activity recognition with commodity ieee 802.11 access points," *IEEE Transactions on Mobile Computing*, 2022.

[70] "CSI Dataset," https://univr-my.sharepoint.com/:u:/g/personal/florenc_demrozi_univr_it/Ed4EfxQXZfVLrG1JJZaaJ3EBXuvle_I6UoybkEpQsTq0Fw?e=GjHC2u, 2023, [Online; accessed 16-March-2023].

[71] N. Semiconductor. (2022) Nordic thingy:52. [Online]. Available: https://www.nordicsemi.com/Products/Development-hardware/Nordic-Thingy-52

[72] M. Schulz, D. Wegemer, and M. Hollick. (2017) Nexmon: The c-based firmware patching framework. [Online]. Available: https://nexmon.org

[73] F. Gringoli, M. Schulz, J. Link, and M. Hollick, "Free your CSI: A channel state information extraction platform for modern wi-fi chipsets," in *Proceedings of the 13th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization*, ser. WiNTECH '19, 2019, p. 21–28. [Online]. Available: https://doi.org/10.1145/3349623.3355477

[74] P. H. Kindt, C. Turetta, F. Demrozi, A. Masrur, G. Pravadelli, and S. Chakraborty, "WiFiEye - seeing over WiFi made accessible," *CoRR*, vol. abs/2204.01830, 2022. [Online]. Available: https://arxiv.org/abs/2204.01830

[75] O. Source. (2022) Python video annotator. [Online]. Available: https://pythonvideoannotator.readthedocs.io/en/master/

[76] IEEE, "Ieee standard for information technology-telecommunications and information exchange between systems–local and metropolitan area networks-specific requirements–part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications amendment 2: Sub 1 ghz license exempt operation," 2017.

[77] K. Qian, C. Wu, Z. Yang, Y. Liu, F. He, and T. Xing, "Enabling contactless detection of moving humans with dynamic speeds using CSI," *ACM Transactions on Embedded Computing Systems (TECS)*, vol. 17, no. 2, pp. 1–18, 2018.

[78] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. Havinga, "A survey of online activity recognition using mobile phones," *Sensors*, vol. 15, no. 1, pp. 2059–2085, 2015.

[79] M. Barandas, D. Folgado, L. Fernandes, S. Santos, M. Abreu, P. Bota, H. Liu, T. Schultz, and H. Gamboa, "Tsfel: Time series feature extraction library," *SoftwareX*, vol. 11, p. 100456, 2020.

[80] O. Source. (2022) Tsfel webpage. [Online]. Available: https://tsfel.readthedocs.io/en/latest/index.html

[81] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.

[82] J. Brownlee, *Master Machine Learning Algorithms: discover how they work and implement them from scratch*. Machine Learning Mastery, 2016.

[83] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.

[84] J. Brownlee, "How to grid search hyperparameters for deep learning models in python with keras," *línea]. Disponible en: https://machinelearningmastery. com/grid-search-hyperparameters-deep-learning-models-python-keras*, 2016.

[85] M. Claesen, J. Simm, D. Popovic, and B. Moor, "Hyperparameter tuning in python using optunity," in *Proceedings of the international workshop on technical computing for machine learning and mathematical engineering*, vol. 1, no. 3, 2014.

[86] D. M. Powers, "Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation," –, 2011.