

Research article

Time matters: Empirical insights into the limits and challenges of temporal generalization in CSI-based Wi-Fi sensing

Andrea Brunello ^a,¹ Angelo Montanari ^b, Raúl Montoliu ^c, Adriano Moreira ^d, Nicola Saccomanno ^b,^{*},¹ Emilio Sansano-Sansano ^c, Joaquín Torres-Sospedra ^{e,f}

^a Dept. of Humanities and Cultural Heritage, University of Udine, Palazzo Caiselli, Vicolo Florio 2/b, Udine, 33100, Italy

^b Dept. of Computer Science, Mathematics, and Physics, University of Udine, Via delle Scienze 206, Udine, 33100, Italy

^c Institute of New Imaging Technologies, University Jaume I, Av. de Vicent Sos Baynat, Castelló de la Plana, 12071, Spain

^d Centro Algoritmi, University of Minho, Campus de Azurém, Guimarães, 4804 - 533, Portugal

^e Dept. of Computer Science, University of València, Avda Universitat s/n, Burjassot, 46100, Spain

^f VALGRAI, Valencian Graduate School and Research Network of Artificial Intelligence, Camí de Vera s/n, València, 46022, Spain

ARTICLE INFO

Dataset link: <https://doi.org/10.5281/zenodo.14212401>

Keywords:

Wi-Fi
Channel State Information
Wireless sensing
Indoor positioning
Human activity recognition
Machine learning

ABSTRACT

Wi-Fi is ubiquitous, and Channel State Information (CSI)-based sensing has often emerged as superior for tasks like human activity recognition (HAR) and indoor positioning (IP). The foundational premise is that similar scenarios exhibit similar CSI patterns. However, establishing such similarities is challenging due to signal attenuation and multipath effects caused by static and dynamic objects, that create complex interaction phenomena. Although acknowledged in literature, a comprehensive study of how these variables affect CSI patterns across scenarios, particularly their long-term impact on real-world applications, is still missing. In fact, many recent works focus on laboratory settings disregarding temporal generalization when testing their solutions. Here, we present a systematic study of the reliability of CSI-based sensing, consolidating key challenges and insights previously scattered in the literature. We provide a clear and independent perspective about the need of considering temporal aspects when developing CSI-based sensing approaches, particularly for real-world applications. To achieve that, we consider two tasks, IP and HAR, combining theoretical modeling with experiments using state-of-the-art methods. We show how tasks dependent on reflections from static objects, like IP, are severely impacted by disturbances that accumulate over time, also in the absence of physical modifications of the environment. In contrast, those relying on reflections from dynamic objects, like HAR, face fewer challenges. Our findings, supported by novel real-world datasets for CSI fingerprint-based IP and CSI stability analysis over time, suggest that future research must consider time as a crucial factor both in the development and test of approaches.

1. Introduction

Wi-Fi access points (APs) are now a fundamental part of the pervasive network infrastructure that connects people and objects to the Internet [1]. Beyond providing Internet connectivity, Wi-Fi has emerged as a powerful tool supporting applications such as indoor

* Corresponding author.

E-mail addresses: andrea.brunello@uniud.it (A. Brunello), angelo.montanari@uniud.it (A. Montanari), montoliu@uji.es (R. Montoliu), adriano@dsi.uminho.pt (A. Moreira), nicola.saccomanno@uniud.it (N. Saccomanno), esansano@uji.es (E. Sansano-Sansano), joaquin.torres@uv.es (J. Torres-Sospedra).

¹ Co-primary and corresponding author of the work.

positioning (IP) and human activity recognition (HAR) [2,3]. While traditional solutions have often relied on the AP Received Signal Strength Indicator (RSSI) measurements [3], the advent of Channel State Information (CSI)-based approaches offered a paradigm shift, promising enhanced accuracy and a broader spectrum of applications [4–7]. CSI popularity is further justified by its support to two operational modes: device-based, requiring user-carried devices for signal acquisition or emission, and device-free, not requiring any user-carried devices [4].

CSI provides communication channel properties as signals travel from the transmitter to the receiver, capturing the complexity of the environment. The foundational principle of CSI-based sensing is the intuition that distinct indoor scenarios, whether they pertain to positioning or the detection of intricate human activities, produce characteristic CSI patterns; in addition, similar scenarios yield similar patterns. Nevertheless, as well known by the literature, determining such similarities presents challenges as wireless signals, when propagating through indoor spaces, are affected by the presence of objects, both static and transient/moving (dynamic), that determine multipath and signal attenuation [8,9]. These elements, in conjunction with fluctuating environmental conditions, can induce complex interaction phenomena. Patterns can be highly influenced, for instance, by: the usage of different recording protocols (e.g., different devices, or a change into their position and orientation); different furniture arrangements; variations in the electromagnetic field, for instance due to electronic equipment; and, different occupancy of the premises, considering both people within them and in nearby areas [10,11].

Although some recent studies propose approaches to deal with these issues [11–17], a closer examination reveals a significant gap: no systematic, independent study has been conducted, so far, on the impact of different phenomena on CSI patterns, particularly regarding their evolution over time, and assessing real-world applications compared to the more controlled settings typically considered in laboratory experiments. In fact, much of the existing literature addresses these challenges in a fragmented manner, often focusing on specific aspects of the problem in isolation or with varying degrees of detail. For example, [11] conducts only a limited experimental evaluation regarding changes over time, concentrating on a device-free positioning task. Similarly, [15] raises the issue of generalization over time in indoor positioning, but does not investigate it in depth. In addition, several recent works still overlook the temporal dimension entirely, raising concerns about the validity of their portrayed – often very good – results [18–24]; these are typically achieved by disregarding the temporal variability when splitting the collected data into training and test sets which, as we will see, can lead to overly optimistic results.

Our work seeks to consolidate all these fragmented insights by providing a systematic, unified and unbiased analysis of the key challenges in CSI-based Wi-Fi sensing, particularly as they relate to real-world applications, where time plays a central role. The goal is to provide the research community with a clear understanding of the key factors that must be considered when developing—and, even more importantly, evaluating—novel CSI-based solutions. This is especially relevant given the large number of recent studies that appear to overlook these issues in their experimental workflow (see, e.g., [25–31]). We do that by focusing on the following research questions:

RQ1 *Do CSI-based Wi-Fi sensing tasks generalize over time?*

RQ2 *How do patterns associated with CSI components evolve through time?*

RQ3 *What challenges does the real world present, and how should future research be directed to address them effectively compared to current state of the art?*

Answering these questions systematically would, in principle, require an examination of all possible CSI-based tasks, which is practically impossible. For this reason, we focus on two representative ones, that we argue to rely on two different types of CSI components: the first is device-free human activity recognition, which leverages *dynamic* CSI components generated by moving objects, such as a person walking; the second is device-based indoor positioning, which relies on *static* components, like those generated by static objects, such as walls. As we will show, these two tasks provide us with a comprehensive view on the issue of CSI temporal generalization.

Our contributions can be summarized as follows:

- (i) we overview related work (Section 2) and spotlight the absence of in-depth, systematic analyses on the long-term temporal shifts in CSI patterns, especially in real-world scenarios;
- (ii) through a detailed modeling of CSI (Section 3), we argue that static components are more vulnerable to temporal changes than dynamic ones;
- (iii) we validate the CSI modeling starting from HAR (Section 5.1). By utilizing methods and data from recent approaches [32,33], we show that, when dynamic CSI components are relied upon, only a partial performance degradation is observed over time;
- (iv) we then focus on indoor positioning (Section 5.2). Utilizing a recently published dataset [34] and state-of-the-art approaches [12,13,16], we confirm that static CSI components, even in controlled scenarios, are highly susceptible to disturbances, leading to poor positioning performance. We corroborate these findings with negative results from a second dataset, collected by us, which models a real-world scenario (Section 4.3);
- (v) using another dataset, also collected by us (Section 4.4), we then analyze CSI signal stability over time considering a prolonged and continuous sampling, within a same premise, under both supposedly varying (weekdays) and non-varying (holidays) environmental conditions (Section 5.3). Our empirical findings suggest that even under seemingly static conditions (holidays), exogenous factors (e.g., weather) significantly influence CSI, accumulating over time.
- (vi) we conclude by thoroughly answering our research questions (Section 6) and providing a forward-thinking analysis, emphasizing the importance of time as a core element in both the development and the assessment of new approaches, and the necessity of comprehensive evaluation in large and diverse real-world scenarios.

2. Related work

Although there are several studies and surveys on CSI-based wireless sensing (e.g., [4,7,35,36]), to the best of the authors' knowledge, only a few contributions study the application of CSI to different tasks, focusing on the current advancements and limitations of the state-of-the-art, concerning the generalization capabilities of the various solutions. For instance, the work by [37] is a very broad, recent survey about CSI and the approaches developed to enable cross-domain generalization. There, also HAR and indoor positioning are considered, and temporal evolution can be seen as a special case of domain generalization, where the new domain is the same scenario but at a different time instant. Yet, the focus of the survey is not on this specific aspect, and the few works mentioned in it that consider this type of domain generalization for indoor positioning do it only partially. Specifically, in [17], a machine learning-based roaming model, CrossSense, is presented; its aim is to reduce the cost of training data collection on new environments, by generating synthetic instances. In [38], the authors propose a self-calibration time-reversal fingerprinting localization approach to mitigate the effects of environmental changes without modifying the fingerprint database, which may become outdated over time. Another recent contribution [39] tries to cope with changes in the environment that may occur at different time instants, training a domain-adaptive classifier and making use of data augmentation via a variational autoencoder. Still, they consider a device-free setting, very dense radio maps, and a high and regular number of packets being transmitted ad-hoc and received by an Intel 5300 NIC device [40]. Actually, the only work in which the authors systematically investigate real-world network conditions and the use of regular ambient traffic for CSI-based sensing tasks seems to be the very recent one of [41]. However, their study does not address spatial or temporal generalization.

In [11], LTLoc is presented as a device-free passive fingerprinting neural network-based localization approach. Recognizing that Wi-Fi signals (and, especially, CSI) are susceptible to various environmental factors that lead to pattern change over time, the authors also design AdaptDNN; it is an adaptive deep neural network relying on metanetwork learning to determine which layers and features of the original neural network need to be transferred to automatically adapt to CSI fingerprints change. The experiments involve an Intel 5300 network card receiving 100 packets per second, over a 6 day long localization task, carried out in a conference room with 56 reference points. Despite the interestingness of the work, the need for continuous domain adaptation over such a short time scale makes it challenging to apply the approach in real-world scenarios, where typically lower packet rates and an inferior number of reference points is to be expected. Similar considerations, compounded by the challenges of collecting radio-map/training data, hold true also for other works based on the idea of transfer learning or domain adaptation, even when based on unsupervised learning [5,36,42].

Then, the work by Rocamora et al. [43] tests CSI on four locations 20 cm apart, using dedicated hardware (2 USRP N210 SDR devices and a laptop) with MIMO and frequency hopping (30 channels), resulting in a 300 MHz effective bandwidth. The influence of environment and temporal dynamics (training and test on two consecutive days) are studied, suggesting that Time Reversal Resonating Strength (TRRS) [12] and a Support Vector Regression TRRS-inspired approach provide a certain degree of robustness to such phenomena. The matter is further studied in [44]. Although promising, the considered setting is far distant from the real-world scenario presented in this work, and in which TRRS does not seem to preserve its generalization properties (Section 5). Another work similar to ours is that of Cominelli et al. [32], which focuses on a single approach for human activity recognition, SHARP [33]. The performance of SHARP is tested on a novel dataset encompassing multiple scenarios, people, activities, access points, frequency bands, and temporal instants (two consecutive days). However, the authors only briefly discussed the influence of time.

Extending the analysis by Cominelli et al. [32] to better understand the influence of time concerning SHARP using their data will be the very first step of our contribution. In addition, our work differs consistently from the previous ones as: we consider both HAR and indoor positioning; we specifically focus on the influence time has on such tasks; we take into account multiple approaches from the state-of-the-art; and, we use and release a novel dataset for CSI based indoor positioning, resembling a real-world scenario. Overall, we aim to provide a foundational and unified perspective on why CSI-based tasks may fail over time, focusing on real-world conditions and applicability, which present several complexities. We advocate that studying this specific problem, often overlooked in current research, is of primary importance.

3. CSI-based sensing

In this section, we first provide background information on CSI sensing. Then, we discuss the different roles played by static and dynamic components in the CSI. While these concepts are well-established in the literature, our aim is to present them in a clear and comprehensive manner to the reader, in order to provide a solid theoretical foundation for our experiments.

3.1. Background and CSI modeling

Within a typical indoor environment, the transmission of a radio signal (e.g., Wi-Fi) is characterized by multiple propagation paths, arising due to reflective interactions with surfaces. Each such path causes attenuation and phase modifications to the originally transmitted signal. Thus, they inherently contain geometric characteristics of the propagation environment and, consequently, characterizing them is pivotal for the conceptualization of sensing algorithms.

CSI is precisely the metric used in Orthogonal Frequency-Division Multiplexing (OFDM) wireless standards (e.g., IEEE 802.11a/n/ac/ax) for describing amplitude and phase variations across subcarrier frequencies, as wireless signals travel

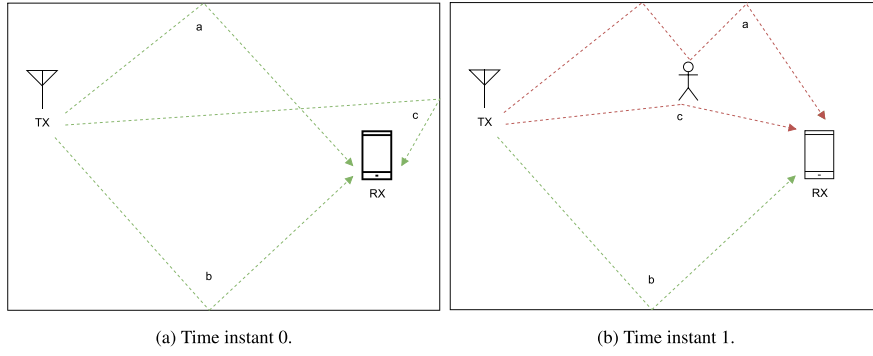


Fig. 1. Simplified example of static (green) and dynamic (red) transmission paths at two sequential times.

from a transmitter to a receiver. At a given time instant, the relationship between the transmitted and the received signal can be formalized as:

$$y = \mathbb{H}x + \eta, \quad (1)$$

where y is a vector representing the signal sensed by the receiver, x is the vector of the originally transmitted signal, and η is an additive vector of Gaussian white noise. $\mathbb{H} \in \mathbb{C}^{N_{\text{Tx}} \times N_{\text{Rx}} \times N}$ is a complex CSI matrix where N_{Tx} is the number of transmitting antennas, N_{Rx} is the number of receiving antennas, and N is the number of subcarriers. Restricting to a single link, such a matrix would contain a value for each subcarrier i representing the Channel Frequency Response (CFR) $\mathbf{h}_i$. The latter can be formalized as:

$$\mathbf{h}_i = A_i e^{j\phi_i}, \quad (2)$$

where A_i stands for the amplitude and ϕ_i the phase for the i th subcarrier.

The above considered modeling of CSI, the CFR, is based on the frequency domain representation of the signals. By means of the Inverse Fast Fourier Transform (IFFT), it is possible to equivalently translate it to the time domain representation. In that case, the Channel Impulse Response (CIR) is considered. In the same manner, through FFT, it is possible to move from the time domain to the frequency domain representation, thus from CIR to CFR.

As previously mentioned, due to the physical complexity of real-world environments, the received signal is typically seen as a sum of different reflections, also known as multipath:

$$\mathbf{h}_i = \sum_{n=1}^N A_n e^{\frac{-2\pi f_i d_n}{c} + j\phi_n}, \quad (3)$$

where A_n , ϕ_n and d_n are, respectively, the amplitude, phase, and path length related to a single multipath route n ; f_i is the frequency of the subcarrier i and c is the speed of light. The set of N multipaths can be partitioned into two subsets based on the objects that reflected the signal [35]. The first set, Ω_s (aka static components), includes all signals influenced only by reflections from static objects, such as walls; the second set, Ω_d (aka dynamic components), includes all dynamic paths, which are affected by a moving object such as a person walking. We thus obtain the following formulation:

$$\mathbf{h}_i = \sum_{n \in \Omega_s} A_n e^{\frac{-2\pi f_i d_n}{c} + j\phi_n} + \sum_{n \in \Omega_d} A_n e^{\frac{-2\pi f_i d_n}{c} + j\phi_n}. \quad (4)$$

In the literature, the elements of Ω_s are often considered to be invariant over time (see, for instance, [35]). Nevertheless, let us focus on the situations depicted in Fig. 1. In Fig. 1(a), the signal transmitted by TX is received, among others, along three paths by RX, namely, a , b and c . These are all static, as they are influenced only by reflections originating from static objects. In Fig. 1(b), instead, a moving person is introduced. As a result, among others, three paths are received by RX: two of them (red colored, a and c) are influenced also by the dynamic object, while the third (b) is, again, static. As a matter of fact, the two sets Ω_s and Ω_d change over time. We make the dependency over time explicit in Eq. (5), drawing inspiration from [45] to highlight such a phenomenon. Note how the attenuation, phase, and path length also depend on the considered time instant t :

$$\mathbf{h}_i(t) = \sum_{n \in \Omega_s(t)} A_n(t) e^{\frac{-2\pi f_i d_n(t)}{c} + j\phi_n(t)} + \sum_{n \in \Omega_d(t)} A_n(t) e^{\frac{-2\pi f_i d_n(t)}{c} + j\phi_n(t)}. \quad (5)$$

3.2. Static and dynamic components in CSI sensing over time

From Eq. (5), it follows that at two consecutive time instants, say t and $t + 1$, a dynamic event (e.g., a person moving) causes alterations in the propagation environment observable through the set of components²

$$\begin{aligned} & \{\Omega_d(t) \cap \Omega_d(t + 1)\} \\ & \cup \{\Omega_s(t) \cap \Omega_d(t + 1)\} \\ & \cup \{\Omega_d(t) \cap \Omega_s(t + 1)\} , \end{aligned} \quad (6)$$

i.e., those that are dynamic at both time instants, and those that were static (resp., dynamic) at time instant t and dynamic (resp., static) at time instant $t + 1$. In the scenario of Fig. 1, the terms are: $\emptyset \cup \{a, c\} \cup \emptyset$. The changes regarding the dynamic components can be naturally derived by contrasting $\mathbf{h}_i(t)$ with $\mathbf{h}_i(t + 1)$, essentially neglecting $\Omega_s^{t,t+1} = \Omega_s(t) \cap \Omega_s(t + 1)$, i.e., the static multipaths that remained as such. Thus, dynamic events can be characterized by a time-relative, differential quantity.

Conversely, the situation is more complex when focusing on static components. Let us consider two time instants, t and t' , with $t' > t$. Also, let us define

$$\Delta_{\Omega_s}(t, t') = \left| \sum_{n \in \Omega_s^{t,t'}} A_n(t) e^{\frac{-2\pi f_i d_n(t)}{c} + j\phi_n(t)} - \sum_{n \in \Omega_s^{t,t'}} A_n(t') e^{\frac{-2\pi f_i d_n(t')}{c} + j\phi_n(t')} \right|. \quad (7)$$

When $t \approx t'$, we have that $\Delta_{\Omega_s}(t, t') \approx 0$. Instead, when $t' \gg t$, then $\Delta_{\Omega_s}(t, t') > 0$, since reflections from the common static elements, even if unaltered by dynamic events, are still affected by other environmental factors. These include variations in the electromagnetic field, for instance linked to weather conditions or the presence of electronic equipment [46,47]. Thus the same static component, when observed in the same environment at distant time points, can be significantly different. Lastly, unlike the differential approach we outlined for dynamic components, there is no direct strategy for extracting $\Omega_s^{t,t+1}$ from sequentially received signals.

In summary, characterizing static components is challenging compared to dynamic ones, as CSI excels in highlighting variations in the propagation environment over consecutive time points (Eq. (6)). Thus, we anticipate that tasks utilizing static CSI components will be more susceptible to disturbances over time compared to those relying on dynamic CSI components.

4. Considered datasets

Here, we present the datasets we consider. First, the one of [32], used for the experiments on human activity recognition. Then, the datasets for indoor positioning, i.e., the one from [34], and our novel, real-world one, for the same task. Finally, our other proposed dataset, designed to study CSI signal temporal stability within an environment.

4.1. Literature HAR dataset

The authors in [32] conducted their research using a dataset collected over two days, as detailed in Table 1a. The dataset includes 7 distinct scenarios, labeled S1 through S7. In each of them, data is collected considering one transmitter and three receiver APs with multiple antennas (MIMO), and encompassing several transmission standards. We focus on 160-MHz 802.11ax which, in their study, provided the best performance. Scenarios S1–S5 were gathered sequentially in the same medium-sized laboratory room, with S5 being gathered on the following day than the others. Note again that these scenarios refer to the same physical environment and neither the receivers nor the transmitter were moved for the entire duration of the experiments. Scenario S6 represents data from a small office environment, while S7 from a large hall. In each scenario, one of three persons (labeled A, B, C) is instructed to carry out 12 activities, summarized in Table 1b. For each activity, 80 s of CSI data are collected.

4.2. Literature IP dataset

The dataset from [34,44] was collected between 17 Nov 2020 and 8 Apr 2021 in a university studio of $\approx 20 \text{ m}^2$ with five work desks. There are 8 reference points (RPs), clustered around two locations, as shown in Fig. 2. The network setup (20 MHz, 802.11n) uses a TP-Link Wi-Fi access point, to which two devices are connected: an Intel Wireless Link 5300 network interface card, equipped with three antennas, mounted on a HummingBoard Pro SBC (receiver device); and a laptop ping the board 2 times per second to generate network traffic. MIMO-OFDM is available, leading to log two streams from the AP on each receiver's antenna. The CSI data was collected over several days, with each day having multiple collection *periods* (up to 8). In each period, the receiver was moved through all 8 RPs, following a protocol for its positioning and to ensure a fixed orientation. At each RP, 180 packets were collected over a span of 5 min. Various environmental conditions, such as the presence of obstructions (e.g., a door that could be opened or closed, boxes placed in different positions, etc.) and people, were considered. The collected data was then partitioned into two splits, A and B, summarized in Table 2. Notably, apart from days and periods, the individual CSI observations lack precise timestamps, preventing pinpointing their exact, absolute measuring time instant.

² In fact, scattering phenomena can generate multiple new components from a single one, thus multisets should be considered. However, for simplicity and clarity, we resort to sets without loss of generality.

Table 1
Experimental scenarios and activities as in [32].

(a) Experimental scenarios				(b) Activities			
Scenario	Person	Env.	Time				
S1	A	Lab	Day 1	A.	Walk	G.	Wave hands
S2	B	Lab	Day 1	B.	Run	H.	Clapping
S3	C	Lab	Day 1	C.	Jump	I.	Lay down
S4	A	Lab	Day 1	D.	Sitting	J.	Wiping
S5	A	Lab	Day 2	E.	Empty room	K.	Squat
S6	A	Office	Day 2	F.	Standing	L.	Stretching
S7	A	Hall	Day 2				

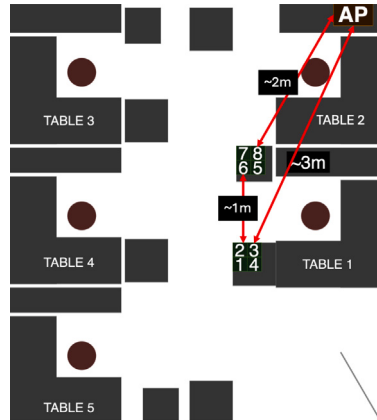


Fig. 2. Collection setup of [34]. Image adapted from [34], distributed under MIT license.

Table 2
Dataset Split A and Split B as considered in [34].

Subsets	Split A	Split B
Training	17 days, 77 periods (Nov 17–Dec 30)	18 days, 80 periods (Nov 17–Jan 5)
Validation	2 days, 8 periods (Jan 5–12)	5 days, 27 periods (Jan 12–26)
Test	3 days, 15 periods (Jan 15–20)	4 days, 18 periods (Mar 15–Apr 8)

4.3. Our novel IP dataset

One of our contributions is a new dataset for CSI device-based indoor positioning modeling a real-world scenario where fingerprinting is typically meant to be employed: premises may be complex and span multiple rooms and floors; users may not be connected to the Wi-Fi network, but simply sniff the already present, potentially irregular network traffic; the devices may be at random orientations; transmitters are placed with the aim of supporting connectivity, not indoor positioning, and may not be under the users' influence; in addition, their position may be unknown. This means typically there is no control over the AP's transmission channel, used frequency bandwidth, or observable CSI packets per second. Rather, user devices must adapt to the environment's radio-frequency characteristics.

Data was collected using an ESP32S2 portable development board [48] based on a standalone, low-cost, single antenna micro-controller, at a private home over two separate days, roughly one month apart from each other (8th Feb 2023 and 1st Mar 2023). To perform the collection we used the ESP32 CSI toolkit and its smartphone labeling app [35,49]. The board was running in passive mode, i.e., without being explicitly connected to any Wi-Fi access point, as opposed to what is often assumed in the literature (see, e.g., [50]), without having a predefined and uniform sampling frequency of the CSI vectors. Thus, in our case, every CSI observation is associated with its exact time of detection timestamp. In the environment, there was one controlled³ AP, whose location (near RP 9) was known, and configuration was nevertheless not changed for the study at hand. The controlled access point was transmitting on channel 7, at 2.4 GHz frequency with a 20 MHz bandwidth. Other APs, also transmitting on channel 7 but located in nearby

³ With controlled AP we mean that we know its location and, for experimental purposes, we can set its channel to match the one used by the board.



Fig. 3. Sensing equipment used for our IP Dataset.

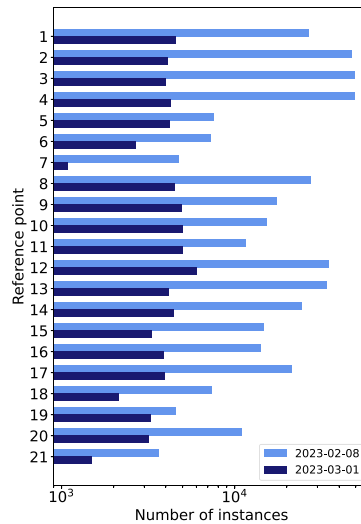


Fig. 4. Number of collected instances.

houses or structures, were collected, although they were not under our control. The person who conducted the experiment carried the smartphone (Fig. 3) in the hand for the duration of the trial (as this is the most natural setting for a person attempting to locate themselves), remaining 5 min at each of the 21 RPs marked on the floor plan of Fig. 5. The RPs represent semantically-meaningful locations (e.g., a table) rather than tiles of a pre-defined grid-based partitioning of the floor plan [51]. The orientation of the smartphone during the experiment was free. The house consists of three different floors, and RPs were present in all of them. The described setting was applied for both days of sampling campaigns. Fig. 4 shows the number of instances collected over the two days at each RP. As a final note, our dataset mirrors a real-world situation not only in its scenario but also in the amount of data collected: in fingerprinting, it is often impractical to gather multiple versions of the radio map over several days across vast premises. Consider, e.g., for the RSSI case, the dataset UJIIndoorLoc [52], where 3 buildings with an overall area of 108703 m² are covered with 933 RPs.

4.4. Our novel CSI signal temporal stability dataset

The second dataset we contribute is specifically designed to support studies on CSI signal stability over time. The hardware, software, and data collection protocol are identical to those used for the previous dataset (Section 4.3). The data was collected in a ~30 m² laboratory room at Jaume I University (Spain). The sensing device was fixed on top of a table, far from the windows,



Fig. 5. Floor plan with reference points for our IP Dataset.

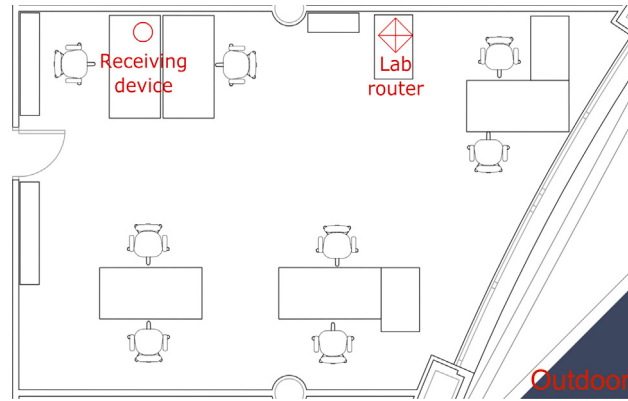


Fig. 6. Laboratory plan. The receiving device is placed at ~ 3.5 m from the AP.

capturing data transmitted by the laboratory AP over channel 7, also placed on a table at a distance of ~ 3.5 m (see Fig. 6). To ensure a comprehensive temporal analysis, data collection spanned over more than three weeks, from Friday, March 3, 2023, to Monday, March 27, 2023. This period included weekends and a local holiday week from March 13, 2023, to March 18, 2023, during which the laboratory remained closed, unoccupied, and with no HVAC system in operation. On workdays, typical laboratory activities occurred, including the presence of researchers and cleaning staff. Overall, we collected 5,769,240 samples, with an average of 230,769 samples per day (269,742 per day on workdays and 194,794 per day on holidays).

5. Multi-task empirical evaluation

Here, we experimentally validate the theoretical stance of Section 3 on the influence of time on CSI. We begin by confirming the results of [32] on device-free HAR and then demonstrate the temporal influence on the data used in that study. Next, for IP, using two reference datasets and employing fingerprinting, we show that various recent state-of-the-art methods fail to generalize over time, particularly in real-world scenarios.

5.1. Device-free HAR

In [32], a device-free HAR task is discussed, which exemplifies applications predominantly influenced by reflections from dynamic objects. The authors make use of SHARP [33], a state-of-the-art environment-independent activity recognition system. It is a supervised deep learning framework focused on a classification task: it takes, as its input, a sequence of CSI signals and then predicts an activity chosen from a predefined set. The main idea behind SHARP is to pre-process the raw CSI data to derive a so-called Doppler vector. The latter, unlike raw CSI, should be independent from environment conditions, and be influenced instead only by the activity being monitored. While a comprehensive description of SHARP is outside the purpose of this paper, we refer the interested reader to the original publication [33].

Model building and results reproduction. We first reproduced the results of the SHARP framework, based on the dataset described in Section 4.1. Three main activity recognition tasks are considered in the work of [32], but here we focus only on the PE case, where SHARP is trained with CSI data of a single person (e.g., person A) in a single environment (e.g., Lab) and then tested on other data from the same person in the same environment; our results considering 802.11ax and from one to four antennas are reported in Fig. 7. They position themselves intermediately between those of the authors of [32] and those observed in the original SHARP

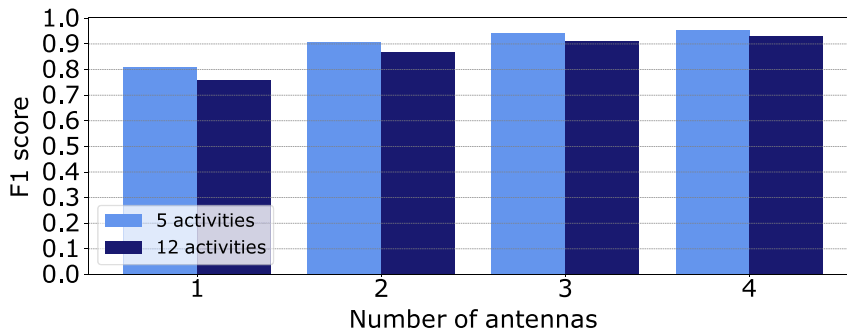


Fig. 7. SHARP applied to the datasets in [32], same person, same environment (PE) case. Performances improve as more and more antennas are employed.

Table 3

Results for HAR PE setting, 4 antennas, same and different time instants. Values: aggregated over the 3 APs (left); and, those of the overall best AP (right).

#Activities	Train	Test (F1, mean)			Test (F1, best AP)		
		S1	S4	S5	S1	S4	S5
5	S1	0.994	0.628	0.619	1.00	0.959	0.994
	S4	0.491	0.999	0.664	0.588	1.00	0.834
	S5	0.622	0.671	0.999	0.995	0.948	1.00
12	S1	0.985	0.210	0.199	0.990	0.318	0.369
	S4	0.213	0.970	0.285	0.290	0.914	0.413
	S5	0.281	0.338	0.981	0.480	0.529	1.00

contribution [33]: this variability/discrepancy is likely due to the strategy we employed to fuse the decision for multiple antenna systems (i.e., the one proposed in [33]). Overall, this experimental step provided us with a validation of our trained models and we can now move to the study of temporal effects.

Study of temporal effects. We now show that the influence of temporal dynamics on HAR via SHARP is relatively limited. To such an extent, we consider all the distinct PE (same person, same environment) experiments involving scenarios S1, S4, S5 (Table 1). Let us recall that these have been collected at sequential time points, with S1 being gathered first, followed on the same day by S4, and on the day after by S5. Following the approach of the original work [32], we randomly split the data of each scenario into training (60%), validation (20%) and test (20%). Then, we perform the following experiments: train on S1 (training and validation split) and test on S1 (test split), S4 (full set), S5 (full set); train on S4 (training and validation split) and test on S1 (full set), S4 (test set), S5 (full set); train on S5 (training and validation split) and test on S1 (full set), S4 (full set), S5 (test set). Results for the 4-antenna case are reported in Table 3. As it can be noticed, the performance, averaging across APs (Table 3, left side), does experience some degradation when different time instants are considered (e.g., when the model trained on S1 is applied to S4 or S5). Still, considering the results pertaining to the best overall performing AP (Table 3, right side), the F1 score remains – often well – above 0.83 when considering 5 activities, with the notable exception of the model trained on S4 and tested on S1, that has an F1 score of 0.588. This outcome aligns with expectations, given that the usage of the Doppler vector in SHARP is specifically designed to mitigate noise and other environmental effects.⁴ Moreover, in our experiments we noticed that: (i) a large overfitting phenomenon is present and linked with the performance degradation,⁵ which may be mitigated by evaluating/validating the model, during training, on a different dataset (e.g., train on S1, validate on S4, test on S5), achieving a further boost in performance (5%–10%, not reported in the table); and, (ii) there is a considerable variability among the employed APs, as witnessed in Table 3 when comparing the F1 score of the overall best AP with the mean F1. For completeness, when 12 activities are taken into account, the performance discrepancy increases considerably, confirming the observation by Cominelli et al. [32] that, when transitioning to a much finer-grained HAR task, achieving complete generalization to new, unseen situations and time variations becomes challenging. As a final remark, recall that our focus here is not on achieving optimal performance but on replicating the results and studying temporal effects.

We now turn our attention to scenarios S1 through S5, but restricting to the activity E. EMPTY ROOM and a single receiver AP. These subsets of the scenarios share identical environmental conditions (the same, empty, mid-sized hall) and differ only in the time sequence in which the corresponding data were collected, which is indicated by their respective lexicographic order. Two questions naturally arise: do these scenarios, despite being virtually identical, exhibit some variation that can be used by the trained SHARP model to discriminate between them? If the answer is yes, could such variations be related to their time of collection?

⁴ For the sake of completeness note that, in the 5 activities case, employing a single antenna leads to a decrease in performance of 0.164 on average.

⁵ Let us recall that we used the default SHARP implementation, simply changing the epoch budget (200) and the early stopping patience (10), without performing any other hyperparameter tuning.

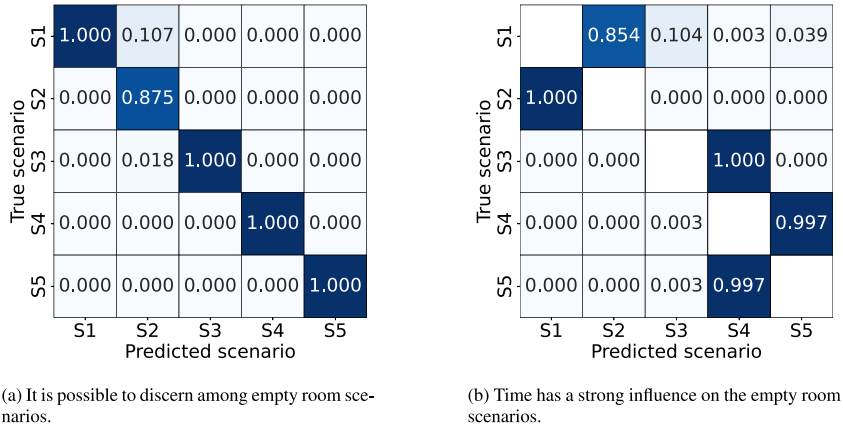


Fig. 8. Temporal influence in empty room scenarios. The same spatial environment (Lab) is considered, at different time points.

Let $S = \{S_1, S_2, S_3, S_4, S_5\}$ be the set of the five datasets associated with the corresponding scenarios. Each dataset S_i consists of n_i instances, where $(x_{i,j}, y_{i,j}, z_{i,j})$ represents the j^{th} instance in S_i with $x_{i,j}$ being the CSI feature vector, $y_{i,j} = i$ the label indicating the dataset to which $x_{i,j}$ belongs, and $z_{i,j} \in \{A, B, C, D, E, F, G, H, I, J, K, L\}$ the label indicating the activity to which $x_{i,j}$ refers. We now build $S' = \cup_{S_i \in S} \{(x_{i,j}, y_{i,j}) \mid (x_{i,j}, y_{i,j}, E) \in S_i\}$. Finally, we split S' into a training (S'_{train} , 60%), validation (S'_{val} , 20%) and test (S'_{test} , 20%) set, in a stratified fashion based on $y_{i,j}$, implying that training, validation and test sets will have the same proportion of data from S1, S2, S3, S4 and S5. We train and evaluate SHARP accordingly, obtaining the results reported in Fig. 8(a). The model clearly distinguishes among the instances belonging to the original datasets. This means that data from the same dataset is more similar to each other than to data from different datasets, despite all in principle representing the exact same situation.

Having established that there is indeed a variability among scenarios S1–S5 that can be successfully exploited by SHARP to distinguish them, could such a variability be attributed to their time of collection? In other words, do scenarios that are collected in closer temporal proximity exhibit greater similarity? To answer that, we proceed as follows. Let S' be the same dataset as in the previous paragraph. The goal is to obtain a set $\{M_1, M_2, M_3, M_4, M_5\}$, where each M_i is a SHARP model trained and then tested as follows:

1. training set is $\{(x_{i,j}, y_{i,j}) \mid (x_{i,j}, y_{i,j}) \in S' \wedge y_{i,j} \neq i\}$;
2. test set is $\{(x_{i,j}, y_{i,j}) \mid (x_{i,j}, y_{i,j}) \in S' \wedge y_{i,j} = i\}$;
3. intuitively, the goal of M_i is to classify instances of dataset S_i by assigning them labels in $\cup_{k \neq i} k$.

Note how each M_i is trained without any knowledge of the label associated to S_i . Thus, as the considered datasets all represent the same physical scenario except for the temporal perspective, we can determine the extent of temporal-related information by looking at how each model M_i classifies the instances of S_i . From Fig. 8(b) it is immediate to see that the instances of each dataset S_i are classified as belonging to one of the closest datasets in time among those observed during training. Thus, time implicitly characterizes data, even in conditions where this is not expected. The sources of this temporal variability require further attention and investigation from the research community with a more specialized dataset, which is still lacking. Potential factors could include fluctuations in ongoing Wi-Fi connections or interference from other wireless signals, such as Bluetooth.

Main takeaways on HAR. We confirmed that temporal factors do impact the CSI samples within the dataset under study. However, this influence only results in a partial decline in coarse-grained HAR performance, even when evaluating instances one day apart. Empirically, this is not surprising; although we did not include the generalization results for datasets S6 and S7 from [32], we refer the reader to [33], where extensive experiments demonstrate that SHARP successfully generalizes to different spatial environments. This ability to adapt to new environments suggests that the key features leveraged by HAR models are inherently robust to spatial variations. Since spatial generalization requires the model to capture fundamental activity-related patterns rather than environment-specific artifacts, it is reasonable to infer that this robustness also extends over time—even across very long periods—within a fixed environment, well beyond the time span considered in our experiments. Still, unlike previous literature, we explicitly considered the temporal aspect, providing specific evidence to support what was expected by the theoretical modeling (Section 3): HAR primarily leverages differential information between CSI vectors, a feature strongly influenced by dynamic objects and their reflections. These reflections exhibit robust patterns, which can be exploited to ensure that the HAR task is not highly susceptible to degradation over time, provided that temporal aspects are properly managed, as in SHARP. Performance degradation becomes more pronounced when fine-grained HAR is taken into account. In such cases, it is essential to investigate the underlying sources of errors, which warrants further attention from the research community also through the development of specialized datasets.

5.2. Device-based fingerprint-based IP

In the previous section, we determined that time has an effect even on environments where no obvious dynamic elements are present (E. EMPTY ROOM). While the temporal factor affected the performance of HAR, it did not lead to a total failure, at least when no very fine-grained activities were considered. What about indoor positioning? Here, we specifically focus on fingerprinting, which is, to-date, still the most commonly employed technique in real-world scenarios [53,54].

Fingerprinting [55] comprises two phases: the offline phase involves collecting CSI vectors from specific RPs and storing these as (CSI vector, position) pairs in a radio-map database; the online phase matches a CSI vector from an unknown location to the radio-map to estimate position. Finding the most similar instances in the radio-map can be done based on distance metric calculation, or employing other techniques, such as machine/deep learning approaches.

Differently from HAR, device-based IP predominantly relies on reflections originated from static objects and that remain static over time. This stems from the fact that dynamic objects, like people moving within a room, are not meaningful to characterize the environment surrounding a device, since they are transient. On the other hand, signals reflected by (nearly) static objects like walls, doors, tables, etc. are (almost) persistent through time, and thus should provide a unique characterization of a specific location.

Considered approaches. Various strategies for CSI-based indoor positioning have been proposed (Section 2), each distinguishable by the utilized features (signal phase, amplitude, or both), feature preprocessing type (e.g., filtering, denoising, auto-correlation, or neural network-based), and the ultimate localization methodology (e.g., k-NN or deep neural network-driven classification/regression). While both the feature preprocessing/extraction and the final localization can employ analytical or machine learning techniques – with every potential combination thereof being proposed in literature, each asserting its own merits – evaluating all possible variations is impracticable. Consequently, we consider a cohort of well-established and methodologically different approaches to analyze CSI temporal effects, intending to encompass the most pertinent alternatives explored in recent years, which are now summarized.

DeepFi. This approach, proposed in [16], harnesses CSI amplitude via a set of Deep Belief Network (DBN) autoencoders, each constructed from unsupervised Restricted Boltzmann Machines (RBMs). During training, the encoders' RBMs are pre-trained layer-wise through a greedy algorithm, with the acquired weights also initializing the decoders' weights by unrolling. Each DBN, fine-tuned for a reconstruction task using the Mean Absolute Error (MAE) loss function, attempts to reconstruct an input CSI amplitude vector. Importantly, individual DBNs are trained separately for each specific location. For online localization, a series of possible reconstructions for a new CSI vector is computed using all pre-trained models. The vector's likelihood of pertaining to a specific location is estimated using the Radial Basis Function (RBF) kernel to compare the original and reconstructed vectors, following the theory that the optimal reconstruction is attained via the DBN modeling the correct location. The method, accommodating multiple CSI vectors by averaging the RBF kernel values, has showcased consistent performance across varying numbers of antennas, sampling grid sizes, and number of CSI vectors. DeepFi was chosen for our study due to its academic recognition and proven superiority over both traditional and Received Signal Strength (RSS)-based methods across diverse scenarios. Hyperparameters were determined via random search using the Optuna library [56].

Neural-Network-Driven Feature Learning. In the study presented in [13], a positioning method using raw single-input multiple-output (SIMO) CSI data is introduced, wherein features are autonomously learned via a deep neural network within an end-to-end architecture aimed at classifying instances into predefined locations. This network, inspired by a manual three-step feature extraction—comprising (i) delay-domain transform via Discrete Fourier Transform (DFT), (ii) DFT-based autocorrelation, and (iii) L2-norm CSI feature normalization—initializes weights to mirror these steps, subsequently refining them through training. Compared with a hand-wired feature design and a neural network applied directly to the data, this approach, particularly when utilizing a sequence of 3 CSI instances, demonstrated superior performance. Our experimentation, adhering to this architecture and employing the Optuna library for hyperparameter tuning, considers combinations of feature preprocessing (none, hand-wired, or learned from CSI data) and varying CSI sequence lengths and APs. Although initially meant for passive device-based IP (localization performed by the AP), the methods were later successfully applied to an active device-based IP (localization performed by the user device) scenario by the original authors [14], supporting their application to our dataset. To handle the MIMO case in the considered literature's IP dataset [34], Section 4.2, we generalize the approach by constructing a separate model for each receiver's antenna and subsequently averaging the predictions.

TRRS. The time-reversal resonating strength (TRRS) leverages the time-reversal signal processing technique, which aims to mitigate the phase distortion of a signal. It is based on the fact that such a distortion can be removed when the signal, at a given time instant, is combined with its time-reversed and conjugate counterpart [12]. Given two, single link CSI samples in the frequency domain $\mathbf{H}, \mathbf{H}' \in \mathbb{C}^N$, the TRRS ϕ between them is expressed by:

$$\phi(\mathbf{H}, \mathbf{H}') = \frac{\max_{\epsilon} \left| \sum_{k=1}^N \mathbf{H} \cdot \mathbf{H}' e^{j\epsilon N} \right|^2}{\langle \mathbf{H}, \mathbf{H} \rangle \langle \mathbf{H}', \mathbf{H}' \rangle}, \quad (8)$$

which compensates for both the sampling frequency and symbol timing offsets. In [12], ϵ is estimated by an FFT-based algorithm; then, the positioning phase is carried out by comparing an instantaneous CSI $\mathbf{H}'$ with the set of CSI samples collected at training time and for which the location is known. If more than one link is present (e.g., in the literature IP dataset), a generalization to multiple links, with the idea of fusing the TRRS of single links, is employed [12]. With respect to the original implementation, which can also output an *unable to localize* verdict, ours always generates, as a positioning estimate, the location observed at training time with the highest TRRS (which is, thus, a similarity measure).

Table 4
Results on positioning, dataset [34].

Ref.	Details	Same time instant		Closer time instant		Farther time instant	
		$F1_{mic}$	$F1_{mac}$	$F1_{mic}$	$F1_{mac}$	$F1_{mic}$	$F1_{mac}$
[13,14]	#AP = 1, #S = 1, Pp = none	1.0	1.0	0.89	0.89	0.60	0.58
	#AP = 1, #S = 1, Pp = wired	1.0	1.0	0.70	0.69	0.60	0.59
	#AP = 1, #S = 1, Pp = learn	1.0	1.0	0.72	0.72	0.61	0.60
	#AP = 1, #S = 3, Pp = none	1.0	1.0	0.74	0.73	0.58	0.57
	#AP = 1, #S = 3, Pp = wired	1.0	1.0	0.69	0.68	0.65	0.64
	#AP = 1, #S = 3, Pp = learn	1.0	1.0	0.69	0.68	0.61	0.60
[12]	TRRS, #AP = 1	1.0	1.0	0.90	0.90	0.59	0.60
[16]	DeepFi, #AP = 1	0.98	0.98	0.72	0.72	0.50	0.51

Where applicable: #AP = number of access points; #S = number of CSI instances; Pp = kind of preprocessing.
mic = micro average; mac = macro average.

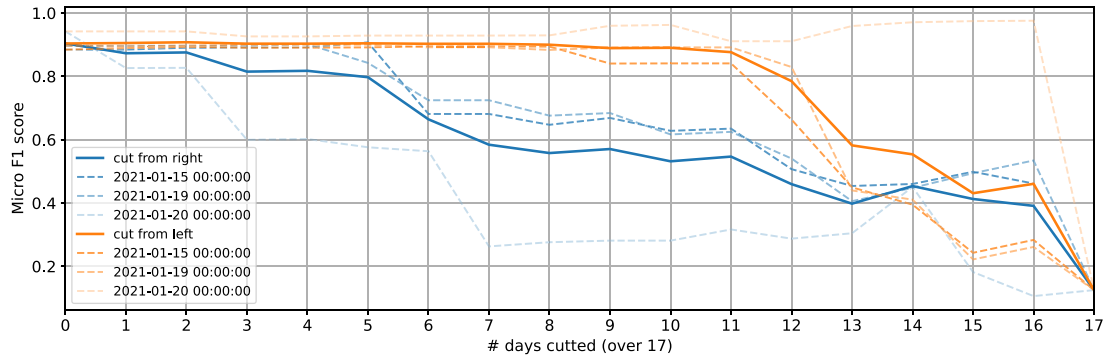


Fig. 9. TRRS performance when trimming the training days of Split A, dataset [34].

5.2.1. Experiments on the literature IP dataset

We begin by focusing on the dataset described in Section 4.2, which encompasses the most controlled scenario. In all the next experiments, we framed the positioning problem as a classification task, where an algorithm predicts a location from a given CSI-based input.

Validation of the models. To validate our implementations, we focus on the Split A of Table 2, where the test set is only a few days apart from the training and validation sets, and for which positioning performances have been presented in [44]. The results of this first experiment are reported in Table 4 (Closer time instant). Based on the considered metrics, almost all the algorithms indeed reached or exceeded the performances presented in [44].⁶ Notably, there was no clear advantage in using learnable feature extraction over a hand-wired approach in the method of [13,14] (first block of rows in the table).

Study of temporal effects. Having confirmed the correctness of our implementation, we now analyze the temporal effects on the positioning task. Previously, we discussed the case of Split A, where the test set is only a few days apart from the training and validation sets. What happens when such a temporal gap is extended? To investigate this, we again use the models trained and validated on Split A and apply them to predict the test instances of Split B, which are from a time period 2 months later. As shown in Table 4 (Farther time instant), the performance metrics decrease significantly. Conversely, training, validating, and testing the models on instances from the same time interval should yield high performance. To verify this, we focused on the training set of Split A and randomly partitioned it into 70% training, 20% validation, and 10% test subsets, in a stratified fashion over the RPs. We then retrained, tuned and tested the models on the new split. The results, presented in Table 4 (Same time instant), align with our expectations, with almost perfect F1 scores.

At this point, it is evident that time significantly influences the CSI patterns used for IP, and that *the accuracy of IP decreases over time*. This finding is highly significant, especially in light of the many recent studies that do not account for the temporal dimension when splitting their experimental data into training and test sets, while sometimes claiming sub-meter positioning accuracy (see, for example, [18,19,22,23]).

Let us now investigate more systematically how the temporal distance between the training and test set affects performance. We focus again on the Split A and, for the sake of clarity and conciseness, we consider a single technique, TRRS.⁷ Recall that the training instances of Split A span 17 days, from November 17th to December 30th, while the test instances were collected, over three

⁶ Not reported in the table. Please refer to the original publication.

⁷ Repeating the experiments with the other techniques yields very similar results.

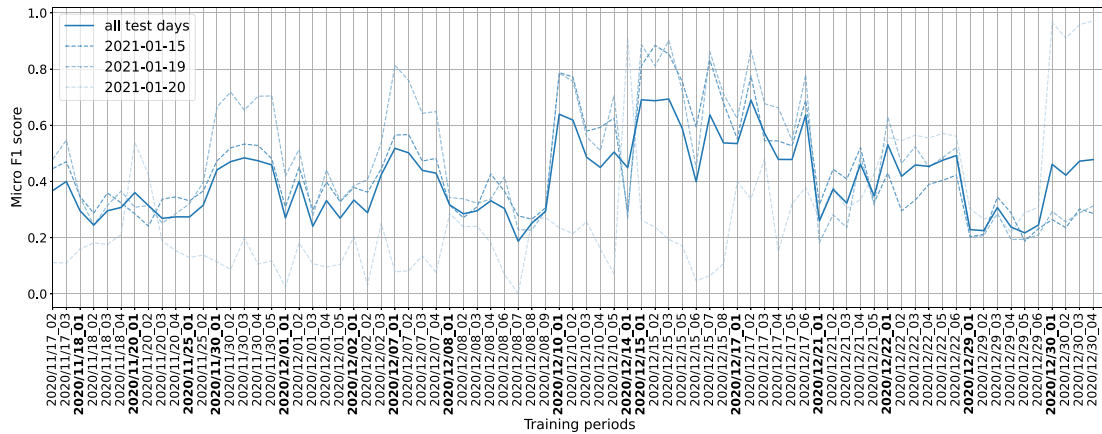


Fig. 10. TRRS performance when considering the 77 training periods of Split A independently, dataset [34].

days, half a month later. What happens if we start removing training days in two different ways, i.e., trimming the dataset from the beginning (starting from November 17th) and from the end (starting from December 30th)? Fig. 9 presents the results. The average performance is depicted for the left (bold orange line) and the right (bold blue line) trimming case. Additionally, dashed lines show the results for each of the three test days separately. Cutting from the end, thus temporally distancing the training set from the test set, results in a sharper decline in overall performance when compared to cutting from the beginning. Notably, the three test days show different behaviors, particularly January 20th, suggesting varying importance of training days for the prediction task, with January 20th highly dependent on the most recent ones. Fig. 10 provides a more in-depth analysis by considering each collection period as an independent training set. Performance varies significantly, and indeed the last training periods are most important for January 20th, but not for the other two test days. This highlights a relevant and counterintuitive point: *the most recently collected data are not always tied to a better positioning performance*.

Could this apparently random variations be linked to differing conditions of obstructions and crowding between training and test data? Let us focus on the most common condition (93% of the time) in the test instances of Split A: *obstructions='Closed door'* and *people='1 on Table 1'*, named condition *C*. This is also the simplest crowding and obstruction condition in the dataset. Firstly, based on the accessory data provided with the dataset, we observe that the periods belonging to training days December 22nd and December 30th share the same condition *C*, yet they behave very differently, particularly in relation to the test day January 20th (see again Fig. 10). This raises a new question: is it true that when the same condition *C* is present in two CSI observations, the TRRS similarity between them is higher? To address this, we compare the similarity between CSI observations from the same RP in three situations: when taken from a same period; when taken from different periods but under the same condition *C*; and when taken from different periods with different conditions. Since it is unfeasible to consider all possible pairs of CSI observations in the dataset, we randomly select a subsample of 20,000 pairs for each case. Again for the sake of conciseness, we focus on RPs 1, 2, 7, and 8, although our findings extend to the other RPs as well. Fig. 11 presents the results. Clearly, CSI observations from the same RP and period exhibit a high degree of similarity. However, similarity values decrease significantly and become more heterogeneous when different periods are considered, even under the same condition *C*. This leads to the observation: *scheduling radio-map sampling campaigns for specific environmental conditions is generally impractical and of questionable usefulness, as their effects on CSI patterns are not easily controllable, offering little guarantees on the resulting IP performance*.

Finally, recall that in the considered dataset, the receiver is equipped with three different antennas, which are consistently positioned and aligned throughout the data collection campaign. Consequently, one might ask whether they deliver similar performance. Fig. 12 considers the same data as *Closer time instant* and *Farther time instant* scenarios in Table 4, but restricting to single antennas. The performances of the three antennas, particularly in all the configurations of the approach [13,14], vary. Anyway, all single-antenna performances are lower than those in Table 4, indicating that multi-antenna and, in general, MIMO, are likely to offer better positioning results.

5.2.2. Experiments on our IP dataset

We now analyze our novel IP dataset (Section 4.3), designed to represent a more real-world scenario than the previous one. Here data was gathered, with no strict protocol, by a single-antenna receiver at a private home, detecting multiple APs. The data collection occurred over two separate days, approximately one month apart.

Validation of the models. Again, we start by considering CSI data collected on a same day (specifically, the first day) of our dataset. We divided such data into a training (70%), validation (20%) and test (10%) datasets, in a stratified fashion over the RPs. In this setting, we expect the performances of the various approaches to be high. They are reported in the *Same time instant* part of Table 5. Based on all metrics, almost all the considered algorithms indeed perform very well. A notable exception is the approach by [13,14], when no pre-processing is performed ('Pp = none') and 'S = 3'. This contrasts with the results in Table 4, where performances for

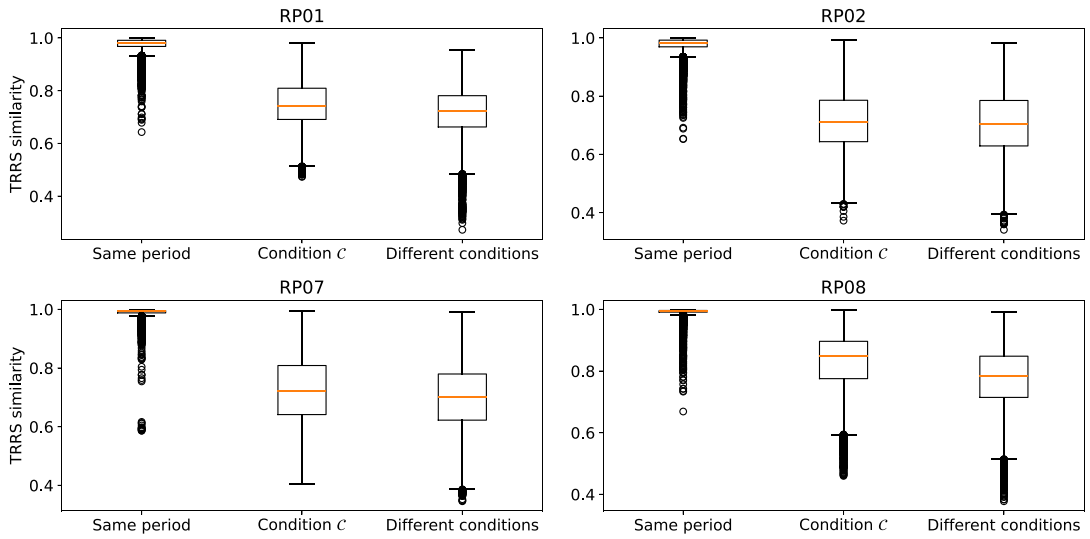


Fig. 11. TRRS similarity between CSI observations at different RPs and conditions, dataset [34].

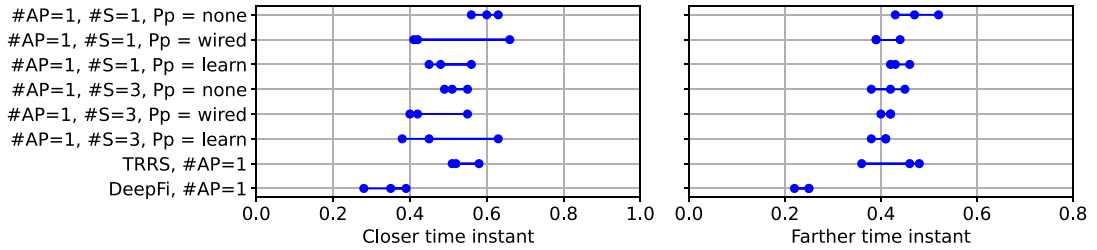


Fig. 12. Results on positioning (micro F1 score) for the single antennas (blue dots), dataset [34].

Table 5
Results on indoor positioning, own dataset.

Ref.	Details	Same time instant			Different time instant		
		$F1_{mic}$	$F1_{mac}$	$F1_{whg}$	$F1_{mic}$	$F1_{mac}$	$F1_{whg}$
[13,14]	#AP = 1, #S = 1, Pp = none	0.72	0.57	0.70	0.15	0.10	0.12
	#AP = 1, #S = 1, Pp = wired	0.96	0.94	0.96	0.08	0.06	0.07
	#AP = 1, #S = 1, Pp = learn	0.99	0.99	0.99	0.08	0.06	0.08
	#AP = 1, #S = 3, Pp = none	0.11	0.06	0.07	0.05	0.03	0.03
	#AP = 1, #S = 3, Pp = wired	0.89	0.84	0.89	0.07	0.05	0.06
	#AP = 1, #S = 3, Pp = learn	0.98	0.96	0.98	0.07	0.06	0.07
	#AP = 2, #S = 1, Pp = none	0.70	0.55	0.68	0.16	0.11	0.13
	#AP = 2, #S = 1, Pp = wired	0.95	0.92	0.95	0.07	0.06	0.07
	#AP = 2, #S = 1, Pp = learn	0.99	0.98	0.99	0.07	0.05	0.06
	#AP = 2, #S = 3, Pp = none	0.06	0.03	0.03	0.07	0.03	0.04
	#AP = 2, #S = 3, Pp = wired	0.87	0.80	0.86	0.06	0.05	0.06
	#AP = 2, #S = 3, Pp = learn	0.97	0.96	0.97	0.07	0.05	0.06
[12]	TRRS, #AP = 1	0.89	0.85	0.89	0.07	0.05	0.06
[16]	DeepFi, #AP = 1	0.97	0.93	0.97	0.10	0.05	0.06

Where applicable: #AP = number of access points; #S = number of CSI instances; Pp = kind of preprocessing.

the same case were instead optimal. This discrepancy could be attributable to the fact that, in the previous dataset, subsequent CSI observations were collected at regular time points, whereas in ours they are more arbitrarily spaced in time, resulting in greater diversity. The latter seems to be better managed by a dedicated data pre-processing phase ('Pp = wired' and 'Pp = learned' in the table). Repeating the experiment on data from the other date produced very similar results, and therefore, those numbers have been omitted.

Study of temporal effects. Now, let us evaluate how much the performance of IP in our real-world setting is affected by temporal factors. From Table 5 we can see that all the considered approaches, which were performing very well when applied to test instances

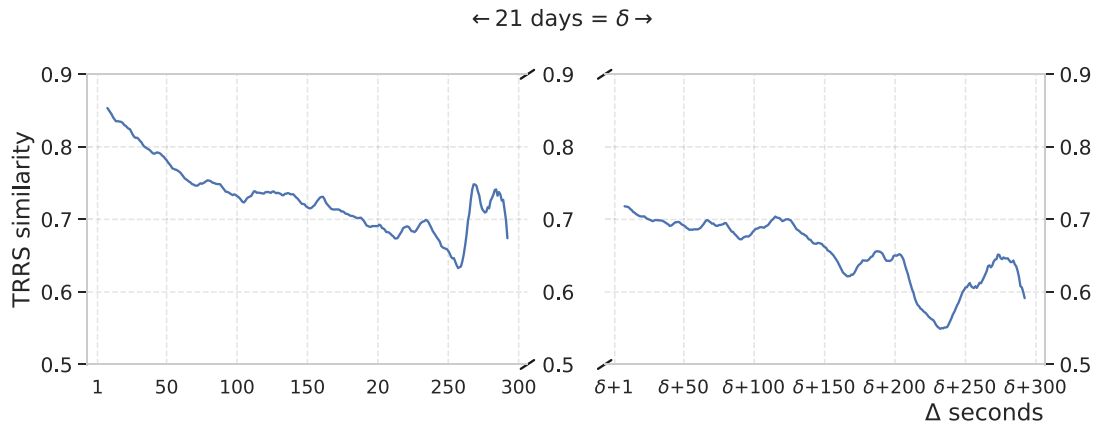


Fig. 13. 95th percentile of TRRS similarity over time distances (Savgol filter, window size 15, polynomial order 3 applied for the ease of reading), own dataset.

collected on the same day of training ones, completely fail when evaluated on instances gathered almost one month later. It is a rather worse result than the one witnessed with the previous dataset; still, it should not come as a surprise, since ours represents a less controlled scenario. We can thus summarize: *when confronted with a real-world scenario, the considered approaches completely fail to generalize over time*. This further remarks the necessity of correctly splitting between training and test data, taking into account the temporal dimension, when developing and testing a new approach.

Let us now, as we did earlier, analyze more in depth the temporal effects: how do CSI patterns related to static objects evolve over time? Our dataset, unlike [44], includes timestamps for each CSI instance, allowing us to investigate finer temporal influences. We focus on RP 8, since it is centrally located in the premises (Fig. 5), and use TRRS. Fig. 13 shows how TRRS similarity changes when calculated between pairs of CSI instances collected at varying temporal distances. The graph presents the 95th percentile of similarities,⁸ highlighting that similarity sharply decreases over the initial 300-s collection interval on the first day. When comparing instances collected nearly a month later to those from the initial day (i.e., temporal distances higher than 300 s), we observe on average even lower values, especially compared to the case where the CSI pairs were collected at most 50 s one from the other. Of course, while the overall trend may be decreasing, the time series representing TRRS is still expected to exhibit periodic or random fluctuations, leading to local increases or decreases relative to the trend. Such fluctuations can arise from various factors, such as temporary environmental changes (e.g., furniture rearrangements, minor structural modifications, or variations in electromagnetic interference from nearby devices), or differences in human activity patterns.

We have determined that, in a real-world scenario, temporal effects significantly impact TRRS similarity, even over short time intervals. Thus, how would such a measure behave when used to discriminate between CSI fingerprints belonging to different RPs? Fig. 14 offers insights by showing the similarity between CSI samples of RP 8 and those from all other RPs at various collection times. When the temporal distance between CSI samples is under 30 s, distinguishing RP 8 samples from the others is straightforward. However, this already becomes visually impossible when considering temporal distances from 30 to 300 s and, similarly, instances collected on the two different days. These results align with the positioning performances in Table 5 (*Same time instant*), and overall mean that *with the considered approaches, maintaining an adequate radio map in a real-world scenario would require continuously gathering new labeled data*, as a significant portion of the temporal variability is already introduced within the first few moments of data collection.

Finally, Fig. 15 generalizes our findings to the other RPs. A clear diagonal is visible in Fig. 15(a), meaning that instances from a given RP x are more similar to those gathered at x itself than those from a different RP $y \neq x$, provided the sampling time difference remains under 30 s. The scenario degrades rapidly in Fig. 15(b), which examines a time interval from 30 to 300 s and barely displays the diagonal. The latter vanishes entirely in Fig. 15(c), which accounts for a nearly one-month time differential.

Main takeaways on IP. The effectiveness of state-of-the-art IP methods is deeply impacted by temporal factors. Performance diminishes as the temporal distance between radio-map instances and the fingerprint to be localized increases. This degradation is limited in controlled settings but becomes severe in real-world ones, even at very short temporal distances—where, in contrast, HAR exhibited significantly less degradation. Experiments also showed that controlling environmental conditions and interferences on CSI patterns is not fully possible, meaning that collecting newer fingerprints does not always guarantee improved positioning performance, especially as we showed that the most recently collected data do not always yield a better positioning performance. Overall, our results underscore the need for approaches to be both developed and tested with proper consideration of the temporal dimension in the data.

⁸ Focusing on the highest similarities is sensible, since in fingerprinting the goal is to identify radio-map instances that are most similar to a given one.

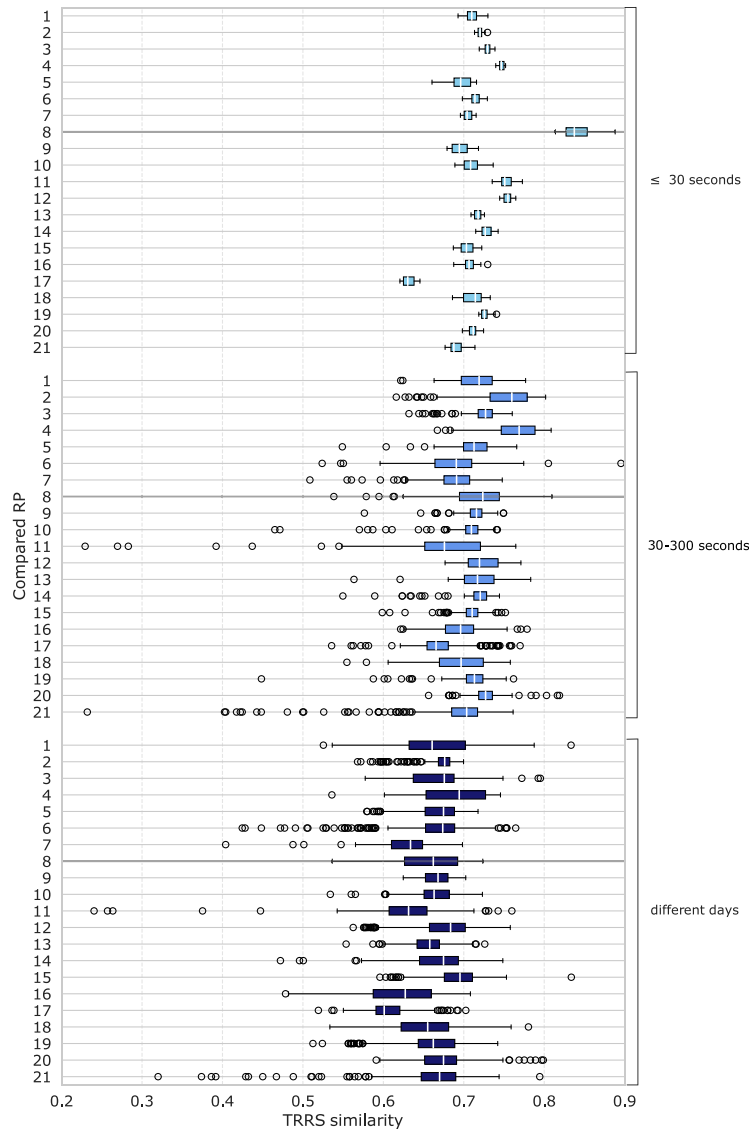


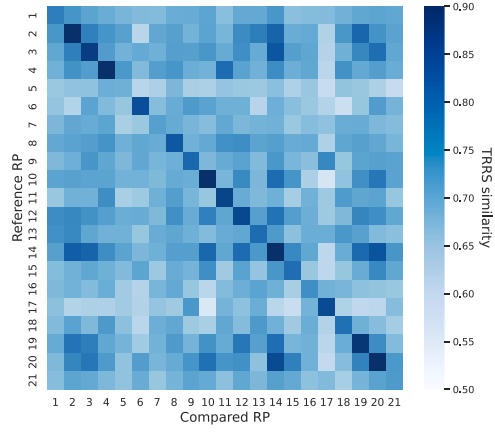
Fig. 14. TRRS similarity between RP 8 and other RPs. Boxes: 1st–3rd quartile. Line: median. Whiskers: farthest point within 1.5x inter-quartile range.

5.3. CSI stability over time

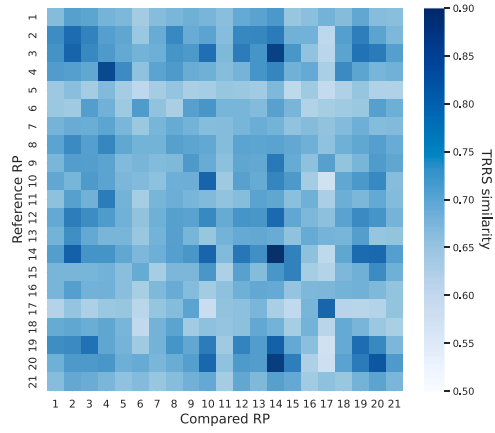
Experiments over IP (Section 5.2) demonstrated that even within a single environment, static CSI components can undergo significant changes over time. However, the datasets used in those experiments were not designed for continuous and extended monitoring of a given environment while precisely distinguishing between endogenous and exogenous factors that may influence CSI. Endogenous factors refer to human-related activities occurring within the observed environment, such as, for instance, people moving, changes in the arrangement of objects or furniture, or the operation of HVAC systems to regulate room temperature. Exogenous factors, on the other hand, originate outside the observed environment and are not directly linked to human activity within it, such as weather conditions or nearby car traffic.

The dataset described in Section 4.4 enables a more detailed investigation. Notably, during holidays and weekends, when the building is unoccupied and the HVAC system is inactive, the influence of endogenous factors is significantly reduced. In contrast, on workdays, human presence introduces changes such as furniture rearrangements, varying crowd levels, and adjusted thermal conditions, thereby intertwining both endogenous and exogenous factors. This observation prompts two questions:

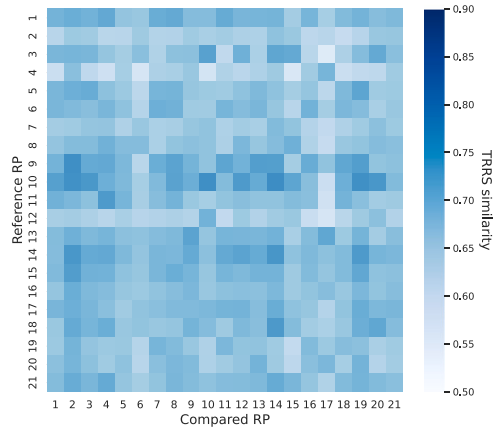
- Over multiple days of continuous data collection, can we detect an impact on the CSI attributable mainly to exogenous factors?
- In scenarios primarily influenced by exogenous factors, does CSI similarity, as measured by TRRS, continue to evolve changing shortly after the reference samples are collected?



(a) Up to 30 sec.



(b) 30–300 sec.



(c) Different dates.

Fig. 15. Median TRRS similarity for instance pairs gathered at different time distances across all RPs.

To address these questions, we analyze three continuous 5-day periods (Monday–Friday, 00:00–23:59). The first two periods (from March 6, 2023, to March 10, 2023, and from March 20, 2023, to March 24, 2023) represent typical weekday conditions, while the third period (from March 13, 2023, to March 17, 2023) corresponds to a holiday scenario. For each period, we proceed as follows. We build a set S_0^* of 3000 randomly sampled CSI instances from the first hour. Then, we randomly sample 3000 CSI instances per hour, obtaining 120 independent sets $S_0, \dots, S_{119}$. For each pair (S_0^*, S_i) , $i = 0, \dots, 119$, we compute all pairwise TRRSs

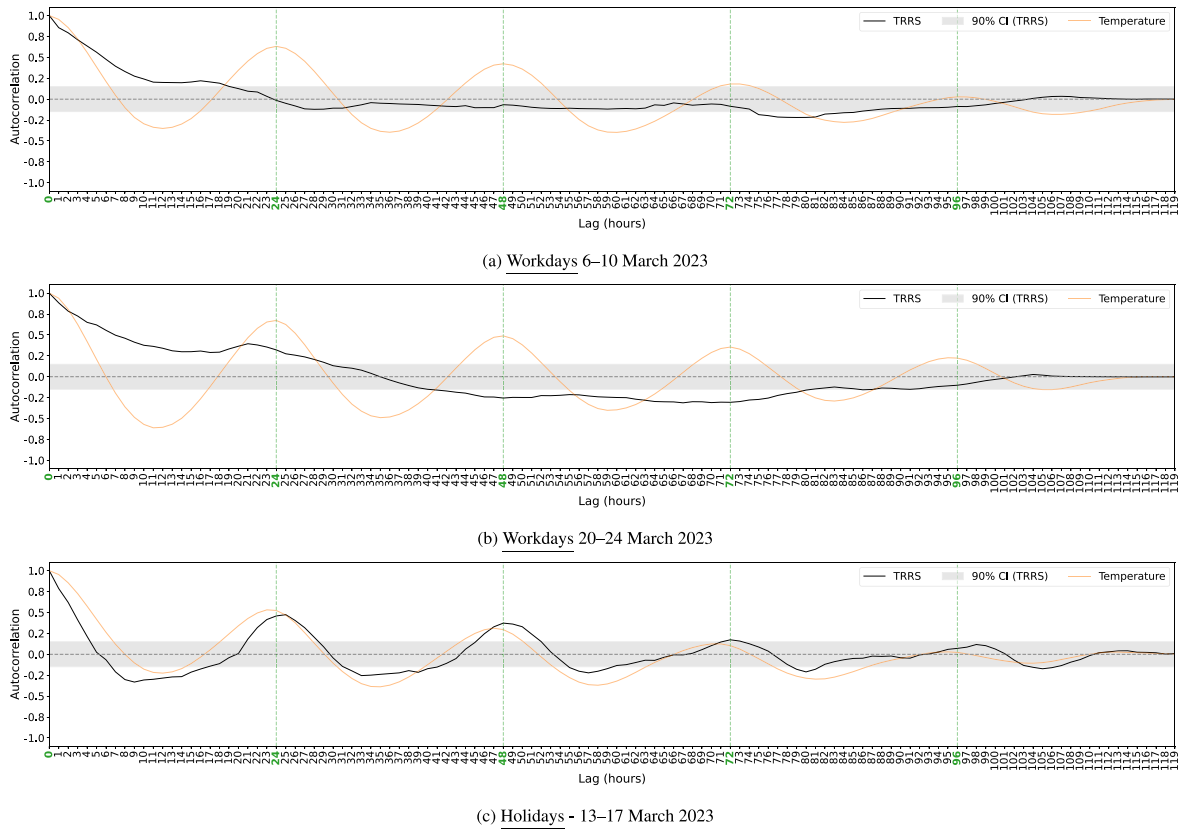


Fig. 16. Correlograms showing the autocorrelation coefficients of the median TRRSs time series (black) and external temperature (orange) over the considered periods. The 90% confidence interval for TRRSs autocorrelation is shaded in gray. Autocorrelation values within this interval are not statistically significant.

between instances of S_0^* and S_i . This resulted in 120 sets of TRRS values. Next, we calculated the median TRRS for each set, forming a time series that tracks how CSI similarity (in terms of TRRS) evolves over time relative to the first hour, spanning a time range from the 1st (index 0) hour to the 120th (index 119) hour. We computed the autocorrelation function (ACF) of the time series, which measures how correlated the values are to themselves at different time lags:

$$ACF(y) = (r_0, \dots, r_{T-1}) \quad (9)$$

$$r_k = \frac{\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}, \quad (10)$$

where r_k is the k th autocorrelation coefficient, that is, the coefficient for a lag of k , y is the TRRSs time series, T is the length of y , and $\bar{y}$ is the mean of the time series. For instance, a strong autocorrelation at lag $k = 24$ would indicate a daily periodicity in TRRS similarity, suggesting recurring, seasonal patterns in CSI behavior over a 24-h cycle. Finally, for each period, we extracted the historical weather temperature data,⁹ as recorded by an outdoor weather station placed at ~7 km distance from the University.

Figs. 16(a) and 16(b) show the correlograms for the two workdays periods. In both of them, a negative trend is present. Shorter lags have a strong positive correlation, meaning that observations closer in time have similar TRRS values. Then, the correlation tapers off as the lags increase. Given our previous experimental results, this sharp decline in correlation over time is expected, and it highlights how both endogenous and exogenous environmental changes accumulate, leading to increasingly different TRRS values over time. Note that the autocorrelation of the TRRS time series appears to be unrelated to that of the external temperature, which is not surprising given that the University's HVAC system was in operation in those periods; any residual effect of the external temperature may have been overshadowed by much stronger endogenous factors. In contrast, Fig. 16(c) shows a clear periodic (seasonal) pattern during holidays, where exogenous factors predominantly influence the system. Here, the TRRS time series exhibits stronger autocorrelation at lags that are multiples of the seasonal frequency (24 h) compared to other lags (e.g., 5–10 h). In other words, CSI samples collected at approximately the same time across different days are more likely to exhibit similar TRRS values (when exogenous factors are predominant). This pattern closely aligns with the seasonality observed in the temperature

⁹ Retrieved from <https://open-meteo.com/>.

autocorrelation time series, which depends on natural daily cycles. Of course, in the holiday period, outdoor weather has an impact also on the indoor conditions, given the non-operative HVAC system. In addition, such an impact is not influenced by endogenous factors, which are at their minimum. As a final observation, note that the autocorrelation values in Fig. 16(c) never return to the high levels observed for temporally proximate samples (e.g., one or two hours apart). This further emphasizes how TRRS values tend to diverge as the temporal separation increases, even though some periodicity may occur.

Main takeaways on CSI stability. Overall, our empirical findings suggest that even under seemingly static conditions, exogenous factors significantly influence CSI, accumulating over time. However, if not properly accounted for in the analysis workflow, these effects may go unnoticed, as they can easily be overshadowed by endogenous ones.

6. Discussion and possible research avenues

By studying HAR and IP, with a particular focus on the real world applicability of these tasks and the influence of temporal changes, we were able to get the following insights:

- Temporal factors significantly impact CSI samples, with effects on both HAR and IP;
- Performance degradation (witnessed when time is properly taken into account during testing) is considerably more marked for IP than HAR, possibly because the former exploits static CSI components instead of the dynamic ones used by the latter;
- Multi-antenna and, in general, MIMO, are likely to offer better results;
- The effect of time is clearly detectable even in situations where the environments are semantically equivalent from the perspective of the task at hand;
- Even assuming an environment entirely devoid of endogenous factors (e.g., furniture rearrangement, people) that may affect CSI, exogenous factors (e.g., weather) still exert a clear influence;
- The most recently collected data are not always tied to a better (positioning) performance;
- Scheduling radio-map sampling campaigns for specific environmental conditions is generally impractical and of questionable usefulness, as their effects on CSI patterns are not easily controllable, offering little guarantees on the resulting IP performance;
- Given the previous points, maintaining an adequate radio map in a real-world scenario would require continuously gathering new labeled data, as a significant portion of the temporal variability is already introduced within the first few moments of data collection;
- For IP, when confronted with a real-world scenario, the considered approaches completely fail to generalize over time.

Based on the collected evidences, we can now answer our initial research questions.

RQ1: Do CSI-based wi-fi sensing tasks generalize over time? In general, tasks experience performance degradation over time. This is particularly true for those involving static CSI components, such as indoor positioning, where recent state-of-the-art methods fail to generalize over time. A less marked degradation is observed in tasks involving dynamic components, like HAR. Overall, the experimental results confirmed the theoretical expectations, underscoring the need for proper evaluation of any proposed approach, particularly by incorporating test data that are temporally distant from the training data — a consideration that remains too much overlooked in the current literature.

RQ2: How do patterns associated with CSI components evolve through time? As is well known, CSI is influenced by the environment. Observations of the same situation within the same, short time span are highly similar, but this similarity sharply decreases over time, even within minutes in real-world data. This occurs even under equivalent environmental conditions, like a same empty room. The causes of this temporal variability warrant deeper investigation by the research community, ideally using more specialized datasets. In this work, we paved the way for understanding the impact on CSI of exogenous factors in an environment devoid of strong endogenous influences, focusing on the weather. Other possible contributing factors may include fluctuations in active Wi-Fi connections, interference from other wireless technologies, such as Bluetooth, or diverse and unaccounted-for phenomena, such as people moving in nearby premises.

RQ3: What challenges does the real world present, and how should future research be directed to address them effectively compared to current state of the art? Our experiments revealed significant performance differences in CSI-based tasks between controlled and real-world environments. In real-world settings, CSI patterns are continuously influenced by unpredictable environmental effects. Thus, approaches that try to address this issue by collecting more data or performing supervised domain adaptation/transfer learning are not guaranteed to succeed, other than being impractical.

Current state-of-the-art shows a disparity between IP and HAR: despite extensive study, IP lags behind HAR in reproducibility and real-world usability. HAR benefits from public datasets (see, e.g., [32,57–60]) and community-accessible methods (see, e.g., [33,61]), with solutions working across different spaces and times being investigated, and promising results already being achieved, at least for what concerns coarse-grained human activity recognition. Conversely, work on indoor positioning is much less systematic and focuses on enhancing localization performance without exploiting any large, well-established, real-world datasets, often relying on new, private, laboratory-condition data [62,63].

Testing solutions in realistic settings is crucial for effective deployment. This calls for the creation of CSI benchmark datasets spanning different times and environments, with pragmatic data collection strategies, as exemplified by our testbed (Section 4.3) and RSS-based Wi-Fi fingerprinting studies [52,64]. The frequent claim in the literature of centimeter-level accuracies [38,62] is

instead typically based on datasets with very proximate training locations, but is such fine sampling granularity feasible in reality, with large premises and possibly very frequent radio map updates?

Backed by benchmark datasets, research can focus on enhancing the generalizability of CSI-based indoor positioning. This is essential for real-world applications, where, as already mentioned, frequent collection of labeled CSI samples is unrealistic and might rely on crowdsourced and unsupervised (i.e., unlabeled data) methodologies. Potential solutions may involve developing analytical methods that utilize features resistant to temporal changes, which requires a deeper theoretical understanding and modeling of CSI. Another promising, and orthogonal direction is leveraging modern deep learning paradigms to autonomously extract time-invariant data representations. This could be achieved by designing deep learning architectures that employ contrastive loss functions, encouraging the model to learn features that remain similar for spatially close instances collected at different times, while distinguishing between instances from different locations. Although gathering the necessary training data for deep learning models might be challenging, simulated and synthetic data could also be utilized effectively to supplement real-world data [65].

7. Conclusions

In this study, we explored CSI-based sensing, systematically and independently examining the temporal variability of CSI patterns from both a theoretical as well as an experimental point of view. Our experiments focused on Human Activity Recognition and Indoor Positioning tasks. For Indoor Positioning, we utilized two datasets: one from existing literature, capturing a controlled environment, and another we contributed, reflecting a real-world scenario. We also proposed and analyzed a dataset specifically meant to discern between endogenous and exogenous factors potentially influencing CSI. Our investigation led to several key insights (summarized in Section 6), highlighting that phenomena occurring during the natural flow of time can significantly impact system performance in ways that are often unpredictable and difficult to account for. This finding underscores the importance of testing systems with their temporal stability in mind, an aspect often overlooked even in the recent literature published in top-tier venues. Specifically, the sample-level random split procedure commonly used in the machine learning community should be avoided when handling CSI data, as it can result in data leakage and provide over-optimistic evaluation results. Our evidences strongly support the need to prioritize temporal factors in future research, development, and deployment of CSI-based sensing and localization solutions, making them essential considerations for robust performance.

CRedit authorship contribution statement

Andrea Brunello: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Angelo Montanari:** Writing – review & editing, Supervision, Funding acquisition. **Raúl Montoliu:** Supervision, Data curation. **Adriano Moreira:** Writing – review & editing, Supervision. **Nicola Saccomanno:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Emilio Sansano-Sansano:** Writing – review & editing, Data curation. **Joaquín Torres-Sospedra:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Andrea Brunello, Nicola Saccomanno, Angelo Montanari report financial support was provided by u-blox AG. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

J. Torres-Sospedra acknowledges funding from Generalitat Valenciana (CIDEXG/2023/17, Conselleria d'Educació, Universitat i Ocupació). E. Sansano-Sansano, J. Torres-Sospedra and R. Montoliu acknowledge funding from the Spanish Ministry of Science and Innovation through Project INDRI under Grants PID2021-122642OB-C42 and PID2021-122642OB-C44. A. Montanari, and N. Saccomanno acknowledge the support from the Interconnected Nord-Est Innovation Ecosystem (iNEST), which received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.5 – D.D. 1058 23/06/2022, ECS00000043). In addition, A. Montanari acknowledges the support from the MUR PNRR project FAIR - Future AI Research (PE00000013) also funded by the European Union Next-GenerationEU. Finally, A. Brunello also acknowledges the support from the University of Udine's Department of Humanities and Cultural Heritage for the project "Localizzazione tramite Wireless Sensing per un'Esperienza Museale Interattiva e l'Analisi Automatica dei Dati Relativi alle Visite" funded via the "Department of Excellence 2023-2027" initiative of the Italian Ministry of University and Research (MUR), and from u-blox AG for the project "Seamless Outdoor-Indoor Positioning Leveraging 5G Networks". This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

Data availability

The novel datasets used in this study is openly available at the following link: <https://doi.org/10.5281/zenodo.14212401>.

References

- [1] Jean Damascène Mazimpaka, Sabine Timpf, Trajectory data mining: A review of methods and applications, *J. Spat. Inf. Sci.* 13 (1) (2016) 61–99.
- [2] Neha Gupta, Suneet K Gupta, Rajesh K Pathak, Vanita Jain, Parisa Rashidi, Jasjit S Suri, Human activity recognition in artificial intelligence framework: A narrative review, *Artif. Intell. Rev.* 55 (6) (2022) 4755–4808.
- [3] Shuang Shang, Lixing Wang, Overview of WiFi fingerprinting-based indoor positioning, *IET Commun.* 16 (7) (2022) 725–733.
- [4] Yongsan Ma, Gang Zhou, Shuangquan Wang, WiFi sensing with channel state information: A survey, *ACM Comput. Surv.* 52 (3) (2019) 1–36.
- [5] Xinpeng Rao, Zhenzhen Luo, Yong Luo, Yugen Yi, Gang Lei, Yuanlong Cao, MFFALoc: CSI-based multi-features fusion adaptive device-free passive indoor fingerprinting localization, *IEEE Internet Things J.* 11 (2023) 15.
- [6] Souvik Sen, Božidar Radunovic, Romit Roy Choudhury, Tom Minka, You are facing the mona lisa: Spot localization using PHY layer information, in: *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services*, ACM, Seoul, South Korea, 2012, pp. 183–196.
- [7] Zheng Yang, Zimu Zhou, Yunhao Liu, From RSSI to CSI: Indoor localization via channel response, *ACM Comput. Surv.* (CSUR) 46 (2) (2013) 1–32.
- [8] Moustafa Abbas, Moustafa Elhamshary, Hamada Rizk, Marwan Torki, Moustafa Youssef, WiDeep: WiFi-based accurate and robust indoor localization system using deep learning, in: *Proceedings of the 2019 IEEE International Conference on Pervasive Computing and Communications*, IEEE, Sapporo, Japan, 2019, pp. 1–10.
- [9] Daqing Zhang, Dan Wu, Kai Niu, Xuanzhi Wang, Fusang Zhang, Jian Yao, Dajie Jiang, Fei Qin, Practical issues and challenges in CSI-based integrated sensing and communication, in: *Proceedings of the 2022 IEEE International Conference on Communications Workshops*, IEEE, Seoul, South Korea, 2022, pp. 836–841.
- [10] Hoonyong Lee, Changbum R Ahn, Nakjung Choi, Toseung Kim, Hyunsoo Lee, The effects of housing environments on the performance of activity-recognition systems using Wi-Fi channel state information: An exploratory study, *Sensors* 19 (5) (2019) 983.
- [11] Zhi Li, Xinpeng Rao, Toward long-term effective and robust device-free indoor localization via channel state information, *IEEE Internet Things J.* 9 (5) (2021) 3599–3611.
- [12] Chen Chen, Yan Chen, Yi Han, Hung-Quoc Lai, Feng Zhang, KJ Ray Liu, Achieving centimeter-accuracy indoor localization on WiFi platforms: A multi-antenna approach, *IEEE Internet Things J.* 4 (1) (2016) 122–134.
- [13] Emre Gönültaş, Sueda Taner, Howard Huang, Christoph Studer, Feature learning for neural-network-based positioning with Channel State Information, in: *Proceedings of the 55th Asilomar Conference on Signals, Systems, and Computers*, IEEE, Pacific Grove, CA, USA, 2021, pp. 156–161.
- [14] Parham Kazemi, Hanan Al-Tous, Christoph Studer, Olav Tirkkonen, User-side indoor localization using CSI fingerprinting, in: *Proceedings of the 23rd IEEE International Workshop on Signal Processing Advances in Wireless Communication*, SPAWC, IEEE, Oulu, Finland, 2022, pp. 1–5.
- [15] Miguel Matey-Sanz, Joaquín Torres-Sospedra, Adriano Moreira, Temporal stability on human activity recognition based on Wi-Fi CSI, in: *Proceedings of the 2023 13th International Conference on Indoor Positioning and Indoor Navigation*, IEEE, Nuremberg, Germany, 2023, pp. 1–6.
- [16] Xuyu Wang, Lingjun Gao, Shiwen Mao, Santosh Pandey, CSI-based fingerprinting for indoor localization: A deep learning approach, *IEEE Trans. Veh. Technol.* 66 (1) (2017) 763–776, <http://dx.doi.org/10.1109/TVT.2016.2545523>.
- [17] Jie Zhang, Zhanyong Tang, Meng Li, Dingyi Fang, Petteri Nurmi, Zheng Wang, CrossSense: Towards cross-site and large-scale WiFi sensing, in: *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, ACM, New York, NY, United States, 2018, pp. 305–320.
- [18] Rongjie Che, Honglong Chen, Channel state information based indoor fingerprinting localization, *Sensors* 23 (13) (2023) 5830.
- [19] Chao Chen, Zhaoli Wu, Xin Wang, Intelligent indoor localization algorithm based on channel state information, *Int. J. Pattern Recognit. Artif. Intell.* 37 (04) (2023) 2359009.
- [20] Khandaker Foysal Haque, Milin Zhang, Francesco Restuccia, SiMWiSense: Simultaneous multi-subject activity classification through Wi-Fi signals, in: *Proceedings of the 24th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks*, IEEE, Boston, MA, USA, 2023, pp. 46–55.
- [21] Fei Luo, Salabat Khan, Bin Jiang, Kaishun Wu, Vision transformers for human activity recognition using WiFi channel state information, *IEEE Internet Things J.* 11 (17) (2024) 28111–28122.
- [22] Fangzhan Shi, Wenda Li, Chong Tang, Yuan Fang, Paul V Brennan, Kevin Chetty, Decimeter-level indoor localization using WiFi round-trip phase and factor graph optimization, *IEEE J. Sel. Areas Commun.* 42 (1) (2023) 177–191.
- [23] Xueting Tang, Jun Yan, A CSI amplitude-phase information based graph construction for GNN localization, in: *Proceedings of the 2023 International Conference on Computer, Information and Telecommunication Systems*, IEEE, Genoa, Italy, 2023, pp. 01–07.
- [24] Siavash Zaravashan, Sadegh Arefzadeh, Sajjad Torabi, Efficient sub-carrier relationship extraction for human activity recognition via eegnet in wireless sensing, in: *Proceedings of the 13th International Conference on Computer and Knowledge Engineering*, IEEE, Mashhad, Iran, 2023, pp. 259–264.
- [25] Abid Hussain, Yueshan Chen, Arif Ullah, Sihai Zhang, WiSigPro: Transformer for elevating CSI-based human activity recognition through attention mechanisms, *Expert Syst. Appl.* 258 (2024) 124976.
- [26] Bin Li, Xin Jiang, Yirui Du, Yanzuo Yu, Ruonan Zhang, Wi-SSR: Wi-fi based lightweight high-resolution model for human activity recognition, *IEEE Sensors J.* 25 (4) (2025) 6556–6571.
- [27] Majid Ghosian Moghaddam, Ali Asghar Nazari Shirehjini, Shervin Shirmohammadi, A WiFi-based method for recognizing fine-grained multiple-subject human activities, *IEEE Trans. Instrum. Meas.* 72 (2023) 1–13.
- [28] Zhenguo Shi, Qingqing Cheng, J Andrew Zhang, Richard Yi Da Xu, Environment-robust WiFi-based human activity recognition using enhanced CSI and deep learning, *IEEE Internet Things J.* 9 (24) (2022) 24643–24654.
- [29] Ishtiaque Ahmed Showmik, Tahsina Farah Sanam, Hafiz Imtiaz, Human activity recognition from Wi-Fi CSI data using principal component-based wavelet CNN, *Digit. Signal Process.* 138 (2023) 104056.
- [30] Jianfei Yang, Xinyan Chen, Han Zou, Dazhuo Wang, Qianwen Xu, Lihua Xie, EfficientFi: Toward large-scale lightweight WiFi sensing via CSI compression, *IEEE Internet Things J.* 9 (15) (2022) 13086–13095.
- [31] Xiaobo Yang, Daosen Zhai, Ruonan Zhang, Haotong Cao, Sahil Garg, Mohammad Mehdi Hassan, Human-to-human interaction behaviors sensing based on complex-valued neural network using Wi-Fi channel state information, *Future Gener. Comput. Syst.* 148 (2023) 160–172.
- [32] Marco Cominelli, Francesco Gringoli, Francesco Restuccia, Exposing the CSI: A systematic investigation of CSI-based Wi-Fi sensing capabilities and limitations, in: *Proceedings of the 2023 IEEE International Conference on Pervasive Computing and Communications*, IEEE, Atlanta, USA, 2023, pp. 81–90.
- [33] Francesca Meneghello, Domenico Garlisi, Nicolò Dal Fabbro, Ilenia Tinnirello, Michele Rossi, SHARP: Environment and person independent activity recognition with commodity IEEE 802.11 access points, *IEEE Trans. Mob. Comput.* 22 (2022) 6160–6175.
- [34] Josyl Mariela Rocamora Reyes, Ivan Wang-Hei Ho, Man-Wai Mak, Wi-Fi channel state information (CSI) dataset for indoor positioning system (IPS) using hummingboard pro, 2023, https://github.com/iVANET-PolyU/CSI_IPS_2020-21/tree/main.
- [35] Steven M. Hernandez, Eyuphan Bulut, WiFi sensing on the edge: Signal processing techniques and challenges for real-world systems, *IEEE Commun. Surv. Tutorials* 25 (2022) 46–76, <http://dx.doi.org/10.1109/COMST.2022.3209144>.
- [36] Oussama Kerdjidi, Yassine Himeur, S S Sohail, Abbes Amira, Fodil Fadli, S Attala, Wathiq Mansoor, Abigail Copiaco, Amjad Gawanmeh, Sami Miniaoui, et al., Uncovering the potential of indoor localization: Role of deep and transfer learning, *IEEE Access* 12 (2024) 31.

- [37] Chen Chen, Gang Zhou, Youfang Lin, Cross-domain WiFi sensing with channel state information: A survey, *ACM Comput. Surv.* 55 (11) (2023) 231:1–231:37, <http://dx.doi.org/10.1145/3570325>.
- [38] Lili Zheng, Bin-Jie Hu, Jinguang Qiu, Manman Cui, A deep-learning-based self-calibration time-reversal fingerprinting localization approach on Wi-Fi platform, *IEEE Internet Things J.* 7 (8) (2020) 7072–7083.
- [39] Xi Chen, Hang Li, Chenyi Zhou, Xue Liu, Di Wu, Gregory Dudek, Fidora: Robust WiFi-based indoor localization via unsupervised domain adaptation, *IEEE Internet Things J.* 9 (12) (2022) 9872–9888.
- [40] Daniel Halperin, Wenjun Hu, Anmol Sheth, David Wetherall, Tool release: Gathering 802.11n traces with channel state information, *ACM SIGCOMM Comput. Commun. Rev.* 41 (1) (2011) 53–53.
- [41] Aryan Sharma, Junye Li, Deepak Mishra, Sanjay Jha, Aruna Seneviratne, Towards energy efficient wireless sensing by leveraging ambient Wi-Fi traffic, *Energies* 17 (2) (2024) 485.
- [42] Xinghang Zhan, Zhenhua Wu, Indoor positioning based on channel state information and deep learning domain adaptation, in: *Proceedings of the 4th International Conference on Computer Graphics, Artificial Intelligence, and Data Processing*, 13105, SPIE, Harbin, China, 2024, pp. 918–930.
- [43] Josyl Mariela B. Rocamora, Ivan Wang Hei Ho, Wan-Mai Mak, Alan Pak Tao Lau, Survey of CSI fingerprinting-based indoor positioning and mobility tracking systems, *IET Signal Process.* 14 (7) (2020) 407–419, <http://dx.doi.org/10.1049/iet-spr.2020.0028>.
- [44] Josyl Mariela Rocamora Reyes, Ivan Wang-Hei Ho, Man-Wai Mak, Wi-Fi CSI fingerprinting-based indoor positioning using deep learning and vector embedding for temporal stability, *Expert Syst. Appl.* 264 (2025) 125802.
- [45] Feng Zhang, Chen Chen, Beibei Wang, K.J. Ray Liu, WiSpeed: A statistical electromagnetic approach for device-free indoor speed estimation, *IEEE Internet Things J.* 5 (3) (2018) 2163–2177.
- [46] Diana Bri, Marta Fernandez-Diego, Miguel Garcia, Francisco Ramos, Jaime Lloret, How the weather impacts on the performance of an outdoor WLAN, *IEEE Commun. Lett.* 16 (8) (2012) 1184–1187.
- [47] Tanim M Taher, Matthew J Misurac, Joseph L LoCicero, Donald R Ucci, Microwave oven signal interference mitigation for Wi-Fi communication systems, in: *Proceedings of the 5th IEEE Consumer Communications and Networking Conference*, IEEE, Las Vegas, Nevada, 2008, pp. 67–68.
- [48] Espressif Systems, ESP32S2 product web page, 2023, <https://www.espressif.com/en/products/socs/esp32-s2>.
- [49] Steven M. Hernandez, Eyuphan Bulut, Lightweight and standalone IoT based WiFi sensing for active repositioning and mobility, in: *Proceedings of the 21st International Symposium on "a World of Wireless, Mobile and Multimedia Networks"*, IEEE, Cork, Ireland, 2020, pp. 156–161.
- [50] Emre Gönültaş, Eric Lei, Jack Langerman, Howard Huang, Christoph Studer, CSI-based multi-antenna and multi-point indoor positioning using probability fusion, *IEEE Trans. Wirel. Commun.* 21 (4) (2021) 2162–2176.
- [51] Andrea Brunello, Angelo Montanari, Nicola Saccomanno, A framework for indoor positioning including building topology, *IEEE Access* 10 (2022) 114959–114974.
- [52] Joaquín Torres-Sospedra, Raúl Montoliu, Adolfo Martínez Usó, et al., UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems, in: *Proceedings of the 2014 International Conference on Indoor Positioning and Indoor Navigation*, IEEE, Busan, South Korea, 2014, pp. 261–270.
- [53] Germán Martín Mendoza-Silva, Joaquín Torres-Sospedra, Joaquín Huerta, A meta-review of indoor positioning systems, *Sensors* 19 (20) (2019) 4507, <http://dx.doi.org/10.3390/s19204507>.
- [54] Shixiong Xia, Yi Liu, Guan Yuan, et al., Indoor fingerprint positioning based on wi-fi: An overview, *ISPR Int. J. Geo- Inf.* 6 (5) (2017) 135, <http://dx.doi.org/10.3390/ijgi6050135>.
- [55] Paramvir Bahl, Venkata N. Padmanabhan, RADAR: an in-building RF-based user location and tracking system, in: *Proceedings of the 19th Annual Joint Conference of the IEEE Computer and Communications Societies on Computer Communications*, IEEE Computer Society, Tel Aviv, Israel, 2000, pp. 775–784, <http://dx.doi.org/10.1109/INFCOM.2000.832252>.
- [56] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, Masanori Koyama, Optuna: A next-generation hyperparameter optimization framework, in: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, Anchorage, AK, USA, 2019, pp. 2623–2631, <http://dx.doi.org/10.1145/3292500.3330701>.
- [57] Iandra Galdino, Julio CH Soto, Egberto Caballero, Vinicius Ferreira, Taiana Coelho Ramos, Célio Albuquerque, Débora Muchaluat-Saade, eHealth CSI: A Wi-Fi CSI dataset of human activities, *IEEE Access* 11 (2023) 71003–71012.
- [58] Shuokang Huang, Kaihan Li, Di You, Yichong Chen, Arvin Lin, Siying Liu, Xiaohui Li, Julie A. McCann, WiMANS: A benchmark dataset for WiFi-based multi-user activity sensing, 2024, p. 39, <http://dx.doi.org/10.48550/ARXIV.2402.09430>, CoRR, arXiv:2402.09430 abs/2402.09430.
- [59] Guo Linlin, Wang Lei, Liu Jialin, Zhou Wei, Lu Bingxian, A novel benchmark on human activity recognition using WiFi signals, in: *Proceedings of the 2017 IEEE International Conference on E-Health Networking, Application and Services*, IEEE, Dalian, China, 2017, pp. 1–6.
- [60] Francesca Meneghello, Nicolò Dal Fabbro, Domenico Garlisi, Ilenia Tinnirello, Michele Rossi, A CSI dataset for wireless human sensing on 80 MHz Wi-Fi channels, *IEEE Commun. Mag.* 61 (9) (2023) 146–152, <http://dx.doi.org/10.1109/MCOM.005.2200720>.
- [61] Francesca Meneghello, Cheng Chen, Carlos Cordeiro, Francesco Restuccia, Toward integrated sensing and communications in IEEE 802.11bf Wi-Fi networks, *IEEE Commun. Mag.* 61 (7) (2023) 128–133, <http://dx.doi.org/10.1109/MCOM.001.2200806>.
- [62] Jingtao Guo, Ivan Wang-Hei Ho, Yun Hou, Zijian Li, FedPos: A federated transfer learning framework for CSI-based Wi-Fi indoor positioning, *IEEE Syst. J.* 17 (3) (2023) 4579–4590, <http://dx.doi.org/10.1109/JSYST.2022.3230425>.
- [63] Wei Liu, Zhiqiang Dun, D-Fi: Domain adversarial neural network based CSI fingerprint indoor localization, *J. Inf. Intell.* 1 (2) (2023) 104–114, <http://dx.doi.org/10.1016/j.jiixd.2023.04.002>.
- [64] Yuming Hu, Xiubin Fan, Zhimeng Yin, Feng Qian, Zhe Ji, Yuanchao Shu, Yeqiang Han, Qiang Xu, Jie Liu, Paramvir Bahl, The wisdom of 1,170 teams: Lessons and experiences from a large indoor localization competition, in: *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*, ACM, Madrid, Spain, 2023, pp. 38:1–38:15.
- [65] Celso M de Melo, Antonio Torralba, Leonidas Guibas, James DiCarlo, Rama Chellappa, Jessica Hodgins, Next-generation deep learning based on simulators and synthetic data, *Trends Cogn. Sci.* 26 (2) (2022) 174–187.