IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 1 October 2024, accepted 28 October 2024, date of publication 12 November 2024, date of current version 21 November 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3496561

SURVEY

A Review of Open Access WiFi Fingerprinting Datasets for Indoor Positioning

XU FENG, KHUONG AN NGUYEN, AND ZHIYUAN LUO
Computer Science Department, Royal Holloway University of London, TW20 0EX Surrey, U.K.
Corresponding author: Xu Feng (Xu.Feng.2023@live.rhul.ac.uk)

ABSTRACT WiFi fingerprinting is one of the most widely used techniques for indoor positioning systems. However, existing fingerprinting datasets came in different shapes and forms with varying levels of information without any standardised format. They were also dispersed across multiple platforms, making it challenging for new researchers to identify and access a suitable dataset to evaluate their own positioning systems. To address this challenge, this paper provides a comprehensive review of more than 50 publicly available WiFi fingerprinting datasets. We examine the most critical elements for fingerprinting, including the size and location of the testbed, the WiFi signal input, the number of locations, the temporal and spatial intervals of data collection, the positioning performance, and more. Surprisingly, it was observed that a large number of reference and access points, the use of 3D coordinates, denser sampling grid, and higher data collection frequencies do not always guarantee improved performance as often reported in the literature. The paper also outlines current challenges, and proposes guidelines for creating new WiFi fingerprint datasets.

INDEX TERMS Indoor positioning, WiFi fingerprinting, open access dataset.

I. INTRODUCTION

Positioning and navigation systems, like GPS, have become indispensable in most aspects of our life, from transportation to logistics operations [1], [2]. However, GPS faces severe challenges in indoor environments due to the inability of satellite signals to penetrate modern buildings. Additionally, the multipath effect, reflections, blockages, and absorption in real-world indoor scenarios can result in fluctuating and unstable GPS signal measurements, leading to unreliable indoor positioning performance [3], [4].

Thanks to the extensive infrastructure of WiFi Access Points (APs) in public spaces and the widespread use of WiFi-enabled smart devices (e.g., cell phones, tablets, and smartwatches), WiFi-based approaches were cost-effective for indoor positioning [5], [6], making WiFi fingerprinting one of the most popular approaches for indoor positioning research [3], [7].

WiFi fingerprinting is a positioning approach that estimates the user's location by employing positioning algorithms to match real-time WiFi signal measurements (fingerprints) to a pre-constructed fingerprint dataset. Thus, the positioning performance largely relies on the quality and granularity of the collected fingerprint dataset. Anomalies, outliers and missing values in the fingerprint dataset can significantly impair the accuracy of the final estimation. Additionally, the application of machine learning and deep learning to WiFi fingerprinting has recently become prominent, which further enhances the significance of data quantity, time and space diversity and heterogeneity as factors in determining positioning accuracy [4], [7], [8]. However, creating a high-quality WiFi fingerprint dataset is both labour-intensive and time-consuming.

To evaluate and validate the newly proposed WiFi fingerprinting methods and ensure their generalisation and transferability, researchers have chosen a cost-effective approach by utilising publicly available large-scale WiFi fingerprint datasets [9], [10], [11]. On the other hand, the research community is increasingly committed to sharing their collected WiFi fingerprint datasets, as shown in Figure 1. However, under these circumstances, there remains a lack of comprehensive taxonomy and consistent standards in the construction, formatting, description, and publication

The associate editor coordinating the review of this manuscript and approving it for publication was Abdel-Hamid Soliman.

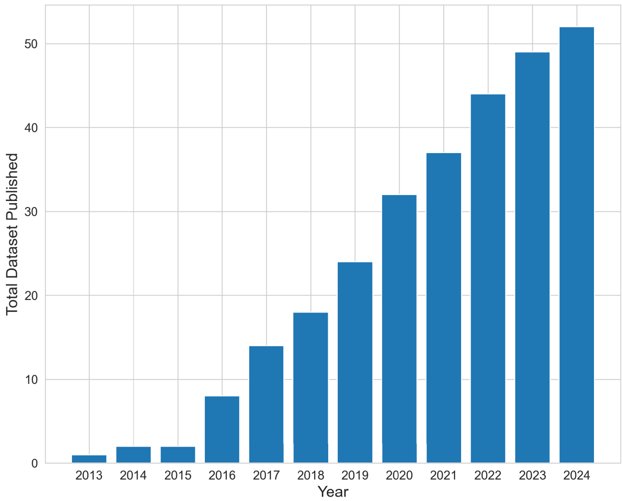


FIGURE 1. The total number of WiFi fingerprint datasets with open access published annually up to June 2024. Datasets with invalid access are not included.

of publicly available WiFi fingerprinting datasets, which presents several critical challenges:

- The fingerprinting datasets made publicly available by the research community were distributed across multiple platforms, making it challenging to access.
- Many published WiFi fingerprint datasets contain expired links or restricted access [12]. Common repositories like the IndoorLoc Platform [13] only have 7 public WiFi fingerprinting datasets, with the most recent published in 2017.
- Many existing datasets do not make clear of important aspects such as the human interference, the signal inputs, and the temporal changes in the WiFi signals.
- No uniform standards for the collection, formatting, publication, and organisation of public WiFi fingerprinting datasets.

As a result, the above issues have led to biased usage of publicly available datasets [14], [15], [16], potentially causing overfitting in the proposed WiFi fingerprinting algorithms.

To address these challenges, this paper conducts a comprehensive analysis and review of over 50 public WiFi fingerprint datasets with open access. It thoroughly investigates the most critical elements of a WiFi fingerprint dataset from a researcher's perspective, including the size and location of the testbed, 2D/3D indoor positioning type, WiFi signal input, access points (APs), receiver devices, the number of reference points (RPs) covered, the number of WiFi fingerprint data samples, data collection temporal and spatial interval, ground truth acquisition, and reported positioning performance. The dataset features are meticulously and extensively extracted and compared even when they are not explicitly provided. This paper also analyses current trends and challenges within each dataset feature and its impact on the performance of WiFi fingerprinting. By ensuring open access to all included WiFi fingerprint datasets, this paper aims to provide valuable insights and guidelines for the effective selection of the existing datasets and the efficient construction and sharing of new ones.

In summary, this paper makes the following contributions:

- This paper conducts a comprehensive analysis and comparison of over 50 publicly available WiFi fingerprint datasets, providing a broad overview of the current landscape in WiFi fingerprint datasets.
- The open accessibility of all included WiFi fingerprint datasets was meticulously validated. Only access links that are valid and up-to-date were retained.
- It identifies and investigates the most critical elements of WiFi fingerprinting in every dataset, even when they are not explicitly provided.
- It analyses how the trends and challenges in the existing WiFi fingerprinting datasets impact the system performance, thereby offering valuable insights and guidelines for the effective selection of existing datasets and the efficient construction and sharing of new ones.
- This paper points out that an increase in reference points and access points, the use of 3D positioning, larger RP intervals, and higher WiFi collection frequencies does not always result in enhanced system performance.
- This paper proposes standards for the collection, formatting, publication, and organisation of public WiFi fingerprint datasets.

The remainder of this paper is structured as follows. Section II outlines the review scope and methodology applied for the inclusion of public WiFi fingerprint datasets. Section III introduces WiFi fingerprinting, its main signal inputs in the literature. Section IV details the basic dataset structure and presents a comprehensive comparison of existing public WiFi fingerprint datasets. Section V lists limitations in current publicly available WiFi datasets, analysed the influence of different dataset features on the reported performance, and proposes standards and guidelines for public WiFi fingerprinting dataset publication. Finally, Section VI draws a conclusion to the whole work.

II. REVIEW SCOPE AND METHODOLOGY

This section introduces the boundaries and research focus of this review and details the systematic methodology utilised to search, select, and analyse open access WiFi fingerprint datasets, ensuring a thorough and comprehensive review of the topic.

A. REVIEW FOCUS

This paper strives to provide a comprehensive review and analysis of the current publicly accessible WiFi fingerprint datasets, aiming to offer valuable insights and practical guidelines for both selecting existing datasets and developing and sharing new ones effectively. Therefore, datasets specifically designed for WiFi fingerprinting and indoor positioning purposes constitute our primary research focus. Outdoor WiFi fingerprint datasets, due to the different signal propagation characteristics, are not included in the main body of this review. To catalyse the development of WiFi fingerprinting

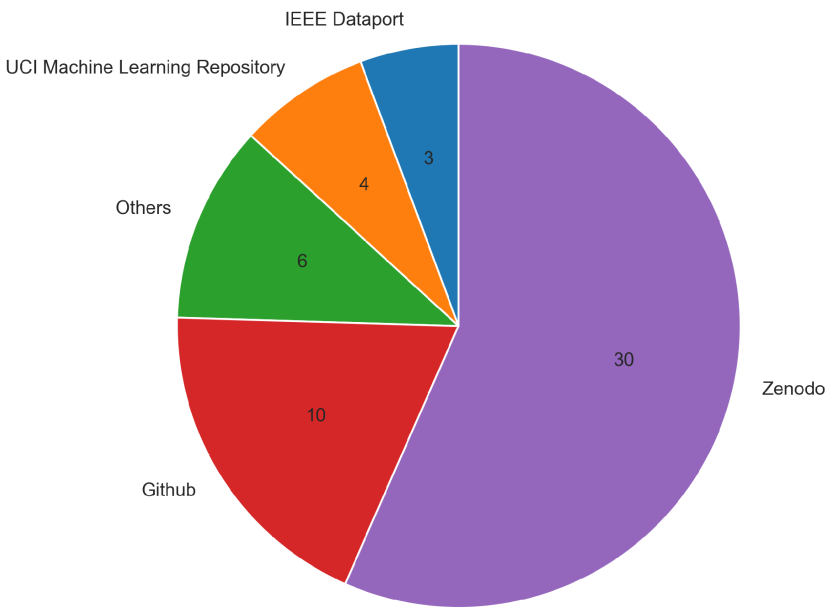


FIGURE 2. Open access WiFi fingerprint dataset published on different public dataset platforms. It is observed that most of the recent WiFi fingerprint datasets were published on Zenodo and Github.

techniques, the scope is restricted to the datasets that have valid and timely public available access.

B. METHODOLOGY

To ensure a comprehensive search and precisely identify the highly relevant publicly available WiFi fingerprint datasets, various combinations of the searching keywords were utilised. The keywords employed include "WiFi", "indoor", "localization", "localisation", "indoor positioning", "navigation", "fingerprint*", and especially, "dataset" and "database". "IndoorLoc" was also used to align with the naming conventions in WiFi fingerprint datasets [11], [14], [15], [17], [18], [19]. These keywords were searched in the title, keywords, and description sections of datasets on well-known public dataset publication platforms such as Data.gov, Kaggle, UCI Machine Learning Repository, IEEE dataport, Zenodo, Github, Google Dataset Search, data.world, figshare, Mendeley data, and Nist public data repository. The distribution of datasets included in this research, published across various public platforms, is shown as Figure 2, indicating that most of the recent WiFi fingerprint datasets were published on Zenodo and Github. The same methodology was applied to search for WiFi fingerprint dataset related publications on Google Scholar, Scopus, and Web of Science to ensure a more thorough investigation. In addition, research publications that include a comparison table covering a select number of public WiFi fingerprint datasets were also incorporated into the scope of the literature search [18], [20], [21], [22], [23].

C. SELECTION CRITERIA

In order to secure clarity and reproducibility of our research, we have established specific criteria for including public WiFi fingerprint datasets and their sources in our analysis:

- Firstly, the inclusion of existing datasets was initially restricted to those with valid and timely open access. The public accessibility of all datasets included in this research was manually meticulously verified, ensuring that they could be downloaded and unzipped by any researcher with no further requirement or membership subscription. For datasets published across diverse platforms, the duplicate access was examined, and only links that met our criteria were kept for this research.
- Secondly, the inclusion was limited to those with dataset feature description or corresponding research publications for reference. We were making every effort to ensure every dataset matched its corresponding publication, even if this was not specified on the dataset release page, the publication heavily predated the dataset release, or the dataset link had expired and been replaced.
- Additionally, datasets not specifically designed for indoor positioning purpose were included if they came with ground truth coordinates of the WiFi signal measurements.
- Datasets incorrectly labelled with "WiFi", containing no ground truth labels, or offering no dataset feature descriptions were all excluded.

Note that sub-datasets sharing the same access link and release page were regarded as the same entity in this research.

III. WiFi FINGERPRINTING AND SIGNAL INPUTS

This section offers an overview of WiFi fingerprinting and its most prominent signal inputs, including WiFi RSS (received

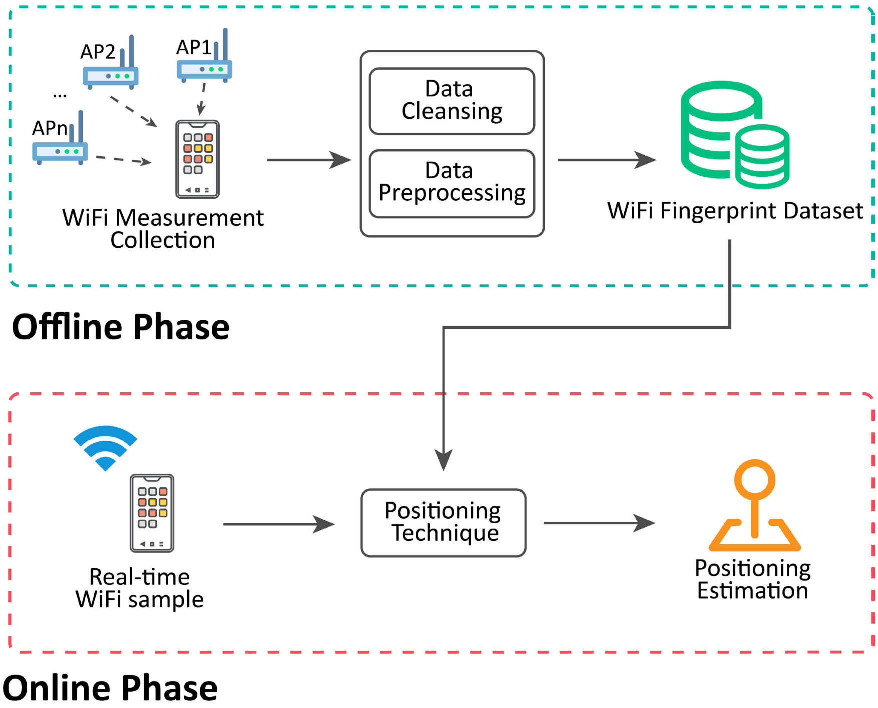


FIGURE 3. An overview of WiFi-based indoor fingerprinting.

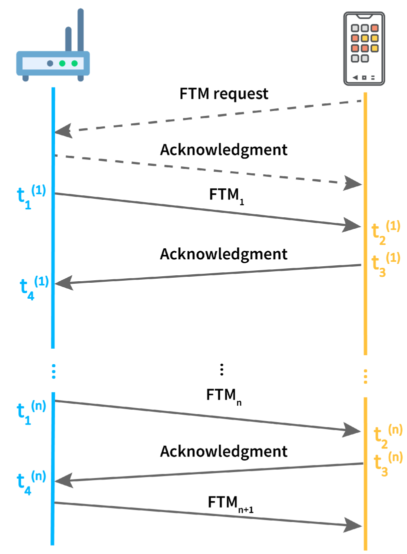


FIGURE 4. An overview of RTT protocol.

signal strength), WiFi RTT (round trip time), and CSI (channel state information).

A. WiFi FINGERPRINTING

WiFi fingerprinting is a positioning method that employs positioning algorithms to match real-time WiFi signal measurements (fingerprints) to a pre-constructed fingerprint dataset. Despite the pervasive penetration of WiFi signals, their propagation still struggles in complex indoor environments due to refraction, reflection, attenuation, blockage, absorption, and multipath interference. The sensitive nature of WiFi signal propagation in complicated indoor scenarios allows it to readily reflect even slight changes in the indoor interior, thereby creating distinguishing and unique fingerprints at different locations [4], [24]. By meticulously collecting unique WiFi fingerprints and the corresponding ground truth coordinates at different locations, a refined level WiFi fingerprint dataset is well established. The positioning algorithm estimates the user's precise location by comparing the real-time fingerprint reported by the user to the fingerprint dataset. As depicted in Figure 3, the fingerprinting method contains an offline phase and an online phases.

In the offline phase, an extensive radio map is constructed, incorporating distinct WiFi fingerprints and their corresponding ground-truth coordinates at all locations within the indoor environment. This dataset then undergoes meticulous preprocessing and data cleansing, which includes missing value imputation, duplicate and outlier detection, data scaling and partitioning [25], [26]. Subsequently, the processed WiFi fingerprint dataset is used to train a positioning model. In the online phase, when a user enters the tracking zone, a new WiFi sample is reported to the system. Following a similar preprocessing routine, the test sample is compared against the training samples in the offline WiFi fingerprint dataset. The positioning estimation of the user's current location is then generated by the positioning technique utilised.

B. WiFi SIGNAL INPUTS

There are currently three prominent types of WiFi signal inputs: RSS, RTT and CSI.

WiFi RSS, also known as received signal strength indicator (RSSI), is one of the most widely used measurements

TABLE 1. A Snapshot of the WiFi dataset proposed in [6], [40]. The value of -200 dBm in (a) and 100,000 millimetres (mm) in (b) indicates that the AP is not visible from the current reference point.

(**a**) WiFi RSS data samples

X	Y	AP1 RSS (dBm)	AP2 RSS (dBm)	...	AP13 RSS (dBm)	LOS APs
1	15	-200	-200	...	-73	12
1	16	-200	-200	...	-70	12
2	0	-200	-200	...	-71	None
2	1	-200	-200	...	-63	12
...	...	...	...	...	...	...
125	15	−74	−47	...	−200	2 3

(**b**) WiFi RTT data samples

X	Y	AP1 RTT (mm)	AP2 RTT (mm)	...	AP13 RTT (mm)	LOS APs
1	15	100,000	100,000	...	5,958	12
1	16	100,000	100,000	...	4,893	12
2	0	100,000	100,000	...	8,716	None
2	1	100,000	100,000	...	10,062	12
...	...	...	...	...	...	...
125	15	10,585	598	...	100,000	2 3

in traditional indoor positioning approaches [1], [7], [27], [28], [29]. As a passive positioning method, it requires the collection of MAC layer WiFi received signal strength and basic service set identifier (BSSID) at every location in the indoor scenario. The signal strength from each AP in the environment at a specific location forms the unique fingerprint. While fingerprinting was originally implemented with the WiFi RSS measures [5], [27], [30], [31], [32], it can also be seamlessly extended to include WiFi RTT and CSI [1], [2], [3], [4], [28], [33], [34], [35], [36].

RTT, which measures the time taken for a WiFi signal to travel from the transmitter to the receiver, directly calculates the distance between these points. As shown in Figure 4, RTT protocol starts with the transmission of a fine time measurement (FTM) request from the initiator (e.g., smartphone) to the responder (e.g., WiFi AP), specifying message count and intervals. Upon receiving the request, the WiFi AP transmits a series of FTM messages and awaits acknowledgment from the smartphone. The responder meticulously timestamps and calibrates each FTM request and acknowledgment receipt. Exchange of these temporal details allows both parties to calculate the round trip time $(t_4^{(1)} - t_1^{(1)})$, propagation time $[(t_4^{(1)} - t_1^{(1)} - t_3^{(1)} + t_2^{(1)})/2]$, and therefore the distance D between the smartphone and WiFi AP is calculated as

$$D = \frac{\frac{1}{n}\sum_{i=1}^{n}((t_4^{(i)} - t_1^{(i)}) - (t_3^{(i)} - t_2^{(i)}))}{2} \times c \tag{1}$$

where n is the total number of FTM round trips, $(t_4^{(i)} - t_1^{(i)})$ is the time it takes for the ith round trip, $(t_3^i - t_2^i)$ is the time delay that occurred within the smartphone, and c is the speed of light.

RTT measurements offer an alternative way to capture the subtleties of WiFi signal propagation. Similar to RSS fingerprints, different locations within the same testbed are characterised by their unique RTT fingerprints. Moreover, due to the speed of light at which the WiFi signals travel, even a slight delay in the propagation path could lead to noticeable changes in the RTT signal measurements especially in none-line-of-sight (NLOS) scenarios [5]. In comparison to RSS, RTT exhibits heightened sensitivity to interior changes, suggesting a more promising fingerprinting performance and positioning accuracy [37], [38], [39].

CSI utilises the propagation properties of WiFi signals to represent signal behaviour in indoor environments. It has two main types: Channel Impulse Response (CIR) and Channel Frequency Response (CFR). CIR provides the magnitude and phase information of WiFi signals in the time domain, describing how the channel alters an impulse signal due to multipath propagation, whereas CFR describes the characteristics of the WiFi channel across frequency subcarriers and can be extracted using an Orthogonal Frequency-Division Multiplexing (OFDM) system. Multipath effects in the complex NLOS indoor environment can be effectively characterised by CSI to enhance positioning performance [1], [3], [7], [28]. This method offers a higher level of granularity compared to traditional WiFi RSS measurements, providing a more robust solution for complicated indoor environments where signals are highly susceptible to interference and reflection.

IV. REVIEW OF THE EXISTING DATASETS

This section provides a detailed overview of the structure of WiFi fingerprint datasets. Next, a comprehensive review is conducted of more than 50 publicly available, state-of-the-art WiFi fingerprint datasets with guaranteed open access. The most important elements of WiFi fingerprint dataset are investigated from a researcher's standpoint, including size and location of the testbed, 2D/3D indoor positioning type, WiFi signal input, access point, receiver device, number of RP covered, number of WiFi fingerprint data samples, data collection temporal and spatial interval, ground truth acquisition, and reported positioning performance.

A. STRUCTURE OF WiFi FINGERPRINT DATASETS

To achieve the best indoor positioning performance and deliver the best indoor positioning system, WiFi fingerprint datasets should consist of several key components, including

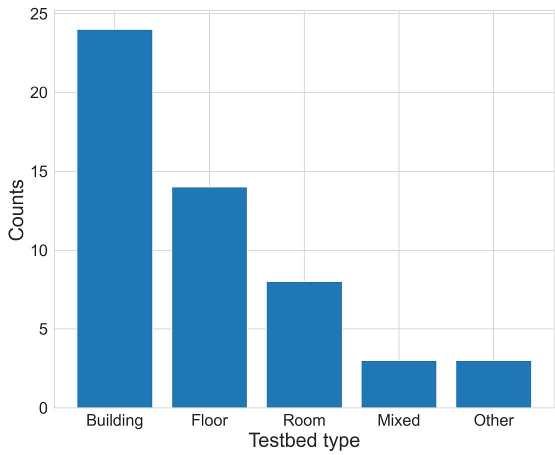


FIGURE 5. **Classification of different testbed types in the included datasets. 'Mixed' indicates that the dataset contains more than 1 testbed type. 'Other' includes building hall, hallway and corridor.**

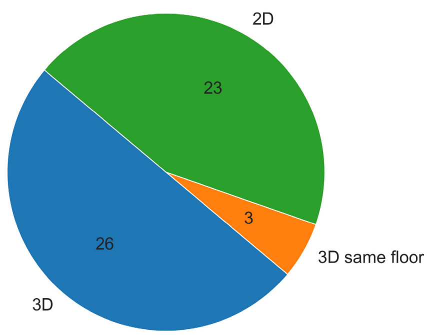


FIGURE 6. **Classification of 2D/3D positioning in the included datasets. '3D' means that the datasets provide floor information for 3D positioning. '3D same floor' indicates that the dataset includes XYZ coordinates for each data sample.**

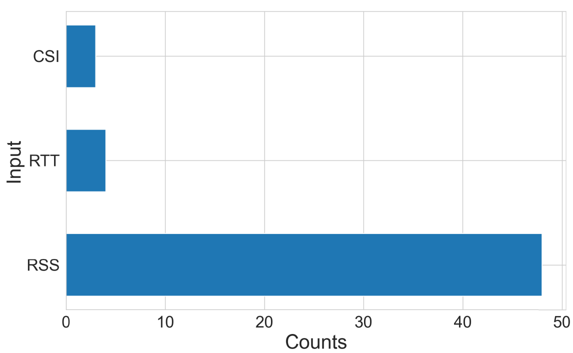


FIGURE 7. **Classification of WiFi signal input in the included datasets.**

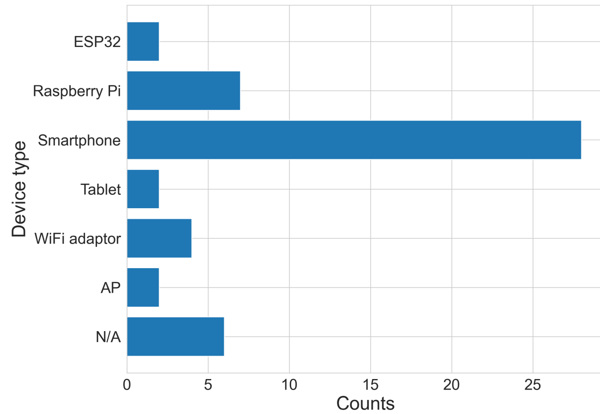


FIGURE 8. **Classification of different receiver devices used in the included datasets.**

primary WiFi signal measures from each AP, different WiFi signal inputs at the same location, multiple WiFi signal measurements at the same location, WiFi AP information like BSSID and line-of-sight (LOS) conditions, a fine grid of RPs, ground truth coordinates. The description of the dataset collection process is also of great significance, such as the description of the testbed type and size, the overall time period for the dataset creation, the devices used for collection, the grid size utilised for dividing RPs, and WiFi signal measurement sampling rate.

However, very few public WiFi fingerprint datasets include all of those components and follow such criteria. Starting with different WiFi signal measurements, aside from RSS, very few datasets include alternative signal types such as CSI and RTT. The comparatively newly released RTT signal measure only appears in four included datasets [40], [41] [42], [43], and CSI is only provided in three included WiFi fingerprint datasets [44], [45] [46]. Providing information on all the APs in the testbed, including those that are undetected at the current reference point (RP), is vital for dataset creation. To indicate those APs that are too far away or in a complete NLOS condition from this RP, default artificial values were employed such as +100dBm [47] and −200dBm [6] for WiFi RSS measurement, and −100,000 mm for RTT measurement, as shown in Table 1. Though the basic WiFi BSSID is seen in all the datasets, the LOS conditions of all APs included are only found in [40] and [48]. Next, the ground truth locations of the RPs where WiFi fingerprint is recorded are one of the most important elements for WiFi fingerprinting. Collected by manually aligning fixed grid [49], [50], robot with IMU, LIDAR, and RGB-D camera [51], [52], or total station [23], ground truth labels are essential for reliable indoor positioning performance evaluation. For 3D WiFi fingerprint dataset, either XYZ axis coordinates [41], [46] for single-floor datasets, or the corresponding FloorID were included in each WiFi sample as illustrated in the example above [15]. While information about the testbed is commonly seen in public datasets, ranging from shopping malls [53] and university buildings [54] to laboratories [55], the collection time period is not as frequently documented. It is also encouraged to provide detailed information of the devices used for collection, whether it was different smartphones [56], Raspberry Pi [43] or WiFi adaptor [57]. The grid size of RP and WiFi sampling rate during dataset collection were included in less than 20 public datasets, as shown in Table 2.

TABLE 2. Overview of existing publicly available WiFi fingerprint datasets.

ID	Year	Testbed size	Location	2D/3D (Floors)*	Signal	# of APs	# of receivers	# of RPs	# of samples	Collection interval	Ground Truth	Performance	Ref.
1	2013	124 m × 57 m, 145 m × 88 m	University building	3D (4&3)	RSS	309, 354	1 tablet	1,476, 584	1,966, 760	1 m	Fixed grid	9.59 m, 14.37 m, 1NN [20]	[74] [73]
2	2014	108,703 m^2	University building	3D (5&4 &4)	RSS	520	25 phones and tablets	933	21,049	N/A	UTM from WGS84	7.9 m, 1NN	[47] [15]
3	2016	185.12 m^2	Office, corridors and hall	2D	RSS	132	1 phone	325	21,795	10 Hz, 60 cm	Fixed grid	4.0 m, KNN [75]	[65] [49]
4	2016	20 m × 26 m	Building floor	2D	RSS	1	5 APs	Whole trajectory	16,576	5 Hz	Robot odometry	82%, N/D	[69] [76]
5	2016	200 m^2	Laboratory	2D	RSS	93	2 phones	34	1,140	N/A	N/A	4.73 m, 1NN	[77] [78]
6	2016	4 buildings	University building	3D (1&4 &6&3)	RSS	816	6 phones	2,007	148,191	0.17-0.25 Hz	Markers, landmarks	5.76 m, N/D	[56] [16]
7	2016	2,106.16 m^2	Building	3D (3)	RSS	31	1 phone	N/A	1,571	N/A	N/A	N/A	[79] [80]
8	2016	120 m^2	Corridor	2D	RSS	168	1 phone	66	1,629	N/A	N/A	N/A	[61]
9	2017	308.4 m^2	Building floor	3D (2)	RSS	448	1 phone	96	63,504	N/A	N/A	3.02 m, 1NN [20]	[81] [82]
10	2017	22,570 m^2	University building	3D (5)	RSS	992	21 phones and tablets	4,537	4,648	N/A	Fingerprinting, manual selection	6.36 m, 1NN [20]	[83] [84]
11	2017	3 buildings	University building	3D (1&6 &4)	RSS	616	9 phones	2,697	170,774	0.17 Hz	Markers, landmarks	2.96 m, N/D	[85] [86]
12	2017	4 office rooms	Office room	2D	RSS	7	1 phone	4 rooms	2,000	N/A	Markers	95.16%, FP-SOGSA	[87] [88]
13	2018	85 m × 125 m	University building	3D (3)	RSS	489	N/A	446	1,428	5 m & 3 m	Fixed grid	6.92 m, 1NN [20]	[89]

* In the 2D/3D (Floors) column, '2D' denotes 2D positioning, while '3D' indicates same-floor 3D positioning. For datasets spanning multiple buildings and floors, additional details are provided after the '3D' label. For example, '3D (3)' signifies a dataset covering one building with three floors, while '3D (5&4&4)' indicates a dataset spanning three buildings with five, four, and four floors, respectively.

Therefore, following the suggested structure guideline would greatly enhance the development and sharing of the public WiFi fingerprint dataset.

B. OVERVIEW OF EXISTING PUBLIC DATASETS

To provide comprehensive in-depth overview of the existing state-of-the-art public datasets, thorough analysis and comparison of important elements in constructing WiFi fingerprint datasets are covered. The elements include size and location of the testbed, 2D/3D indoor positioning type, WiFi signal input, access point, receiver device, number of RP covered, number of WiFi fingerprint data samples, data collection temporal and spatial interval, ground truth acquisition, and reported positioning performance. Following the research scope and methodology outlined in Section II, 52 publicly available datasets with guaranteed open access are included. Note that the datasets are listed according to their publication links, and subsets within the same release are regarded as the same dataset.

The comprehensive overview and comparison of the included WiFi fingerprint datasets are shown in Table 2. The corresponding links, file size and further details of the datasets are listed in Table 3. In the comparison Table 2, 'N/A' indicates that the relevant element was not specified or could not be found in the dataset description page or related research paper, 'N/D' in column 'Performance' means the algorithms used to provide reported performance was not described. The RSS signal measures in the included WiFi datasets are all given in dBm values. Additionally, the counts of reference points (# of RPs) and data samples (# of samples) include all training, testing, validation, and evaluation subsets. In the 2D/3D (Floors) column, '2D' denotes 2D positioning, while '3D' without any additional details indicates same-floor 3D positioning. For datasets spanning multiple buildings and floors, additional details are provided after the '3D' label. For instance, '3D (3)' signifies a dataset covering one building with three floors, while '3D (5&4&4)' indicates a dataset spanning three buildings with five, four, and four floors, respectively. All datasets were manually downloaded, and their open access links were tested as shown in Table 3, ensuring researchers can reliably use these links to access the public WiFi fingerprint datasets. Details related to previous expiring links could be found in the Notes column in Table 3.

TABLE 2. ***(Continued.)*** **Overview of existing publicly available WiFi fingerprint datasets.**

14	2018	30,000 m^2	University building	3D	RSS	630	4 phones	900+	up to 2,667 per scenario	1-0.33 Hz	N/A	2.14 m, spherical error 95% [90]	[91] [92]
15	2018	306 m^2	Building hall	3D (2)	RSS	515	2 phones	969	5,157	60 cm	Fixed grid	0.62 m, LSTM	[93] [17]
16	2018	9,000 m^2	Shopping mall	3D (3)	RSS	845	1 phone	200	161,596	0.17 Hz	LIDAR SLAM, markers	0.7 m, N/D	[62] [63]
17	2019	336 m^2	Building floor	2D	RSS	6	1 phone	365	36,500	N/A	Robot with LIDAR, RGBD camera	0.66 m, SRL-KNN [51]	[68] [51]
18	2019	1,000 m^2	Laboratory	2D	RSS	11	1 RPi	189	5,783	1 m	Fixed grid	2.67 m, 1NN [20]	[94], [95] [96], [97]
19	2019	44,000 m^2	University building	3D (16)	RSS	589	N/A	1,840	9,494	N/A	N/A	7.50 m, CNNLoc	[98] [11]
20	2019	2,646 m^2	Building floor	3D (3)	RSS	172	3 tablets and 1 phone	1000	25364	2 m	Fixed grid	1.24 m, 1NN	[99] [14]
21	2019	6,000 m^2	Building	3D (3)	RSS	709	1 phone	92 eval	211,183	0.25 Hz	Markers	1.7 m, N/D	[100] [101]
22	2019	24,000 m^2	University building	3D (2)	RSS	460	2 phones	489	7,565	3 m	Fixed grid	4.51 m, KNN	[102] [103]
23	2020	25,000+ m^2, 40,000+ m^2	2 Malls	3D (3&2)	RSS	885, 1,291	5 phones	69, 113 (shop level)	22,707 mall 1, 12,539 mall 2	N/A	N/A	40 m in the 90th percentile, FreeLoc	[53] [104]
24	2020	135 m × 62 m, 88 m × 137 m	University building	3D (4&3)	RSS	652, 801	N/A	3,116, 2,787	10,385, 9,291	N/A	N/A	1.94 m, 2.69 m, 1NN [20]	[105] [106]

* In the 2D/3D (Floors) column, '2D' denotes 2D positioning, while '3D' indicates same-floor 3D positioning. For datasets spanning multiple buildings and floors, additional details are provided after the '3D' label. For example, '3D (3)' signifies a dataset covering one building with three floors, while '3D (5&4&4)' indicates a dataset spanning three buildings with five, four, and four floors, respectively.

The detailed investigation and analysis of the key elements of the existing datasets are presented as follows:

1) TESTBED SIZE AND LOCATION

It is observed in Table 2 that when selecting the ideal testbed for WiFi fingerprinting, 24 out of 52 existing datasets chose entire buildings with 17 being university buildings, 2 being shopping malls, and one being a museum, as shown in Figure 5. Buildings, with their complex and complicated interiors, present a challenging environment for implementing WiFi-based indoor positioning systems. Factors such as walls, stairs and furniture of diverse materials increase the severity of propagation issues indoor environment, including refraction, reflection, attenuation, blockage, absorption, and multipath interference. However, these factors also contribute to the construction of unique and distinguishing fingerprint at different locations. The large-scale building testbed, consisting of multiple space types (e.g., rooms, corridors, halls, and basements), is also suitable for challenging long trajectory recordings [23], [41], [43], [52], [57], [58], and [59]. Additionally, multi-floor buildings naturally provide the conditions necessary for evaluating both 2D indoor positioning and floor prediction performance. The reason why other public spaces like train stations were less common due to the need for extra authorisation and the constant presence of large crowds.

Although sufficient for evaluating the generalization and transferability of indoor positioning systems, constructing fingerprint data for such large-scale real-world scenarios requires significant human effort. For example, the offline data construction phase took 5 months for the 5,432 m^2 building in [54], compared to just 3 days for a 92 m × 15 m building floor in [40]. Consequently, 15 of the included datasets were collected in building floors. Smaller testbeds were also found, such as office rooms [60], apartments [40], and single corridors [46], [61].

2) 2D/3D POSITIONING

Despite the importance of accuracy in 2D user positioning, researchers have also emphasized 3D positioning in the literature. We observe that 29 of the included datasets offer 3D ground truth labels, as shown in Figure 6, including 26 datasets that offered floor information as part of the 3D data. The 3D location labels vary from simple XYZ axis coordinates [41], [58] to 2D coordinates with a floor identifier [23], [57]. In Table 2, 3D(3) indicates

TABLE 2. ***(Continued.)*** **Overview of existing publicly available WiFi fingerprint datasets.**

25	2020	100 m × 18 m	Building floor	2D	RSS	157	1 tablet	230	1,717	N/A	N/A	4.95 m, 1NN [20]	[107]
26	2020	50 m × 20 m	Laboratory	2D	RSS	8	N/A	1,071	11,710	1 m	Fixed grid	3.24 m, 1NN [20]	[55] [72]
27	2020	33, 31, 79 m^2	Meeting room and laboratory	2D	RSS	3	1 RPi	59, 22, 56	300	0.5 m, N/A, N/A	Fixed grid	1.83 m, 1.41 m, 1.39 m, KNN	[108] [109]
28	2020	9,564 m^2	University building	3D (3)	RSS	436	1 WiFi adaptor	Whole trajectory	25,790	0.2 Hz	Total station	3.45 m, KNN	[57] [23]
29	2020	3,100 m^2	University building	3D (5)	RSS	470	1 phone	82 eval	172,576	0.25 Hz	Markers	0.86 m, N/D	[110] [111]
30	2020	50 m × 30 m	Office floor	3D	RTT	12	1 WiFi adaptor	Whole trajectory	29,527	N/A	LIDAR	4 m, 98% of the time, ANN	[41] [58]
31	2021	5,432 m^2, 4,184 m^2	University building	3D (6&3)	RSS	613, 775	1 phone	10,633, 9,291	10,683 9,447	N/A	N/A	6.55 m, 9.07 m, 1NN [20]	[54]
32	2021	4 m × 5 m, 4.95 m × 9.45 m	Office room	2D	RSS	4, 4	1 WiFi adaptor	54, 28	12,960 6,720	4 Hz, 50 cm, 45 cm	Markers	1.39 m, adaptive FKF	[112] [113]
33	2021	46.275 m × 37.27 m	Building floor	2D	RSS	220	1 phone	140	3,385	50 Hz	N/A	1.23m, KNN	[64] [114]
34	2021	Ground floor and the basement	University building	3D (2)	RSS	189	5 phones	39 eval	253,558	100 Hz	N/A	4.4 m, 75th percentile, N/D	[115] [116]

* In the 2D/3D (Floors) column, '2D' denotes 2D positioning, while '3D' indicates same-floor 3D positioning. For datasets spanning multiple buildings and floors, additional details are provided after the '3D' label. For example, '3D (3)' signifies a dataset covering one building with three floors, while '3D (5&4&4)' indicates a dataset spanning three buildings with five, four, and four floors, respectively.

that the dataset supports 3D positioning and it contains 3 building floors in total [62], [63]. The challenge with using XYZ coordinates is that the recording of the height of the receiver device requires additional effort in dataset construction. When data is collected using a fixed tripod, the height label remains constant, resulting in a 2D dataset. However, when the user moves naturally with varying receiver heights, the tracking and labelling of the precise ground truth becomes increasingly complex. For this reason, 25 out of the 29 3D datasets primarily provide floor identifiers.

3) SIGNAL INPUT

The presence of RSS in 49 WiFi datasets in Table 2 clearly illustrates that RSS is one of the most widely used measurements in traditional WiFi-based indoor positioning approaches [1], [7], [27], [28], [29]. Over the decade span of the included datasets, RSS has consistently been a focal point in the research area due to its ease of access. In contrast, the newer RTT measurement, available only on a limited number of commercial WiFi routers [24], appears in only four datasets [40], [41], [42], [43]. For example, due to the limited number of RTT-enabled APs in the testbed, datasets in [40] only include up to 13 APs in a building floor testbed. In addition, only 3 of the included datasets provide WiFi CSI [44], [45], [46] due to the challenging in acquiring CSI, as discussed in Section III. The severe shortage of public WiFi datasets that include RTT and CSI highlights the urgent need for their future development and publication.. The overall classification of different WiFi signal inputs is shown in Figure 7.

4) APs AND RECEIVER DEVICES

The number of APs and receiver devices (i.e., # of APs, receivers in Table 2) illustrates the heterogeneity and comprehensiveness of the datasets. In most of the building and floor testbeds documented in the literature, hundreds of APs were detected and recorded with their unique BSSID, including pre-installed WiFi routers, printers, and hotspots. However, this often means that these datasets do not provide detailed information about the specific brand and type of each AP included. For those provided AP information, please refer to the Notes column in Table 3. Another challenge associated with a large number of APs is the need for additional preprocessing and normalization methods when using these datasets. Regarding the receiver devices used, as shown in Figure 8, 30 out of the 52 included datasets utilised mobile devices such as smartphones and tablets, 7 used Raspberry Pi (RPi), 4 used WiFi adaptors, and the remainder used other devices like ESP32 [46], [60]. Brands of smart devices found in the datasets are Xiaomi, BQ, Huawei, Huawei, LG, Celkon, Samsung, HTC, Sony, Nexus, Orange, OnePlus, Asus, Google [15], [40], [53], [64]. Due to the large number and variety of smart devices employed, detailed information on each individual device is beyond the scope of this comparison.

TABLE 2. ***(Continued.)*** **Overview of existing publicly available WiFi fingerprint datasets.**

35	2021	108 m × 106 m, 92 m × 32 m, 92 m × 34 m, 90 m × 42 m, 35 m × 15 m, and 80 m × 92 m	Exhibition hall and 5 buildings	2D	RSS	N/A	5 phones	N/A	N/A	2 m hall, 1 m building	Fixed grid	N/A	[117] [22]
36	2021	Hundreds of buildings	Building	3D	RSS	N/A	N/A	N/A	N/A	N/A	N/A	5.8 m, KNN	[118]
37	2022	16.71 m × 10.76 m	Office room	2D	RSS	5	1 ESP32	173	7,875	20 Hz, 1 m	Fixed grid	2.60 m, KNN	[60] [50]
38	2022	18.5 m × 107.5 m	Building floor	3D (1)	RSS	2,711	6, 1 RPi	49	7,446,538	N/A	N/A	6 m, N/D	[119] [120]
39	2022	46.2 m^2	Laboratory	2D	CSI	2	1 AP	8	16,000	100 Hz	N/A	90%, CNN	[44] [121]
40	2022	92 m × 15 m, 4.5 m × 5.5 m, 7.7 m × 9.4 m	Building floor, office room, apartment	2D	RSS, RTT, LOS	13, 3, 4	1 phone	642, 37, 110	77,040, 4,440, 13,200	0.6 m, 0.455 m, 0.48 m	Fixed grid, markers	0.781 m, 0.394 m, 0.562 m, KNN	[40] [6], [24], [33], [36]
41	2022	35 m × 17.4 m [122]	Building floor	2D	RSS	27	N/A	250	18,750	1.33 Hz	Fixed grid	1.76 m, WKNN	[123] [124]
42	2022	8,000 m^2	University building	3D (3&1 &1)	RSS	105	9 phones	1,802	23,925	1.2 m, 1.2 m, 0.5 m; 1 Hz	Fixed grid	2.3 m, KNN	[125] [18]
43	2022	668 m^2 (4 buildings)	University building	3D (2&2 &1&1)	RSS	169	2 WiFi adaptors	417	109,800	1 m	Fixed grid	2.9 m, 75th Percentile, KNN	[66] [21]
44	2022	3,608 m^2	Building floor	2D	CSI, RSS	6	1 RPi	360	635,000	N/A	N/A	1.38 m, 1D CNN	[45] [19]

* In the 2D/3D (Floors) column, '2D' denotes 2D positioning, while '3D' indicates same-floor 3D positioning. For datasets spanning multiple buildings and floors, additional details are provided after the '3D' label. For example, '3D (3)' signifies a dataset covering one building with three floors, while '3D (5&4&4)' indicates a dataset spanning three buildings with five, four, and four floors, respectively.

5) NUMBER OF RPs AND DATA SAMPLES, AND DATASET COLLECTION INTERVAL

The number of RPs and samples (i.e., # of RPs, and # of samples in Table 2) demonstrates the general scale of the datasets, while the data collection interval illustrates the methodology and organisation of real-world dataset collection. An RP is a location where the WiFi fingerprint data samples and ground truth labels are recorded. These could be scattered points within the testbed or a series of WiFi fingerprints recorded along an entire walking trajectory, as seen in 6 out of the 52 included datasets. The density of RPs in the testbed and the number of data samples at each RP are indicated by the temporal and spacial intervals in the dataset collection. They indicate the spatial and temporal granularity of the fingerprint dataset from a construction perspective. RPs that are too close together may result in highly similar fingerprints and thus less accurate positioning performance [48]. The frequency of WiFi fingerprint recording relies on the number of APs in the background and the need for alignment with other sensors [59], [65]. Multiple data samples at each RP ensure comprehensive coverage of all APs in the testbed and capture the fluctuation of WiFi signals over short periods [24]. Therefore, statistical features of short-term WiFi signal measurements can be utilised by indoor positioning systems and LOS detection algorithms [21], [66] [36]. For model training and evaluation, RPs and data samples are divided into training, testing, evaluation or validation subsets that don't overlap [45], [66], especially for IPIN Indoor Localization Competitions [67]. To enhance clarity and provide a general understanding of the datasets, Table 2 includes the number of RPs and data samples across all non-overlapping training, testing, validation, and evaluation subsets.

6) GROUND TRUTH ACQUISITION

As one of the most important parts of WiFi fingerprint dataset construction, ground truth label acquisition requires significant attention, time and labour, especially for manually collected datasets. It is observed in Table 2, 23 of the included datasets indicated that ground truth labels were obtained by manually measuring and recording coordinates. Methods employed include using fixed tiles or grids on the testbed floor, aligning with landmarks, and utilising markers such as post-it notes. While manual collection requires the least financial investment in ground truth collection devices, it incurs substantial costs in terms of time and labour. In contrast, 4 of the included datasets utilised robots for their location ground truth acquisition. Specifically, a 3-wheel

TABLE 2. ***(Continued.)*** **Overview of existing publicly available WiFi fingerprint datasets.**

45	2023	32.8 m × 22.6 m	Building floor	2D	RSS	310	4 phones and 1 tablet	1800	8,899	60 cm	Laser measure, markers	3.34 m, 1NN	[126] [127]
46	2023	20 m × 50 m	Building floor	2D	RSS	9	4 RPis	21	10,575	0.62 Hz	Video camera	2.19 m, KNN	[70] [71]
47	2023	15 m × 14.5 m, 35 m × 6 m, 18 m × 5.5 m	Lecture theatre, corridor, office room	2D	RSS, RTT, LOS	5, 4, 5	1 phone	120, 114, 108	7,200, 6,840, 6,480	0.6 m	Fixed grid, markers	0.612 m, 0.729 m, 0.612 m, RF	[42] [39], [48]
48	2023	Building, and surroundings	University building	3D (3)	RSS	414	8 phones	92 eval	468,274	100 Hz	N/A	30.1 m, 75th percentile, N/D	[67] [116]
49	2023	Ground floor, mezzanine and the basement	Museum	3D (3)	RSS	321	2 phones	N/A	95,337	50 Hz	N/A	N/A	[67]
50	2024	18 m × 18 m, 20 m × 16 m, 10 m × 18 m, 4 m × 40 m	Building floor	2D	RSS	18	1 RPi	Whole trajectory	100,000	100 Hz	Robot with LIDAR, RGBD camera	0.82 m, BiLSTM	[59] [52]
51	2024	18 m × 18 m, 20 m × 16 m, 10 m × 18 m	Building floor	2D	RSS, RTT	8	1 RPi	Whole trajectory	1,000,000	100 Hz	Robot with LIDAR, RGBD camera	0.70 m, BiLSTM	[43] [52]
52	2024	52 m^2	Hallway	3D	CSI	1	1 ESP32	Whole trajectory	138,879	100 Hz	ORB-SLAM3	0.197 m, EfficientNetV2	[46] [128]

* In the 2D/3D (Floors) column, '2D' denotes 2D positioning, while '3D' indicates same-floor 3D positioning. For datasets spanning multiple buildings and floors, additional details are provided after the '3D' label. For example, '3D (3)' signifies a dataset covering one building with three floors, while '3D (5&4&4)' indicates a dataset spanning three buildings with five, four, and four floors, respectively.

robot equipped with IMU, LIDAR, sonar sensors and an RGB-D camera was used in [68], a Turtlebot3 with LIDAR and an RGBD camera was employed in [43] and [59], and robot odometry was utilised in [69]. Other methods reported in the literature include using the bMS3D mobile mapping system with 6 DOF LIDAR SLAM [62], LIDAR [41], ORB-SLAM3 [46], registering reference tags via video camera [70] and using Leica TS06 Plus total station [57].

7) REPORTED POSITIONING PERFORMANCE

To provide a general understanding of the proposed WiFi fingerprint datasets and establish baseline evaluation performance for user reference, machine learning models are commonly employed in the literature to generate positioning estimates. Among the included WiFi fingerprint datasets, several popular positioning models are predominantly used as positioning algorithms. Note that some of datasets were proposed as a supplementary material for novel positioning systems, thus the corresponding reported positioning performance is extracted based on them.

The k-Nearest Neighbours (KNN) algorithm is a simple yet widely used machine learning model for evaluating WiFi fingerprint datasets, appearing in 26 out of the 52 included datasets. KNN identifies the k closest training examples in the feature space and bases its predictions on the majority label or the average coordinates of these neighbours. The office room dataset constructed by both ESP32 WiFi transmitters and receivers achieved an accuracy of 2.60 m using KNN [50], [60]. KNN was also applied to a vehicle indoor positioning dataset, combing WiFi RSS, inertial measurement unit (IMU), and odometry data, demonstrating an accuracy of 2.19 m [70], [71]. Baseline positioning errors of 0.781 m, 0.394 m, and 0.562 m were achieved by the KNN model in a university building floor, an office room, and in an apartment, respectively [6], [40]. A soft range limited KNN (SRL-KNN) which incorporates a range factor related to the physical distance between the user's previous position and the reference location in the dataset, demonstrated a positioning accuracy of 0.66 m [51], [68]. 1NN, a more straightforward version of the KNN algorithm, is also frequently used in the literature. In [20], a comprehensive comparison of the baseline performance of several public WiFi fingerprint datasets was performed. The use of 1NN as a positioning model was validated across various indoor scenarios, ranging from a 2D 50 m x 20 m laboratory testbed [55], [72] to 3D datasets spanning more than seven university building floors [73]. The baseline positioning performance of the UjiIndoorLoc dataset, collected in a 108,703 square meters testbed with 25 smartphones and tablets, was also reported using 1NN with an accuracy of 7.9 meters [15], [47].

Neural network models, including artificial neural network (ANN), convolutional neural network (CNN) and Long Short-Term Memory (LSTM), have been adopted for performance evaluation of public WiFi fingerprint datasets. ANNs are initial forms of deep neural network (DNN), consisting of few layers of interconnected nodes or neurons.

TABLE 3. Links and notes of existing publicly available WiFi fingerprint dataset.

ID	Links	File size	Notes	Ref.
1	http://www.cs.tut.fi/tlt/pos/Software.htm	845 KB	TUT1, TUT2 datasets. The 4-floor building was remeasured again and labeled as BUILDING1_NEW in the dataset. Details were found in [20].	[74] [73]
2	https://doi.org/10.24432/C5MS59	42.73 MB	The most popular dataset in WiFi IPS. Validation (or testing) samples 4 months after Training. 508h duration in collecting training set from timestamp.	[47] [15]
3	https://doi.org/10.24432/C5DW43	13.72 MB	WiFi, Magnetic field, accelerometers, orientation. Phone: held chest level, screen up.	[65] [49]
4	https://github.com/herolab-uga/indoor-rssi-mobile-robot	2.81 MB	The dataset was designed to enhance the teleoperation of Unmanned Ground Vehicles (UGVs). Five small USB wireless adapters with detachable external antennas were attached to the UGV robot. Contains LOS and NLOS conditions. 82% in providing useful network connectivity. information to the operator. For UGV. Testbed empty.	[69] [76]
5	http://indoorloc.uji.es/	495 KB	Created to be WiFi+Magnetic. The magnetic data not available. It is called 'Databases collected at GEOTEC Lab' on the website.	[77] [78]
6	https://doi.org/10.5281/zenodo.2791530	482 MB	IPIN2016. Needs preprocessing for usage. WiFi, magnetic, IMU, pressure/sound/illumination, GNSS, 3D orientation. Number of APs and samples were counted by the authors. Phone: stable in front of his face or chest, keeping the arm relaxed downwards with the phone low at his hand.	[56] [16]
7	https://doi.org/10.24432/C55K6K	2.40 MB	Magnetometer, WiFi, Bluetooth RSSI. The dataset URL in the original paper no longer exists, please use the new link.	[79] [80]
8	http://indoorloc.uji.es/ipin2016tutorial/	1.1 MB	For IPIN 2016 Tutorial 2.	[61]
9	https://doi.org/10.5281/zenodo.3748719	251 MB	LIB dataset. Data were collected among the bookshelves from two floors (3rd and 5th) of a university building. Updated and extended in 2020, now has 103,584 data samples collected in 25 months.	[81] [82]
10	https://doi.org/10.5281/zenodo.1001662	17.9 MB	TUT3, TUT4 datasets. Details were found in [84] (especially floor number), [83] and [20]. Some details may be found different in [20].	[83] [84]
11	https://doi.org/10.5281/zenodo.2823924	572 MB	IPIN2017. Needs preprocessing for usage. WiFi, BLE, magnetic, IMU, pressure/sound/illumination, GNSS. Number of APs and samples were counted by the authors. Phone: stable in front of his face or chest, keeping the arm relaxed downwards with the phone low at his hand, phoning.	[85] [86]
12	https://doi.org/10.24432/C51880	60.5 KB	Classification dataset. Positioning model: Particle Swarm Optimization & Gravitational Search Algorithm (FPSOGSA).	[87] [88]

A comparative study was proposed in [58] to investigate the range estimation performance of ANN on a WiFi RTT dataset collected in a office floor [41]. The range estimation achieved an accuracy of 4 m, 98 % of the time. CNNs are specialized neural networks that employ convolution operations in one or more of their layers to capture the spatial and temporal dependencies in data with a grid-like topology. In the context of WiFi fingerprint data, one-dimensional convolutional neural networks (1D-CNNs) have been employed. The CNNLoc method, which leverages 1D-CNNs, achieved a positioning accuracy of 7.60 m on the 44,000-square-meters UTSIndoorLoc dataset [11]. In addition, 1D-CNNs were utilised to provide baseline performance in the CSUIndoorLoc, a CSS-RSS fingerprint dataset collected in a university building, demonstrating a positioning error of 1.38 m [19]. Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) architecture that is particularly designed for sequential data learning, such as continuous WiFi fingerprints recorded during trajectory walking. In the XJTLUIndoorLoc RSS dataset, LSTM achieved a positioning accuracy of 0.62 m in a building hall indoor environment [17]. Bidirectional LSTM (BiLSTM) networks were proposed in [52] for evaluating positioning performance on WiFi-BLE and RSS-RTT datasets, achieving positioning accuracies of 0.82 m and 0.70 m, respectively.

V. PUBLIC WiFi FINGERPRINT DATASET GUIDELINE

This section outlines the limitations of current public WiFi fingerprint datasets and proposes a standard for future open access dataset publications.

A. LIMITATIONS

Despite the continuous publication and sharing of new WiFi fingerprint datasets with the research community each year, several prevalent issues in the dataset publication have drawn our attention. Addressing these challenges in existing publicly available WiFi datasets would greatly enhance the open and collaborative environment in the research community and accelerate the development of WiFi fingerprinting-based indoor positioning systems.

Several limitations have been identified in the available public WiFi fingerprint datasets. Firstly, there is a significant shortage of RTT and CSI fingerprint datasets, with only four and three public datasets use WiFi RTT and CSI as signal inputs, respectively. More public datasets are needed to fill this gap. In addition, key WiFi signal related information such as LOS condition of the APs is rarely

TABLE 3. ***(Continued.)*** **Links and notes of existing publicly available WiFi fingerprint dataset.**

13	https://doi.org/10.5281/zenodo.1161525	2.9 MB	TUT5 dataset. Details were found in [89] and [20].	[89]
14	https://data.nist.gov/od/id/mds2-3145	4.89 GB	One office building; two industrial shop and warehouse types of buildings and a subterranean structure. RPs determined by a surveying contractor. WiFi, cellular, GPS and IMU data. In PBS format. Has 4 phones and 38 scenarios, and each has different features. Different phone detected different number of APs per scenario.	[91] [92]
15	https://github.com/ZzhKlaus/2018-SURF-Trajectory-Estimation	13.5 MB	WiFi, geomagnetic coordinates, and IMUs. Collected at LOS rectangle areas. Testbed empty.	[93] [17]
16	https://doi.org/10.5281/zenodo.2823964	208 MB	IPIN2018. Needs preprocessing for usage. WiFi, magnetic, IMU, pressure/sound/illumination, GNSS. Number of APs and samples were counted by the authors. Testbed not empty.	[62] [63]
17	https://dx.doi.org/10.21227/1yd5-rn96	375.75 MB	The open-access datasets with CSI fingerprints can be found at https://www.kaggle.com/datasets/brosnanyuen/wifi-rssi-indoor-localization. The positioning accuracy of the robot is 0.07 m $\pm$ 0.02 m. 5 AP provide 2 distinct MAC address for 2.4 GHz and 5 GHz channels respectively. Used robot collector.	[68] [51]
18	https://doi.org/10.5281/zenodo.3342526	2.39 MB	MINT1 dataset. Wi-Fi fingerprints were collected simultaneously from multiple, synchronised, WiFi interfaces. For autonomous industrial vehicles. Testbed not empty. There are two links that refer to the same dataset as introduced by [97], with only one included.	[94], [95] [96], [97]
19	https://github.com/XudongSong/UTSIndoorLoc-dataset	21.8 MB	Few details were found in [11] and the dataset description page.The rest extracted by the authors.	[98] [11]
20	https://drive.google.com/drive/folders/1_z1qhoRIcpineP9AHkfVGCfB2Fd_e-fD	21.1 MB	Dataset GitHub link expired at the time of this paper's revision. Be really careful when copy and paste this link. Please consider using the Google Drive link in corresponding paper.	[99] [14]
21	https://doi.org/10.5281/zenodo.3606765	831 MB	IPIN2019. Needs preprocessing for usage. WiFi, magnetic, IMU, pressure/sound/illumination, GNSS, 3D orientation. Number of APs and samples were counted by the authors. Phone: mostly stable in front of his face or chest. 6,000 m^2 indoor, 1,000 m^2 outdoor.	[100] [101]
22	https://dx.doi.org/10.21227/49yg-5d21	27.9 MB	Training and testing with different devices. Magnetic also collected. AP: 72 Cisco Aironet APs included. Only 2 floors included in the dataset)	[102] [103]
23	https://doi.org/10.5281/zenodo.3698238	278.2 MB	Three floors for mall 1 and two floors for mall 2. Accuracy regardless of smartphone used. RP is each shop. Validated even 74 days later. Used shop distance to measure errors. Randomly placed phones. Testbed not empty.	[53] [104]
24	https://doi.org/10.5281/zenodo.3819917	54.5 MB	TUT6,TUT7 datasets. Not well specified in the related paper and data description page. More details are only found in [20].	[105] [106]

reported in the literature. In terms of dataset publication, the description sections are often insufficient, with some fundamental elements only available in the related research papers. However, it appears that these datasets typically provide limited information about their features and often lack details about the research papers in which they were utilised. The disorgnised presentation of dataset features poses a major challenge for researchers seeking to select the most suitable dataset. To make matters worse, mismatches have been identified among the dataset description pages, related research papers, and their citations. As a result, the authors have had to invest considerable effort in extracting some critical dataset features in creating the comparison table. Furthermore, many datasets become inaccessible 4 to 5 years after the publication of the related papers, likely due to expired links.

For the datasets that were excluded, several major issues were identified. Most of the excluded datasets either lack a legitimate open access link or do not providing adequate information and description of their features. These shortcomings result in limited usability due to restricted access and incomplete dataset elements. Next, some datasets suffer from problems related to expiring links and lack of maintenance. For instance, several datasets published on https://crawd ad.org/ can no longer be found. And a previous open-source indoor positioning systems and datasets repository (http://lsr.cs.upb.ro/tool/1) also has faced similar issue [12]. Moreover, datasets that are wrongly named or tagged with 'WiFi' further complicate the selection of a suitable WiFi fingerprint dataset.

B. INFLUENCE OF DATASET FEATURES ON THE REPORTED PERFORMANCE

Before constructing of a WiFi fingerprint dataset, it is essential to understand the key features that may influence its performance. Although not all of the datasets reviewed provide comprehensive details about their data collection processes and general features, this review still offers valuable insights into factors affecting dataset performance.

Intuitively, one might assume that deploying as many APs as possible would create more unique fingerprints for each location, leading to better accuracy. However, as shown in Figure 9a, a higher number of APs does not necessarily improve dataset performance. This is because, in a large-scale testbed, most background APs are often undetected at a given

TABLE 3. ***(Continued.)*** **Links and notes of existing publicly available WiFi fingerprint dataset.**

25	https://doi.org/10.5281/zenodo.3778646	1.64 MB	DSI1, DSI2 datasets. Detailed description is included in the dataset.	[107]
26	https://doi.org/10.5281/zenodo.3751042	967 KB	SIM dataset. Only available at [55].	[55] [72]
27	https://github.com/pspachos/RSSI-Dataset-for-Indoor-Localization-Fingerprinting	3.19 MB	Offers different interference level datasets. Only LOS. AP: Raspberry Pi 3 Model Bs.	[108] [109]
28	https://github.com/IS2AI/WiFine	77.4 MB	290 trajectories. Testbed not empty.	[57] [23]
29	https://doi.org/10.5281/zenodo.4314992	703 MB	IPIN2020. Needs preprocessing for usage. WiFi, magnetic, IMU, pressure/sound/illumination, GNSS, 3D orientation. Number of APs and samples were counted by the authors. Phone: mostly stable in front of his face or chest in training, realistic in val and eval.	[110] [111]
30	https://dx.doi.org/10.21227/h5c2-5439	102 MB	Different AP used for train, val and eval. Mainly for range estimation. AP: PC with Intelr Wireless-AC8260 Wi-Fi 802.11ac, 2 × 2 Dual-Band chipset. Used robot collector.	[41] [58]
31	https://doi.org/10.5281/zenodo.5174851	53.6MB	TIE1, SAH1 datasets. Contain coordinates X,Y,Z, and floor id. No related paper found. Not well specified in the related paper and data description page. More details are only found in [20]. Current dataset published 4 years after the relevant paper.	[54]
32	https://doi.org/10.6084/m9.figshare.13607540.v2	3.36 MB	Result based on adaptive federated Kalman filtering (FKF), and path loss based distance. Could be found via multiple sources.	[112] [113]
33	https://doi.org/10.5281/zenodo.4744380	133 MB	WiFi, BLE and IMU. Phone: chest height. Further details could be found on the official website of OpenHPS. Reported performance result only based on WiFi. Testbed empty.	[64] [114]
34	https://doi.org/10.5281/zenodo.5948678	990 MB	IPIN2021. Needs preprocessing for usage. WiFi, BLE, magnetic, IMU, pressure/sound/illumination, GNSS, 3D orientation. Number of APs and samples were counted by the authors. Phone: mostly stable in front of his face or chest in training, realistic in val and eval.	[115] [116]
35	https://github.com/ImAshRayan/Wi-MEST	88.1 MB	User diversity and device heterogeneity were considered. Long time span. Unable to read the disrupted CSV files for further details. Testbed not empty.	[117] [22]
36	https://kaggle.com/competitions/indoor-location-navigation	59.96 GB	WiFi, geomagnetic field, iBeacons, IMU. Android smartphone was held flat in front of the surveyor's body.	[118]
37	https://doi.org/10.5281/zenodo.7193602	22.0 MB	31 new RPs with 775 samples for testing. AP: ESP32 micro controller board. Current dataset published 5 years after the relevant paper.	[60] [50]

location, resulting in a fingerprint that primarily composed of default or artificial WiFi readings.

Moreover, larger RPs number and wider interval between them do not guarantee a stable dataset performance. A greater number of more sparsely positioned RPs is expected to produce more distinctive fingerprints. However, as illustrated in Figure 9b and 9c, no clear correlation was found between the configuration and arrangement of RPs, and the reported dataset performance in the included WiFi fingerprint datasets. This could be because that the number and interval of RPs in most existing datasets are constrained within a certain range due to the efficiency in manual data collection.

Furthermore, the use of higher frequency in the collection of WiFi signal measurements has minimal influence on the final dataset performance, as shown in Figure 9d. A higher sampling rate merely produce a greater number of similar WiFi fingerprints within a fixed time period, which is insufficient for capturing the fluctuating patterns of WiFi signals. The temporal variations in WiFi signals can only be effectively captured by recording each data sample at the same RP over a longer time interval.

As shown in Figure 10, datasets intended for 3D positioning tend to exhibit more unstable performance. This is because WiFi signal measurements do not vary significantly in the vertical direction, especially in buildings where the floor height is less than 4 metres, while the area spans over 90 metres by 15 metres. For multi-storey buildings, the dataset faces the similar challenges as those presented by a large number of APs.

On the other hand, it is intuitive that a smaller testbed, a larger number of data samples, and more informative WiFi signal measures (e.g., CSI, RTT) contribute to greater stability and improved dataset performance.

C. GUIDELINE FOR PUBLIC WiFi FINGERPRINT DATASET PUBLICATION

To prevent similar issues in the future creating and sharing public WiFi fingerprint datasets, we propose a detailed step-by-step guideline to assist researchers in their dataset publication efforts.

Firstly, after selecting the appropriate location type and determining the real-world testbed for dataset collection, researchers must conduct a thorough survey of the entire testing area. This includes checking the interior and basic structure of the selected building, assessing if there are enough background APs, and mapping the entire area suitable for WiFi fingerprinting tasks. Ideally, this should involve providing a detailed floor map and the square footage of

TABLE 3. ***(Continued.)*** **Links and notes of existing publicly available WiFi fingerprint dataset.**

38	https://doi.org/10.5281/zenodo.6928554	6.20 GB	The dataset contains a long-term dataset and 12 site-survey datasets. The long-term dataset was collected by WiFi receivers at fixed locations, used as training dataset.	[119] [120]
39	https://doi.org/10.5281/zenodo.5885636	2.6 GB	The main idea is to reserve users' privacy during CSI-based positioning. Reported positioning result is the positioning precision. AP: Ettus USRP N300.	[44] [121]
40	https://doi.org/10.5281/zenodo.11558192	12.6 MB	Contains WiFi RTT measurements and LOS conditions of all APs at every RP. Only RTT-enabled APs were included. Contains a scenario with both LOS and NLOS conditions, a scenario with complete LOS conditions, and a scenario with only 1 NLOS AP. AP: Google WiFi Router AC-1304. Testbed empty.	[40] [6], [33]
41	https://github.com/m-nabati/WiFi-RSSI-Localization-Dataset	2.31 MB	Though the related paper proposed two datasets, only the first one was made public.	[123] [124]
42	https://github.com/renwudao24/SODIndoorLoc	38.0 MB	FloorID ranges from 1 to 3 in CETC331 subset, FloorID is 4 for HCXY and SYL subsets. Corridor, office room and meeting room identified. Height of the phone recorded.	[125] [18]
43	https://doi.org/10.5281/zenodo.7260097	592 MB	Dual-band. Two buildings were only used for testing. The baseline result was from F04 Device 1, 5GHz, using mean value to generate radio maps.	[66] [21]
44	https://github.com/EPIC-CSU/csi-rssi-dataset-indoor-nav	202 MB	The height from the Raspberry Pi to the ground is fixed at 120 cm. Number of RP and samples were counted by the authors. Every room door has two RPs. Best result from AP1. Testbed not empty.	[45] [19]
45	https://doi.org/10.5281/zenodo.7641701	10.9 MB	The continuous network measurements were taken at each location for a minimum of 10 s and later matched to the coordinates based on the temporal proximity, where valid measurements are taken 3 or less seconds before the marked position.	[126] [127]
46	https://doi.org/10.5281/zenodo.7826540	34.6 MB	Wi-Fi, wheel encoder (displacement), and IMU. AP:3 × ORiNOCO AP200, 6 Cisco Aironet 1100 series. For vehicle. Testbed empty.	[70] [71]
47	https://doi.org/10.5281/zenodo.11558792	1.62 MB	Contains WiFi RTT measurements and LOS conditions of all APs at every RP. Only RTT-enabled APs were included. Contains a scenario with both LOS and NLOS conditions, a scenario with complete LOS conditions, and a scenario with complete NLOS conditions. AP: Google WiFi Router AC-1304. Testbed empty.	[42] [39], [48]
48	https://doi.org/10.5281/zenodo.7612915	1.53 GB	IPIN2022. Needs preprocessing for usage. WiFi, BLE, magnetic, IMU, pressure/sound/illumination, GNSS, 3D orientation. Number of APs and samples were counted by the authors. Phone: mostly stable in front of his face or chest in training, realistic in val and eval.	[67] [116]
49	https://doi.org/10.5281/zenodo.8362205	276 MB	IPIN2023. Needs preprocessing for usage. WiFi, BLE, magnetic, IMU, pressure/sound/illumination, GNSS, 3D orientation. Number of APs and samples were counted by the authors. Phone: mostly stable in front of his face or chest in training, realistic in val and eval. Each regular training trial has been collected 4 times.	[67]

TABLE 3. ***(Continued.)*** **Links and notes of existing publicly available WiFi fingerprint dataset.**

50	https://doi.org/10.5281/zenodo.10862916	741 MB	WiFi and BLE. Phone: 75cm height. Data only measured at the Router Side. The paper mentioned WiFi RSS+RTT, and BLE RSS, however, this dataset only contains WiFi and BLE RSS. AP: ESP32C3 chip-set. Used robot collector.	[59] [52]
51	https://doi.org/10.5281/zenodo.10883013	40.5 MB	Phone: 75cm height. Data only measured at the Router Side. The paper mentioned WiFi RSS+RTT, and BLE RSS, however, this dataset only contains WiFi RSS and RTT. AP: ESP32C3 chip-set. Used robot collector.	[43] [52]
52	https://doi.org/10.5281/zenodo.10715595	121 MB	CSI-based dataset in a small testbed. Six trajectories, 4 train, 1 val, 1 test. AP: ESP32-S3-WROOM-1U3 microcontroller, ALFA Network APA-M25 antenna. Testbed empty.	[46] [128]

the testbed. If the pre-installed APs do not deliver the desired WiFi signal type, additional WiFi routers should be positioned similarly to the existing APs.

Secondly, careful attention must be given to the division of the RPs and the acquisition of ground truth labels. Whether the dataset involves 3D indoor positioning, which includes height information represented by floor IDs or Z-axis values, or 2D mapping, accurate labeling of ground truth coordinates is critically important. Simple tools such as measuring tapes, post-it notes, and the existing tiles on the building floor can greatly assist researchers who may have access to advanced ground truth measuring methods. It is important to note that trajectory tracking datasets may result in lower positioning performance due to rapid movement and insufficient WiFi measurements at each location.

Thirdly, the methodology for recording WiFi fingerprints must be clearly defined. Researchers should begin by selecting the devices for data collection, specifying their orientation, and determining how the user will hold them. It is crucial to decide which signal types will be collected, the frequency of WiFi measurements, the number of measurements required at each RP, and the duration of the entire collection process. All these details must be meticulously documented for future reference before starting data collection to ensure consistency and reproducibility. Introduced in Section IV-A, to indicate APs that are not heard from current RP, default

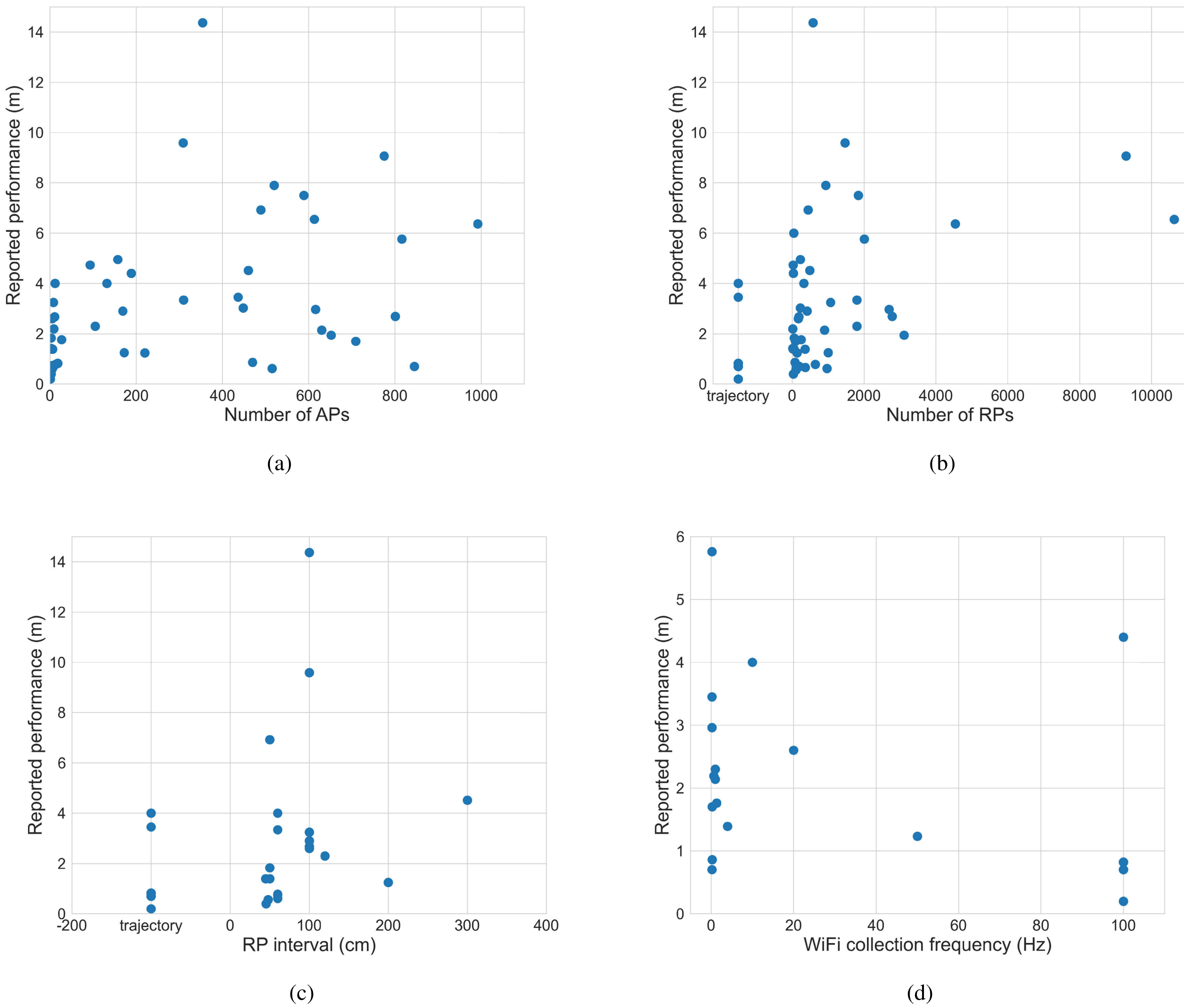


FIGURE 9. **The influence of various dataset features on the reported performance. The 'trajectory' value indicates that the dataset consists solely of trajectory recordings.**

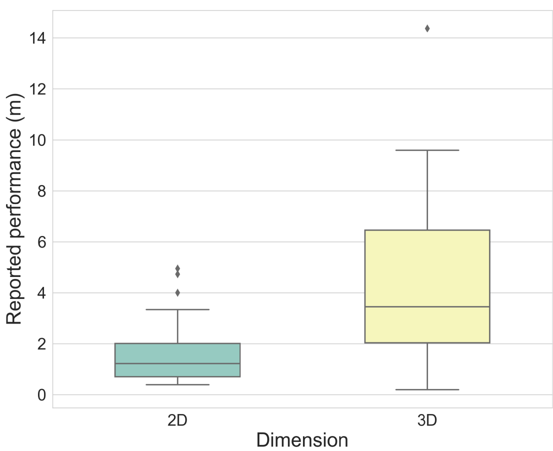


FIGURE 10. **The reported performance of 2D and 3D WiFi fingerprint datasets.**

artificial values should be utilised, as shown in Table 1. Additionally, it is important to record the real-world position of the (0,0) starting point in the testbed, the orientations of the XYZ axes, the LOS conditions for each AP, and environmental factors such as pedestrian traffic and the types of rooms covered.

Finally, it is crucial to publish the WiFi fingerprint dataset, along with comprehensive collection details, on a reliable and open-access platform for future use by the research community. Providing the detailed dataset collection information discussed above in the dataset description is essential for developing new positioning algorithms and replicating existing results. Zenodo, used as the primary publish platform for 30 out of 52, is highly recommended for publishing open-access WiFi fingerprint dataset. Zenodo is an open access and easy to use repository developed by CERN (European Organization for Nuclear Research) as part of the OpenAIRE project to support the European Commission's Open Data policy. With features like versioning, GitHub integration, and usage statistics, Zenodo ensures that the published datasets are citable, accessible, and trackable, promoting open science for all [129].

VI. CONCLUSION

This paper provides a comprehensive and detailed review of over 50 publicly available WiFi fingerprint datasets, emphasizing their crucial role in the advancement of indoor positioning systems. Through our meticulous analysis, we underscore the challenges faced by researchers, including the dispersed publication of datasets across various platforms, inconsistencies in dataset organization and accessibility, and the ineffective and biased selection of public WiFi fingerprint datasets. Notably, all the datasets included in this review have been manually tested to ensure that they have current and accessible open access links.

This paper begins by defining the research scope and methodology of this review. followed by a detailed introduction to the background of WiFi fingerprinting and its various signal inputs. We then conduct an in-depth analysis of the open access WiFi fingerprint datasets, examining critical elements from a researcher's standpoint, including the size and location of the testbed, 2D/3D indoor positioning type, WiFi signal inputs, access points (APs), receiver devices, the number of RP covered, the number of WiFi fingerprint data samples, data collection interval (both temporal and spatial), ground truth acquisition methods, and reported positioning performance. The dataset features are meticulously and extensively extracted and compared even when they are not explicitly provided. Recent trends in these factors and their impact on WiFi fingerprinting performance are thoroughly discussed. Interestingly, we found that an increased number of reference points and access points, the use of 3D coordinates, larger RP intervals, and higher WiFi collection frequencies do not necessarily lead to better reported performance. However, a smaller testbed, a larger number of data samples, and more informative WiFi signal measures tend to contribute to more stable and accurate dataset performance. Finally, we summarise the limitations of existing WiFi fingerprint datasets, and propose standards and guidelines for future open access WiFi fingerprinting dataset publication, recommending Zenodo as the preferred platform. By addressing these challenges and setting guidelines for future public dataset publication, we aim to drive advancements in WiFi fingerprinting technologies, thereby enhancing the performance of indoor positioning systems.

REFERENCES

[1] F. Liu, J. Liu, Y. Yin, W. Wang, D. Hu, P. Chen, and Q. Niu, "Survey on WiFi-based indoor positioning techniques," *IET Commun.*, vol. 14, no. 9, pp. 1372–1383, Jun. 2020.

[2] N. Singh, S. Choe, and R. Punmiya, "Machine learning based indoor localization using Wi-Fi RSSI fingerprints: An overview," *IEEE Access*, vol. 9, pp. 127150–127174, 2021.

[3] S. Shang and L. Wang, "Overview of WiFi fingerprinting-based indoor positioning," *IET Commun.*, vol. 16, no. 7, pp. 725–733, Apr. 2022.

[4] V. Bellavista-Parent, J. Torres-Sospedra, and A. Perez-Navarro, "New trends in indoor positioning based on WiFi and machine learning: A systematic review," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Nov. 2021, pp. 1–8.

[5] C. Yang and H.-r. Shao, "WiFi-based indoor positioning," *IEEE Commun. Mag.*, vol. 53, no. 3, pp. 150–157, Mar. 2015.

[6] X. Feng, K. A. Nguyen, and Z. Luo, "WiFi round-trip time (RTT) fingerprinting: An analysis of the properties and the performance in non-line-of-sight environments," *J. Location Based Services*, vol. 17, no. 4, pp. 307–339, Oct. 2023.

[7] X. Feng, K. A. Nguyen, and Z. Luo, "A survey of deep learning approaches for WiFi-based indoor positioning," *J. Inf. Telecommun.*, vol. 6, no. 2, pp. 163–216, Apr. 2022.

[8] W. Liu, Q. Cheng, Z. Deng, H. Chen, X. Fu, X. Zheng, S. Zheng, C. Chen, and S. Wang, "Survey on CSI-based indoor positioning systems and recent advances," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Sep. 2019, pp. 1–8.

[9] K. Kandasamy, W. Neiswanger, J. Schneider, B. Poczos, and E. P. Xing, "Neural architecture search with Bayesian optimisation and optimal transport," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018.

[10] M. T. Hoang, B. Yuen, X. Dong, T. Lu, R. Westendorp, and K. Reddy, "Recurrent neural networks for accurate RSSI indoor localization," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 10639–10651, Dec. 2019.

[11] X. Song, X. Fan, C. Xiang, Q. Ye, L. Liu, Z. Wang, X. He, N. Yang, and G. Fang, "A novel convolutional neural network based indoor localization framework with WiFi fingerprinting," *IEEE Access*, vol. 7, pp. 110698–110709, 2019.

[12] I. E. Radoi, D. Cirimpei, and V. Radu, "Localization systems repository: A platform for open-source localization systems and datasets," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Sep. 2019, pp. 1–8.

[13] R. Montoliu, E. Sansano, J. Torres-Sospedra, and O. Belmonte, "Indoorloc platform: A public repository for comparing and evaluating indoor positioning systems," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Sep. 2017, pp. 1–8.

[14] P. Roy, C. Chowdhury, D. Ghosh, and S. Bandyopadhyay, "JUIndoorLoc: A ubiquitous framework for smartphone-based indoor localization subject to context and device heterogeneity," *Wireless Pers. Commun.*, vol. 106, no. 2, pp. 739–762, May 2019.

[15] J. Torres-Sospedra, R. Montoliu, A. Martínez-Usó, J. P. Avariento, T. J. Arnau, M. Benedito-Bordonau, and J. Huerta, "UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Oct. 2014, pp. 261–270.

[16] J. Torres-Sospedra, A. Jiménez, S. Knauth, A. Moreira, Y. Beer, T. Fetzer, V.-C. Ta, R. Montoliu, F. Seco, G. Mendoza-Silva, O. Belmonte, A. Koukofikis, M. Nicolau, A. Costa, F. Meneses, F. Ebner, F. Deinzer, D. Vaufreydaz, T.-K. Dao, and E. Castelli, "The smartphone-based offline indoor location competition at IPIN 2016: Analysis and future work," *Sensors*, vol. 17, no. 3, p. 557, Mar. 2017.

[17] Z. Zhong, Z. Tang, X. Li, T. Yuan, Y. Yang, M. Wei, Y. Zhang, R. Sheng, N. Grant, C. Ling, X. Huan, K. S. Kim, and S. Lee, "XJTLUIndoorLoc: A new fingerprinting database for indoor localization and trajectory estimation based on Wi-Fi RSS and geomagnetic field," in *Proc. 6th Int. Symp. Comput. Netw. Workshops (CANDARW)*, Nov. 2018, pp. 228–234.

[18] J. Bi, Y. Wang, B. Yu, H. Cao, T. Shi, and L. Huang, "Supplementary open dataset for WiFi indoor localization based on received signal strength," *Satell. Navigat.*, vol. 3, no. 1, p. 25, Nov. 2022.

[19] L. Wang and S. Pasricha, "A framework for CSI-based indoor localization with ID convolutional neural networks," in *Proc. IEEE 12th Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Sep. 2022, pp. 1–8.

[20] J. Torres-Sospedra, D. P. Quezada Gaibor, J. Nurmi, Y. Koucheryavy, E. S. Lohan, and J. Huerta, "Scalable and efficient clustering for fingerprint-based positioning," *IEEE Internet Things J.*, vol. 10, no. 4, pp. 3484–3499, Feb. 2023.

[21] A. Abdullah, M. Haris, O. A. Aziz, R. A. Rashid, and A. S. Abdullah, "UTMInDualSymFi: A dual-band Wi-Fi dataset for fingerprinting positioning in symmetric indoor environments," *Data*, vol. 8, no. 1, p. 14, Jan. 2023.

[22] I. Ashraf, S. Din, S. Hur, and Y. Park, "Wi-Fi positioning dataset with multiusers and multidevices considering spatio-temporal variations," *Comput., Mater. Continua*, vol. 70, no. 3, pp. 5213–5232, 2022.

[23] Y. Khassanov, M. Nurpeiissov, A. Sarkytbayev, A. Kuzdeuov, and H. A. Varol, "Finer-level sequential WiFi-based indoor localization," in *Proc. IEEE/SICE Int. Symp. Syst. Integr. (SII)*, Jan. 2021, pp. 163–169.

[24] X. Feng, K. A. Nguyen, and Z. Luo, "An analysis of the properties and the performance of WiFi RTT for indoor positioning in non-line-of-sight environments," in *Proc. 17th Int. Conf. Location Based Services*, 2022.

[25] D. Quezada-Gaibor, L. Klus, J. Torres-Sospedra, E. S. Lohan, J. Nurmi, C. Granell, and J. Huerta, "Data cleansing for indoor positioning Wi-Fi fingerprinting datasets," in *Proc. 23rd IEEE Int. Conf. Mobile Data Manage. (MDM)*, Jun. 2022, pp. 349–354.

[26] C. Fan, M. Chen, X. Wang, J. Wang, and B. Huang, "A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data," *Frontiers Energy Res.*, vol. 9, Mar. 2021, Art. no. 652801.

[27] S. He and S.-H. G. Chan, "Wi-Fi fingerprint-based indoor positioning: Recent advances and comparisons," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 466–490, 1st Quart., 2016.

[28] J. M. Rocamora, I. Wang-Hei Ho, W. Mak, and A. P. Lau, "Survey of CSI fingerprinting-based indoor positioning and mobility tracking systems," *IET Signal Process.*, vol. 14, no. 7, pp. 407–419, Sep. 2020.

[29] W. K. Zegeye, S. B. Amsalu, Y. Astatke, and F. Moazzami, "WiFi RSS fingerprinting indoor localization for mobile devices," in *Proc. IEEE 7th Annu. Ubiquitous Comput., Electron. Mobile Commun. Conf. (UEMCON)*, Oct. 2016, pp. 1–6.

[30] C. BASRI and A. El Khadimi, "Survey on indoor localization system and recent advances of WiFi fingerprinting technique," in *Proc. 5th Int. Conf. Multimedia Comput. Syst. (ICMCS)*, Sep. 2016, pp. 253–259.

[31] S. Xia, Y. Liu, G. Yuan, M. Zhu, and Z. Wang, "Indoor fingerprint positioning based on Wi-Fi: An overview," *ISPRS Int. J. Geo-Information*, vol. 6, no. 5, p. 135, Apr. 2017.

[32] A. Khalajmehrabadi, N. Gatsis, and D. Akopian, "Modern WLAN fingerprinting indoor positioning methods and deployment challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1974–2002, 3rd Quart., 2017.

[33] X. Feng, K. A. Nguyen, and Z. Luo, "A multi-scale feature selection framework for WiFi access points Line-of-sight identification," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2023, pp. 1–6.

[34] Y. Wang, C. Xiu, X. Zhang, and D. Yang, "WiFi indoor localization with CSI fingerprinting-based random forest," *Sensors*, vol. 18, no. 9, p. 2869, Aug. 2018.

[35] G. Guo, R. Chen, F. Ye, X. Peng, Z. Liu, and Y. Pan, "Indoor smartphone localization: A hybrid WiFi RTT-RSS ranging approach," *IEEE Access*, vol. 7, pp. 176767–176781, 2019.

[36] X. Feng, K. A. Nguyen, and Z. Luo, "WiFi access points line-of-sight detection for indoor positioning using the signal round trip time," *Remote Sens.*, vol. 14, no. 23, p. 6052, Nov. 2022.

[37] C. Gentner, M. Ulmschneider, I. Kuehner, and A. Dammann, "WiFi-RTT indoor positioning," in *Proc. IEEE/ION Position, Location Navigat. Symp. (PLANS)*, Apr. 2020, pp. 1029–1035.

[38] H. Cao, Y. Wang, J. Bi, S. Xu, M. Si, and H. Qi, "Indoor positioning method using WiFi RTT based on LOS identification and range calibration," *ISPRS Int. J. Geo-Inf.*, vol. 9, no. 11, p. 627, Oct. 2020.

[39] X. Feng, K. A. Nguyen, and Z. Luo, "A dynamic model switching algorithm for WiFi fingerprinting indoor positioning," in *Proc. 13th Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Sep. 2023, pp. 1–6.

[40] X. Feng, K. A. Nguyen, and Z. Luo, "WiFi RTT RSS dataset for indoor positioning," Zenodo, doi: 10.5281/zenodo.11558192. Accessed: Jul. 14, 2024.

[41] N. Dvorecki, O. Bar-Shalom, L. Banin, and Y. Amizur, "Intel open Wi-Fi RTT dataset," *IEEE Dataport*, doi: 10.21227/h5c2-5439. Accessed: Jun. 11, 2024.

[42] X. Feng, K. A. Nguyen, and Z. Luo, "WiFi RSS & RTT dataset with different LOS conditions for indoor positioning," Zenodo, 2024, doi: 10.5281/zenodo.11558792.

[43] B. Yuen, Y. Bie, D. Cairns, G. Harper, J. Xu, X. Dong, and T. Lu, "ESP32C3 WiFi FTM RSSI indoor localization," Zenodo, 2024, doi: 10.5281/zenodo.10883013.

[44] M. Cominelli, F. Gringoli, and R. Lo Cigno, "CSI-based device-free localization and obfuscation (1.0)," Zenodo, 2022, doi: 10.5281/zenodo.5885636.

[45] L. Wang and S. Pasricha. (2022). *Csuindoorloc*. [Online]. Available: https://github.com/EPIC-CSU/csi-rssi-dataset-indoor-nav

[46] J. Strohmayer and M. Kampel, "HALOC dataset | WiFi CSI-based long-range person localization using directional antennas (1.0.0)," in *Proc. 12th Int. Conf. Learn. Represent. (ICLR)*. Vienna, Austria: Zenodo, 2024, doi: 10.5281/zenodo.10715595.

[47] J. Torres-Sospedra, R. Montoliu, A. Martínez-Usó, J. P. Avariento, and T. J. Arnau, *UJIIndoorLoc*. Irvine, CA, USA: UCI Machine Learning Repository, 2014, doi: 10.24432/C5MS59.

[48] X. Feng, K. A. Nguyen, and Z. Luo, "A Wi-Fi RSS-RTT indoor positioning model based on dynamic model switching algorithm," *IEEE J. Indoor Seamless Positioning Navigat.*, vol. 2, pp. 151–165, 2024.

[49] P. Barsocchi, A. Crivello, D. La Rosa, and F. Palumbo, "A multisource and multivariate dataset for indoor localization methods based on WLAN and geo-magnetic field fingerprinting," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Oct. 2016, pp. 1–8.

[50] T. Bravenec, M. Gould, T. Fryza, and J. Torres-Sospedra, "Influence of measured radio map interpolation on indoor positioning algorithms," *IEEE Sensors J.*, vol. 23, no. 17, pp. 20044–20054, Sep. 2023.

[51] M. T. Hoang, Y. Zhu, B. Yuen, T. Reese, X. Dong, T. Lu, R. Westendorp, and M. Xie, "A soft range limited K-Nearest neighbors algorithm for indoor localization enhancement," *IEEE Sensors J.*, vol. 18, no. 24, pp. 10208–10216, Dec. 2018.

[52] B. Yuen, Y. Bie, D. Cairns, G. Harper, J. Xu, C. Chang, X. Dong, and T. Lu, "Wi-Fi and Bluetooth contact tracing without user intervention," *IEEE Access*, vol. 10, pp. 91027–91044, 2022.

[53] J.-A. López-Pastor, A.-J. Ruiz-Ruiz, A.-J. García-Sánchez, and J.-L.Gómez-Tornero, "Wi-Fi RSSI fingerprint dataset from two malls with validation routes in a shop-level for indoor positioning (version 1)," Zenodo, 2020, doi: 10.5281/zenodo.3698238.

[54] E. S. Lohan, J. Torres-Sospedra, and A. Gonzalez, "WiFi RSS measurements in tampere university multi-building campus, 2017 (version 1)," Zenodo, 2021, doi: 10.5281/zenodo.5174851.

[55] J. Torres-Sospedra, D. Quezada-Gaibor, G. Mendoza-Silva, J. Nurmi, Y. Koucheryavy, and J. Huerta, "Supplementary materials for 'new cluster selection and fine-grained search for k-means clustering and Wi-Fi fingerprinting' (1.0)," Zenodo, 2020, doi: 10.5281/zenodo.3751042.

[56] A. R. Jiménez, G. M. Mendoza-Silva, R. Montoliu, F. Seco, and J. Torres-Sospedra, "Datasets and supporting materials for the IPIN 2016 competition track 3 (smartphone-based, off-site) (1.0)," Zenodo, 2016, doi: 10.5281/zenodo.2791530.

[57] Y. Khassanov, M. Nurpeiissov, A. Sarkytbayev, A. Kuzdeuov, and H. A. Varol. (2020). *Wifine*. [Online]. Available: https://github.com/IS2AI/WiFine

[58] N. Dvorecki, O. Bar-Shalom, L. Banin, and Y. Amizur, "A machine learning approach for Wi-Fi RTT ranging," in *Proc. Int. Tech. Meeting Inst. Navigat.*, Feb. 2019, pp. 435–444.

[59] B. Yuen, Y. Bie, D. Cairns, G. Harper, J. Xu, C. Chang, X. Dong, and T. Lu, "WiFi and Bluetooth RSSI SQI indoor localization," Zenodo, 2024, doi: 10.5281/zenodo.10862916.

[60] T. Bravenec, M. Gould, T. Frýza, and J. Torres-Sospedra, "Supplementary materials for 'influence of measured radio environment map interpolation on indoor positioning algorithms' (1.0)," Tech. Rep., 2022, doi: 10.5281/zenodo.7193602.

[61] R. Montoliu. (2016). *IPIN 2016 tutorial 2: 'Fingerprinting-Based Indoor Positioning*. [Online]. Available: http://indoorloc.uji.es/ipin2016tutorial

[62] A. R. Jiménez, G. M. Mendoza-Silva, M. Ortiz, A. Perez-Navarro, J. Perul, F. Seco, and J. Torres-Sospedra, "Datasets and supporting materials for the ipin 2018 competition track 3 (smartphone-based, off-site) (1.0)," Zenodo, 2018, doi: 10.5281/zenodo.2823964.

[63] V. Renaudin, "Evaluating indoor positioning systems in a shopping mall: The lessons learned from the IPIN 2018 competition," *IEEE Access*, vol. 7, pp. 148594–148628, 2019.

[64] M. Van de Wynckel and B. Signer, "OpenHPS: Single floor fingerprinting and trajectory dataset (2021-05-09)," Zenodo, 2021, doi: 10.5281/zenodo.4744380.

[65] P. Barsocchi, A. Crivello, D. Rosa, and F. Palumbo, *Geo-Magnetic Field and WLAN Dataset for Indoor Localisation From Wristband and Smartphone*. Irvine, CA, USA: UCI Machine Learning Repository, 2017, doi: 10.24432/C5DW43..

[66] A. Abdullah, M. Haris, O. A. Aziz, R. A. Rashid, and A. S. Abdullah, "UTMInDualSymFi: A dual-band Wi-Fi dataset for fingerprinting positioning in symmetric indoor environments," *Data*, vol. 8, no. 1, p. 14, 2023.

[67] J. Torres-Sospedra, I. Silva, C. Pendao, A. Moreira, F. Meneses, A. Costa, M. J. Nicolau, A. Gonzalez-Perez, A. R. Jiménez, and A. Pérez-Navarro, "Datasets and supporting materials for the ipin 2022 competition track 3 (smartphone-based, off-site)," Zenodo, 2023, doi: 10.5281/zenodo.7612915.

[68] M. T. Hoang, X. Dong, T. Lu, B. Yuen, and R. Westendorp, "WiFi RSSI indoor localization," *IEEE Dataport*, 2019, doi: 10.21227/1yd5-rn96.

[69] S. Caccamo, R. Parasuraman, F. Båberg, and P. Ögren. (2016). *Indoor-RSSI-Mobile-Robot*. [Online]. Available: https://github.com/herolab-uga/indoor-rssi-mobile-robot

[70] I. Silva, C. Pendão, J. Torres-Sospedra, and A. Moreira, "Dataset for vehicle indoor positioning in industrial environments with Wi-Fi, inertial, and odometry data (1.0)," Zenodo, 2023, doi: 10.5281/zenodo.7826540.

[71] I. Silva, C. Pendão, J. Torres-Sospedra, and A. Moreira, "Industrial environment multi-sensor dataset for vehicle indoor tracking with Wi-Fi, inertial and odometry data," *Data*, vol. 8, no. 10, p. 157, Oct. 2023.

[72] J. Torres-Sospedra and A. Moreira, "Analysis of sources of large positioning errors in deterministic fingerprinting," *Sensors*, vol. 17, no. 12, p. 2736, Nov. 2017.

[73] S. Shrestha, J. Talvitie, and E. S. Lohan, "Deconvolution-based indoor localization with Wlan signals and unknown access point locations," in *Proc. Int. Conf. Localization GNSS (ICL-GNSS)*, Jun. 2013, pp. 1–6.

[74] J. T. Shrestha, Shweta and E. S. Lohan. (2013). *Tut1 and Tut2: Indoor WLAN Measurement Data*. [Online]. Available: http://www.cs.tut.fi/tlt/pos/Software.htm

[75] M. Agostini, A. Crivello, F. Palumbo, and F. Potortì, "An open-source framework for smartphone-based indoor localization," in *Proc. AI* AAL@ AI* IA*, 2017, pp. 74–86.

[76] S. Caccamo, R. Parasuraman, F. Båberg, and P. Ögren, "Extending a UGV teleoperation FLC interface with wireless network connectivity information," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2015, pp. 4305–4312.

[77] J. Torres-Sospedra, D. Rambla, R. Montoliu, O. Belmonte, and J. Huerta, *UJIIndoorLoc-Mag*. Irvine, CA, USA: UCI Machine Learning Repository, 2015, doi: 10.24432/C5D311.

[78] J. Torres-Sospedra, R. Montoliu, G. M. Mendoza-Silva, O. Belmonte, D. Rambla, and J. Huerta, "Providing databases for different indoor positioning technologies: Pros and cons of magnetic field and Wi-Fi based positioning," *Mobile Inf. Syst.*, vol. 2016, no. 1, pp. 1–22, 2016.

[79] J. Tams and Z. Tth, "Hybrid indoor positioning dataset from wifi rssi, Bluetooth and magnetometer," *UCI Mach. Learn. Repository*, 2016, doi: 10.24432/C55K6K.

[80] Z. Tóth and J. Tamás, "Miskolc IIS hybrid IPS: Dataset for hybrid indoor positioning," in *Proc. 26th Int. Conf. Radioelektronika (RADIOELEKTRONIKA)*, Apr. 2016, pp. 408–412.

[81] G. M. Mendoza-Silva, P. Richter, J. Torres-Sospedra, E. S. Lohan, and J. Huerta, "Long-term wi-fi fingerprinting dataset and supporting material (2.2)," 2020, data set. https://doi.org/10.5281/zenodo.3748719

[82] G. Mendoza-Silva, P. Richter, J. Torres-Sospedra, E. Lohan, and J. Huerta, "Long-term WiFi fingerprinting dataset for research on robust indoor positioning," *Data*, vol. 3, no. 1, p. 3, Jan. 2018.

[83] E. S. Lohan, J. Torres-Sospedra, P. Richter, H. Leppäkoski, J. Huerta, and A. Cramariuc, "Crowdsourced wifi database and benchmark software for indoor positioning," Zenodo, 2017, doi: 10.5281/zenodo.1001662..

[84] E. Lohan, J. Torres-Sospedra, H. Leppäkoski, P. Richter, Z. Peng, and J. Huerta, "Wi-Fi crowdsourced fingerprinting dataset for indoor positioning," *Data*, vol. 2, no. 4, p. 32, Oct. 2017.

[85] A. R. Jiménez, G. M. Mendoza-Silva, F. Seco, and J. Torres-Sospedra, "Datasets and supporting materials for the IPIN 2017 competition track 3 (smartphone-based, off-site) (1.0)," Zenodo, 2017, doi: 10.5281/zenodo.2823924.

[86] J. Torres-Sospedra, A. Jiménez, A. Moreira, T. Lungenstrass, W.-C. Lu, S. Knauth, G. Mendoza-Silva, F. Seco, A. Pérez-Navarro, M. Nicolau, A. Costa, F. Meneses, J. Farina, J. Morales, W.-C. Lu, H.-T. Cheng, S.-S. Yang, S.-H. Fang, Y.-R. Chien, and Y. Tsao, "Off-line evaluation of mobile-centric indoor positioning systems: The experiences from the 2017 IPIN competition," *Sensors*, vol. 18, no. 2, p. 487, Feb. 2018.

[87] R. Bhatt, *Wireless Indoor Localization*. Irvine, CA, USA: UCI Machine Learning Repository, 2017, doi: 10.24432/C51880.

[88] J. G. Rohra, B. Perumal, S. J. Narayanan, P. Thakur, and R. B. Bhatt, "User localization in an indoor environment using fuzzy hybrid of particle swarm optimization & gravitational search algorithm with neural networks," in *Proc. 6th Int. Conf. Soft Comput. Problem Solving (SocProS)*, vol. 1. Cham, Switzerland: Springer, 2017, pp. 286–295.

[89] P. Richter, E. S. Lohan, and J. Talvitie, "Wlan (WiFi) RSS database for fingerprinting positioning (1.0.0)," Zenodo, 2018, doi: 10.5281/zenodo.1161525.

[90] N. Moayeri, C. Li, and L. Shi, "PerfLoc (Part 2): Performance evaluation of the smartphone indoor localization apps," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Sep. 2018, pp. 1–8.

[91] N. Moayeri, "Smartphone data for development of indoor localization apps," Nat. Inst. Standards Technol., 2018, doi: 10.18434/mds2-3145. Accessed: Jun. 11, 2024.

[92] N. Moayeri, M. O. Ergin, F. Lemic, V. Handziski, and A. Wolisz, "PerfLoc (Part 1): An extensive data repository for development of smartphone indoor localization apps," in *Proc. IEEE 27th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Sep. 2016, pp. 1–7.

[93] Z. Zhong, Z. Tang, X. Li, T. Yuan, Y. Yang, W. Meng, Y. Zhang, R. Sheng, N. Grant, C. Ling, X. Huan, K. S. Kim, and S. Lee. (2018). *Xjtluindoorloc*. [Online]. Available: https://github.com/ZzhKlaus/2018-SURF-Trajectory-Estimation

[94] A. Moreira, M. J. Nicolau, I. Silva, J. Torres-Sospedra, C. Pendão, and F. Meneses, "Wi-Fi fingerprinting dataset with multiple simultaneous interfaces (1.0)," Zenodo, 2019, doi: 10.5281/Zenodo.3342526.

[95] J. Torres-Sospedra, A. Moreira, G. M. Mendoza-Silva, M. J. Nicolau, M. Matey-Sanz, I. Silva, J. Huerta, and C. Pend?o, "Exploiting different combinations of complementary sensor's data for fingerprint-based indoor positioning in industrial environments: Supplementary materials (1.0)," Zenodo, 2019, doi: 10.5281/zenodo.3333466.

[96] A. Moreira, I. Silva, F. Meneses, M. J. Nicolau, C. Pendao, and J. Torres-Sospedra, "Multiple simultaneous wi-fi measurements in fingerprinting indoor positioning," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Aug. 2017, pp. 1–8.

[97] J. Torres-Sospedra, A. Moreira, G. M. Mendoza-Silva, M. J. Nicolau, M. Matey-Sanz, I. Silva, J. Huerta, and C. Pendão, "Exploiting different combinations of complementary sensor's data for fingerprint-based indoor positioning in industrial environments," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Sep. 2019, pp. 1–8.

[98] X. Song, X. Fan, C. Xiang, Q. Ye, L. Liu, Z. Wang, X. He, N. Yang, and G. Fang. (2019). *Utsindoorloc*. [Online]. Available: https://github.com/XudongSong/UTSIndoorLoc-dataset

[99] P. Roy, C. Chowdhury, D. Ghosh, and S. Bandyopadhyay. (2019). *Juindoorloc*. [Online]. Available: https://drive.google.com/drive/folders/1_z1qhoRIcpineP9AHkfVGCfB2Fd_e-fD

[100] A. R. Jiménez, A. Perez-Navarro, A. Crivello, G. M. Mendoza-Silva, F. Seco, M. Ortiz, J. Perul, and J. Torres-Sospedra, "Datasets and supporting materials for the IPIN 2019 competition track 3 (smartphone-based, off-site) (1.0)," Zenodo, 2019, doi: 10.5281/zenodo.3606765.

[101] F. Potorti, S. Park, A. Crivello, F. Palumbo, M. Girolami, P. Barsocchi, S. Lee, J. Torres-Sospedra, A. R. J. Ruiz, and A. Perez-Navarro, "The IPIN 2019 indoor localisation competition—Description and results," *IEEE Access*, vol. 8, pp. 206674–206718, 2020.

[102] J. L. Salazar González, L. M. S. Morillo, J. A. Á. García, F. Enríquez, and A. R. J. Ruiz, "Energy-efficient indoor localization wifi-fingerprint dataset," *IEEE Dataport*, 2019, doi: 10.21227/49yg-5d21.

[103] J. L. S. González, L. M. S. Morillo, J. A. Álvarez-García, F. E. De Salamanca Ros, and A. R. J. Ruiz, "Energy-efficient indoor localization WiFi-fingerprint system: An experimental study," *IEEE Access*, vol. 7, pp. 162664–162682, 2019.

[104] J.-A. López-Pastor, A.-J. Ruiz-Ruiz, A.-J. García-Sánchez, and J.-L. Gómez-Tornero, "An automatized contextual marketing system based on a Wi-Fi indoor positioning system," *Sensors*, vol. 21, no. 10, p. 3495, May 2021.

[105] E. S. Lohan, "Additional tau datasets for Wi-Fi fingerprinting-based positioning (v1, 11.05.2020)," Zenodo, 2020, doi: 10.5281/zenodo.3819917.

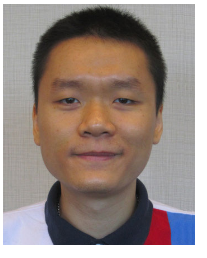
[106] J. Torres-Sospedra, D. Quezada-Gaibor, G. M. Mendoza-Silva, J. Nurmi, Y. Koucheryavy, and J. Huerta, "New cluster selection and fine-grained search for k-means clustering and Wi-Fi fingerprinting," in *Proc. Int. Conf. Localization GNSS (ICL-GNSS)*, 2020, pp. 1–6.

[107] A. Moreira, I. Silva, and J. Torres-Sospedra, "The DSI dataset for Wi-Fi fingerprinting using mobile devices (1.0)," Zenodo, 2020, doi: 10.5281/zenodo.3778646.

[108] P. Spachos. *RSSI Fingerprinting Dataset*. Accessed: Jun, 11, 2024. [Online]. Available: https://github.com/pspachos/RSSI-Dataset-for-Indoor-Localization-Fingerprinting

[109] S. Sadowski, P. Spachos, and K. N. Plataniotis, "Memoryless techniques and wireless technologies for indoor localization with the Internet of Things," *IEEE Internet Things J.*, vol. 7, no. 11, pp. 10996–11005, Nov. 2020.
[110] J. Torres-Sospedra, D. Q. Gaibor, A. R. Jiménez, A. Pérez-Navarro, and F. Seco, "Datasets and supporting materials for the IPIN 2020 competition track 3 (smartphone-based, off-site) (1.0)," Tech. Rep., 2020, doi: 10.5281/zenodo.4314992.
[111] F. Potortì, "Off-line evaluation of indoor positioning systems in different scenarios: The experiences from IPIN 2020 competition," *IEEE Sensors J.*, vol. 22, no. 6, pp. 5011–5054, Mar. 2022.
[112] T. Ayabakan, "Wi-Fi RSSI indoor data set," Tech. Rep., 2021, doi: 10.6084/m9.figshare.13607540.v2.
[113] T. Ayabakan and F. Kerestecioglu, "RSSI-based indoor positioning via adaptive federated Kalman filter," *IEEE Sensors J.*, vol. 22, no. 6, pp. 5302–5308, Mar. 2022.
[114] M. V. de Wynckel and B. Signer, "Indoor positioning using the OpenHPS framework," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Nov. 2021, pp. 1–8.
[115] J. Torres-Sospedra, F. A. Polo, F. Parralejo, V. B. Parent, F. Alvarez, A. Pérez-Navarro, A. R. Jimenez, and F. Seco, "Datasets and supporting materials for the IPIN 2021 competition track 3 (smartphone-based, off-site) (1.0)," Tech. Rep., 2021, doi: 10.5281/zenodo.5948678.
[116] F. Potortì, A. Crivello, S. Lee, B. Vladimirov, S. Park, Y. Chen, L. Wang, R. Chen, F. Zhao, and Y. Zhuge, "Offsite evaluation of localization systems: Criteria, systems and results from IPIN 2021-22 competitions," *IEEE J. Indoor Seamless Positioning Navigat.*, vol. 2, pp. 92–129, 2024.
[117] I. Ashraf, S. Din, S. Hur, and Y. Park. *Wi-Mest*. [Online]. Available: https://github.com/ImAshRayan/Wi-MEST
[118] B. Yu, Q. Xu, and Y. Shu. (2021). *Indoor Location & Navigation*. [Online]. Available: https://kaggle.com/competitions/indoor-location-navigation
[119] I. Silva, C. Pendão, and A. Moreira, "Continuous long-term Wi-Fi fingerprinting dataset for indoor positioning (full version) (1.1.0)," Tech. Rep., 20220, doi: 10.5281/zenodo.6928554.
[120] I. Silva, C. Pendão, and A. Moreira, "Collection of a continuous long-term dataset for the evaluation of Wi-Fi-fingerprinting-based indoor positioning systems," *Sensors*, vol. 22, no. 22, p. 8585, Nov. 2022.
[121] M. Cominelli, F. Gringoli, and R. Lo Cigno, "AntiSense: Standard-compliant CSI obfuscation against unauthorized Wi-Fi sensing," *Comput. Commun.*, vol. 185, pp. 92–103, Mar. 2022.
[122] M. Nabati, S. A. Ghorashi, and R. Shahbazian, "Joint coordinate optimization in fingerprint-based indoor positioning," *IEEE Commun. Lett.*, vol. 25, no. 4, pp. 1192–1195, Apr. 2021.
[123] M. Nabati, S. A. Ghorashi, and R. Shahbazian. (2022). *WiFi-RSSI-Localization-Dataset*. [Online]. Available: https://github.com/m-nabati/WiFi-RSSI-Localization-Dataset
[124] M. Nabati and S. A. Ghorashi, "A real-time fingerprint-based indoor positioning using deep learning and preceding states," *Expert Syst. Appl.*, vol. 213, Mar. 2023, Art. no. 118889.
[125] J. Bi, Y. Wang, B. Yu, H. Cao, T. Shi, and L. Huang. (2022). *Sodindoorloc*. [Online]. Available: https://github.com/renwudao24/SODIndoorLoc
[126] L. Klus, R. Klus, E. S. Lohan, J. Nurmi, C. G. Canut, M. Valkama, J. Talvitie, S. Casteleyn, and J. Torres-Sospedra, "Supplementary materials for," tuji1 dataset: Multi-device dataset for indoor localization with high measurement density' (1.0)," Tech. Rep., 2023, doi: 10.5281/zenodo.7641701.
[127] L. Klus, R. Klus, E. S. Lohan, J. Nurmi, C. Granell, M. Valkama, J. Talvitie, S. Casteleyn, and J. Torres-Sospedra, "TUJI1 dataset: Multi-device dataset for indoor localization with high measurement density," *Data Brief*, vol. 54, Jun. 2024, Art. no. 110356.
[128] J. Strohmayer and M. Kampel, "WiFi CSI-based long-range person localization using directional antennas," in *Proc. 2nd Tiny Papers Track (ICLR)*, 2024.
[129] (2013). *European Organization For Nuclear Research and OpenAIRE*. Zenodo. [Online]. Available: https://www.zenodo.org/

XU FENG received the B.S. degree from Zhejiang University, Hangzhou, Zhejiang, China, in 2020. He is currently pursuing the Ph.D. degree in computer science with the Royal Holloway University of London, U.K. His research interest includes machine learning for WiFi technologies, with applications for indoor positioning.



KHUONG AN NGUYEN received the B.Sc. (Hons.) and Ph.D. degrees from Royal Holloway, University of London, U.K., and the M.Phil. (Cantab) degree from the University of Cambridge, U.K. He is currently a Senior Lecturer in machine learning at Royal Holloway, University of London. His research interests include machine learning theory and its applications in navigation and smart healthcare.

ZHIYUAN LUO received the Ph.D. degree from Heriot-Watt University. He is currently a Professor in machine learning with the Royal Holloway University of London. His research interests include machine learning, data analysis, networked systems, agent-based computing, and applications of these algorithms and techniques.