Datasets for Indoor Positioning with Single-AP Wi-Fi Fingerprinting

Ritesh Kumar[1], Joaquín Torres-Sospedra[2] and Vijay Kumar Chaurasiya[1]

[1]Information Technology, IIIT Allahabad, 211015 Prayagraj, India
[2]Centro ALGORITMI, Universidade do Minho, Campus de Azurem, 4800-058 Guimarães, Portugal

Abstract
With the advent of smartphones and wearable devices, indoor positioning systems are relevant to the ageing in place and in-home monitoring paradigms. In terms of positioning, Wi-Fi is a widely known and used technology for this purpose. While fingerprinting in large areas benefits from a large wireless network with several Access Point (AP) providing network access, the availability of Wi-Fi signals with full or partial Line-of-Sight (LOS) conditions at home is mostly limited to a single AP. This work presents two datasets collected in an urban flat and in a home-like laboratory to support single AP positioning. In contrast to other Wi-Fi datasets where the number of samples is limited in each reference point, the two datasets have several Received Signal Strengh (RSS) readings per reference and evaluation point in order to test advanced positioning methods based on Single-AP (S-AP) Wi-Fi fingerprinting.

Keywords
Single-AP positioning, Wi-Fi fingerprinting, Open datasets, *k*-NN, ESP32

1. Introduction

While Global Navigation Satellite Systems (GNSSs) and Regional Navigation Satellite Systems (RNSSs) are widely used outdoors to support navigation, logistics, and transportation, among other Location Based Services (LBS), there is still not any *de-facto* to bring accurate and reliable technology indoors for any environment and context [1, 2, 3].

Given the variety of indoor spaces from relatively small residential homes and urban flats to large shopping malls or factories, it is difficult to find a positioning technology meeting all requirements in all possible scenarios. While autonomous vehicle navigation needs low latency and high accuracy, LBS in a shopping area is much less demanding.

Among all the possible applications, a relevant one considers in-home monitoring in, for instance, flats in dense urban areas [4, 5, 6]. Among the several positioning technologies available for this purpose, the ones based on Wi-Fi are less intrusive and do not require installing additional infrastructure to support positioning. Some recent works are exploiting the information coming from a single W-Fi AP to provide reliable positioning considering not only the actual single measurements but also their variability among other advanced features [7, 8, 9].

WIPHAL 2023: Work-in-Progress in Hardware and Software for Location Computation, June 06–08, 2023, Castellon, Spain
Pro2017003@iiita.ac.in (R. Kumar); info@jtorr.es (J. Torres-Sospedra); vijayk@iiita.ac.in (V. K. Chaurasiya)
0000-0002-5557-3776 (R. Kumar); 0000-0003-4338-4334 (J. Torres-Sospedra); 0000-0002-6146-7128 (V. K. Chaurasiya)

While several datasets exist for RSS-based fingerprinting indoors [10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22] and outdoors [23, 24], none of them covers the particular case of in-home monitoring with a single AP and sparse reference data. As suggested by Saccomanno et al. [25], we would like to explore extracting reliable spatial knowledge from multiple measurements.

The main contributions of this paper include:

- Full description of two new datasets for S-AP Wi-Fi fingerprinting;
- Definition of baseline methods for both datasets;

The remaining of this work is organised as follows. Section 2 described the procedures to collect the datasets as well as the data format. Section 3 provides some examples of usage of the dataset, giving the baseline method for this dataset. Discussion and conclusions are provided in Section 4.

2. Database description

2.1. Locations of datasets

Data were collected in two environments: the urban flat (FLAT) and a research laboratory (LAB) environment. Both scenarios are partially seen in Fig. 1 (a) and (b).

2.2. Static Data Acquisition with ESP32

(a) (b) (c)

Figure 1: Data collection setup in (a) Urban Flat and (b) Laboratory with (c) an ESP32 board

Data were collected with an ESP32 board (see Fig. 1). In this case, the ESP32 development board was configured in station mode to receive Wi-Fi router RSS signals, which are then transmitted to a computer's serial port. The raw data is obtained on a computer via terminal software and stored in CSV files.

The data was collected using the ESP32 mounted on the tripod in the two environments. Only data coming from the router within the scenario were stored. The dataset was captured in four directions (Up, Down, Right, and Left) in each location in two distinct sets: Reference Points (RPs) and Testing Points (TPs).

In the dataset collection procedure, each RP' data values were recorded for 5 min in each direction, i.e., a total of 20 (5×4) min of data per point. The TP' were recorded for 2 min each, i.e., a total of 8 (2×4) min per point.

The two selected scenarios were complex and the number of points to survey was high. Therefore, data were recorded on alternate days, with some furniture items being shifted from one room to another. Two individuals' movements were recorded in reference and test data points. While the reference points were arbitrarily assigned symmetrically to capture the most critical information with the fewest possible effort in both scenarios, most of the test points were distributed in a grid distribution of ≈0.6 m and ≈0.75 m for the Urban Flat and Laboratory, respectively.

2.3. Description of files

We have collected two datasets, one considering an urban flat `FLAT` and one considering a laboratory setting `URBAN`. Each dataset consists of two `csv` files:
↬`DATASET_SingleRSSI_TRAIN.csv`: This file contains the reference data (or radio map) collected with the ESP32 to train the positioning models. Each row represents a new sample (fingerprint) where the position is provided in columns 1 and 2 and the third column contains the RSS value. The file structure is described in Fig. 1 with an excerpt of data included in the file `FLAT_SingleRSSI_TRAIN.csv`.

Table 1
Contents of `FLAT_SingleRSSI_TRAIN.csv`

x	*y*	*rss*
3.75	1.05	-49
3.75	1.05	-87
3.75	1.05	-52
3.75	1.05	-50
3.75	1.05	-89
3.75	1.05	-51
3.75	1.05	-51
3.75	1.05	-51
3.75	1.05	-51
3.75	1.05	-51
⋮	⋮	⋮
1.25	8	-48

↬`DATASET_SingleRSSI_TEST.csv`: This file contains the evaluation data collected statically in several evaluation points with the ESP32 to test the positioning models in a similar fashion than reference/training data was collected. Each row represents a new sample (fingerprint) where the position is provided in columns 1 and 2 and the third column contains the RSS value.

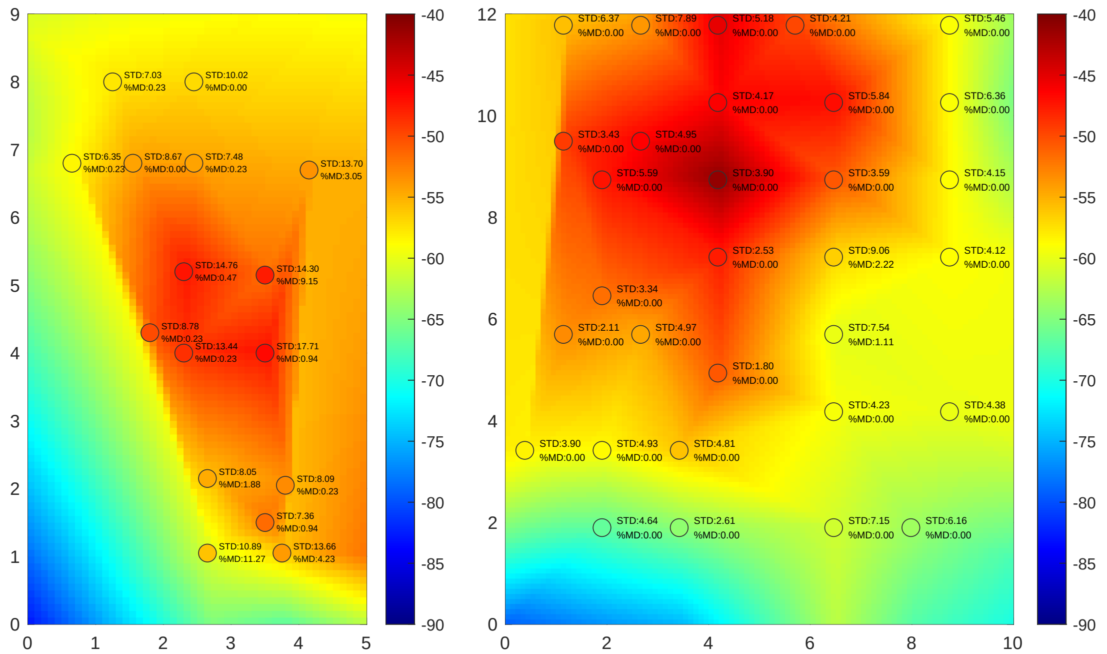
Table 2
Features of datasets FLAT and LAB

	Training/Reference Set				Test/Evaluation Set			
dataset	$\|\mathcal{T}\|$	$\|\mathcal{RP}\|$	Samples per $\mathcal{RP}$	$NA_{\mathcal{T}}$	$\|\mathcal{E}\|$	$\|\mathcal{TP}\|$	Samples per $\mathcal{RP}$	$NA_{\mathcal{E}}$
FLAT	8507	16	413.88±21.73	2.14%	10790	64	160.61±25.54	0.65%
LAB	6222	31	274.42±16.06	0.11%	10279	86	125.47±17.68	0.00%

2.4. Analysis of the datasets

The main features of the two datasets are included in Table 2, where the number of samples for training and testing ($\|\mathcal{T}\|$ and $\|\mathcal{E}\|$), the number of RPs ($\|\mathcal{RP}\|$), the number of TPs ($\|\mathcal{TP}\|$), as well as the average number of samples (measurements or S-AP fingerprints) per point and percentage of missing data/measurements ($NA_{\mathcal{T}}$ and $NA_{\mathcal{E}}$), are provided.

Fig. 2 provides an interpolation of the radio map (training/reference set) where some statistics (mean as point colour, the standard deviation of the valid RSS measurements and the percentage of missing data) are provided for each the RP in the two environments. The extrapolation used is a Gaussian Process Regression with a squared exponential kernel function. This figure is included to show difficulty in positioning just with a single RSS measurement.



(a) (b)

Figure 2: Interpolated radio map and extended information for (a) FLAT and (b) LAB datasets

3. Examples of usage

3.1. The k-NN algorithm

Wi-Fi fingerprinting based on the k-Nearest Neighbors (k-NN) algorithm was introduced by Bahl and Padmanabhan in 2000, since then it has been widely used because its simplicity and smooth integration with RSS values [26].

k-NN is a generic non-parametric supervised learning method introduced in 1951 [27], which is not only devoted to indoor positioning but also to gene selection [28], text classification [29], image classification [30], EEG classification [31], among others. k-NN may be implemented as a data classifier or regressor, where the output is a class membership or the average of the values of k nearest neighbours, respectively.

For Wi-Fi fingerprinting, including S-AP, k-NN provides a position estimate by selecting the k closest (most similar) samples from the training dataset to the operational sample. Then, the geometric centroid (a simple average) is applied to the location of the set of k closest samples.

To compute the set of closest samples, a distance metric (e.g. *Euclidean Distance*) or a similarity metric (e.g. *Cosine Similarity*) is commonly used [32, 33, 34, 35]. As a fingerprint in S-AP Wi-Fi fingerprinting is just one value, most of these distance functions and similarity metrics are equivalent. For simplicity and compliance of our algorithms with other multi-AP datasets, we assume that *City Block distance* (i.e. the absolute difference between two values) is appropriate for our datasets.

3.2. Analysis of k in k-NN

As a use case, we analyse the impact of the value of k in k-NN for the two datasets. As mentioned before, we assume a *City Block distance* to compare the two values (i.e. the absolute difference between two RSS values). The mean positioning error (i.e., the Euclidean distance between the current and estimated positions) for $k \in [1, \ldots, 100]$ in both datasets are provided in Fig. 3.

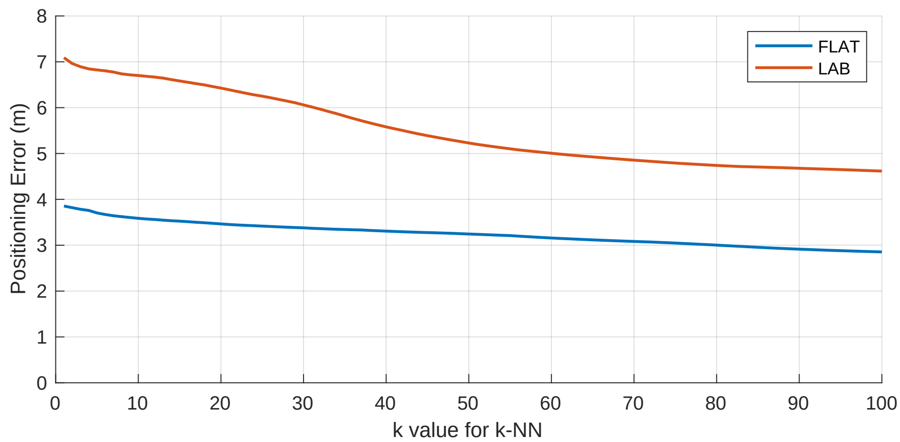


Figure 3: Analysis of k for k-NN with the introduced datasets

According to Fig. 3, the mean positioning error is significantly reduced over the first iterations.

Table 3
Positioning error (m) for both datasets using *k*-NN with $k = 100$

dataset	*mean*	*25th perc.*	*50th perc.*	*75th perc.*	*90th perc.*	*95th perc.*	*99th perc.*	*max.*
FLAT	2.85	1.43	2.49	4.00	5.52	6.68	7.62	8.07
LAB	4.62	2.99	4.48	6.10	7.55	8.51	9.53	12.14

Thus, Table 3 provides advanced error statistics (see [36, 37, 38]) on both datasets with $k = 100$, which can serve as a baseline for new developments. In addition, we provide the performance of the individual positioning errors as a CDF plot for $k = 1$ and $k = 100$ in both datasets (see Fig 4).

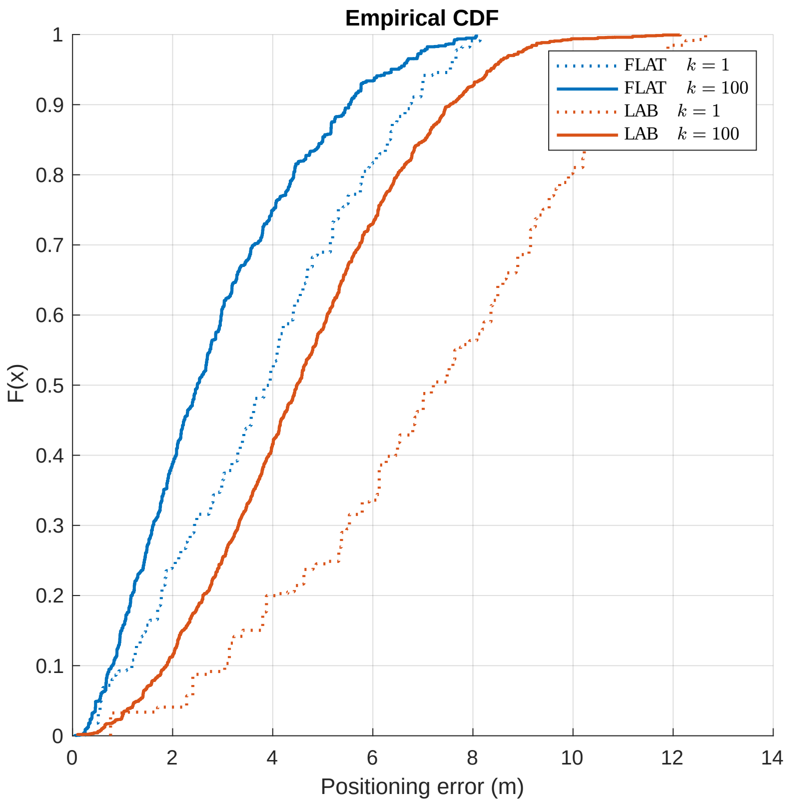


Figure 4: CDF plots of the individual positioning errors for *k*-NN with $k = 1$ and $k = 100$ in both datasets

Given the challenge of positioning with a single measurement from a single AP, the results obtained with mean errors of 2.58 m and 4.62 m introduce a challenge for further research.

4. Dicussion and Conclusions

Single-AP (S-AP) is gaining popularity given that most urban homes are not extensive and are covered by only one AP that has partial LOS in several places and, therefore, can be trusted.

Most of the existing Wi-Fi datasets either cover large environments (with hundreds of APs) or they do not resemble urban/residential medium-size flats.

This paper introduces two new datasets collected in two different places (an urban flat and a medium-sized laboratory) to fill this gap in the literature. Despite the dimensions of the operational areas in both scenarios and the information is limited to just one single AP, the simple baseline method based on k-NN provides an accuracy between 2.85 m to 4.62 m with $k = 100$.

Acknowledgments

R. Kumar and V. Kumar Chaurasiya acknowledges funding from the Ministry of Education, GoI, that has supported this work.

References

[1] C. Laoudias, A. Moreira, S. Kim, S. Lee, L. Wirola, C. Fischione, A survey of enabling technologies for network localization, tracking, and navigation, IEEE Communications Surveys and Tutorials 20 (2018) 3607–3644. doi:10.1109/COMST.2018.2855063.

[2] G. M. Mendoza-Silva, J. Torres-Sospedra, J. Huerta, A meta-review of indoor positioning systems, Sensors 19 (2019) 4507.

[3] P. S. Farahsari, A. Farahzadi, J. Rezazadeh, A. Bagheri, A survey on indoor positioning systems for iot-based applications, IEEE Internet of Things Journal 9 (2022) 7680–7699. doi:10.1109/JIOT.2022.3149048.

[4] J. Torres, O. Belmonte, R. Montoliu, S. Trilles, A. Calia, How feasible is wifi fingerprint-based indoor positioning for in-home monitoring?, in: Int. Conf. on Intelligent Environments (IE), 2016, pp. 68–75. doi:10.1109/IE.2016.19.

[5] R. Montoliu, E. Sansano-Sansano, A. Gascó, O. Belmonte, A. Caballer, Senior monitoring: a real case of applying a wifi fingerprinting-based indoor positioning method for people monitoring, in: Short Paper Proceedings of the Tenth Int. Conf. on Indoor Positioning and Indoor Navigation, volume 2498 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2019, pp. 95–102. URL: https://ceur-ws.org/Vol-2498/short13.pdf.

[6] A. R. Kabbinale, A. Bansal, K. S. Gopalan, Framework to ascertain effectiveness of ultra-wideband for in-home device control, in: 2022 IEEE Wireless Communications and Networking Conference (WCNC), 2022, pp. 1713–1718. doi:10.1109/WCNC51071.2022.9771747.

[7] L. Zhang, E. Ding, Y. Hu, Y. Liu, A novel csi-based fingerprinting for localization with a single AP, EURASIP J. Wirel. Commun. Netw. 2019 (2019) 51. URL: https://doi.org/10.1186/s13638-019-1371-y. doi:10.1186/s13638-019-1371-y.

[8] I. Fahmy, S. Ayman, H. Rizk, M. Youssef, Monofi: Efficient indoor localization based on single radio source and minimal fingerprinting, in: Int. Conf. on Advances in Geographic Information Systems, 2021, pp. 674–675. URL: https://doi.org/10.1145/3474717.3486808.

[9] Y. Wu, R. Chen, W. Fu, W. Li, H. Zhou, G. Guo, Indoor positioning based on tightly coupling of pdr and one single wi-fi ftm ap, Geo-spatial Information Science 0 (2022) 1–16. URL: https://doi.org/10.1080/10095020.2022.2072776.

[10] T. King, S. Kopf, T. Haenselmann, C. Lubberger, W. Effelsberg, CRAWDAD dataset mannheim/compass, 2008. URL: https://crawdad.org/mannheim/compass/20080411.

[11] S. Shrestha, J. Talvitie, E. S. Lohan, Deconvolution-based indoor localization with WLAN signals and unknown access point locations, 2013. URL: http://www.cs.tut.fi/tlt/pos/MEASUREMENTS_WLAN_FOR_WEB.zip.

[12] C. Laoudias, R. Piché, C. Panayiotou, Kios wifi rss dataset, 2013. https://www.researchgate.net/profile/Christos-Laoudias/publication/256482916_KIOS_WiFi_RSS_dataset.

[13] J. Torres-Sospedra, R. Montoliu, A. Martínez-Usó, J. P. Avariento, T. J. Arnau, M. Benedito-Bordonau, J. Huerta, UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems, in: 2014 Int. Conf. on Indoor Positioning and Indoor Navigation, 2014, pp. 261–270. doi:10.1109/IPIN.2014.7275492.

[14] A. Razavi, M. Valkama, E.-S. Lohan, K-means fingerprint clustering for low-complexity floor estimation in indoor mobile localization, in: IEEE Globecom Workshops, 2015.

[15] A. Cramariuc, H. Huttunen, E. S. Lohan, Clustering benefits in mobile-centric WiFi positioning in multi-floor buildings, in: 2016 Int. Conf. on Localization and GNSS, 2016.

[16] P. Richter, E. S. Lohan, J. Talvitie, WLAN (WiFi) rss database for fingerprinting positioning, Zenodo (2018). URL: https://zenodo.org/record/1161525.

[17] X. Song, X. Fan, C. Xiang, Q. Ye, L. Liu, Z. Wang, X. He, N. Yang, G. Fang, A novel convolutional neural network based indoor localization framework with wifi fingerprinting, IEEE Access 7 (2019) 110698–110709. doi:10.1109/ACCESS.2019.2933921.

[18] F. Alhomayani, M. H. Mahoor, Outfin, a multi-device and multi-modal dataset for outdoor localization based on the fingerprinting approach, Scientific Data 8 (2021).

[19] J. Bi, Y. Wang, B. Yu, H. Cao, T. Shi, L. Huang, Supplementary open dataset for wifi indoor localization based on received signal strength, Satellite Navigation 3 (2022) 25. URL: https://doi.org/10.1186/s43020-022-00086-y. doi:10.1186/s43020-022-00086-y.

[20] J. L. Salazar González, L. M. Soria Morillo, J. A. Álvarez García, F. Enríquez De Salamanca Ros, A. R. Jiménez Ruiz, Energy-efficient indoor localization wifi-fingerprint system: An experimental study, IEEE Access 7 (2019). doi:10.1109/ACCESS.2019.2952221.

[21] A. N. Nor Hisham, Y. H. Ng, C. K. Tan, D. Chieng, Hybrid wi-fi and ble fingerprinting dataset for multi-floor indoor environments with different layouts, Data 7 (2022). URL: https://www.mdpi.com/2306-5729/7/11/156. doi:10.3390/data7110156.

[22] E. S. Lohan, J. Torres-Sospedra, A. Gonzalez, WiFi RSS measurements in Tampere University multi-building campus, 2017, Zenodo (2021). URL: https://doi.org/10.5281/zenodo.5174851.

[23] M. Aernouts, R. Berkvens, K. Van Vlaenderen, M. Weyn, Sigfox and lorawan datasets for fingerprint localization in large urban and rural areas, Data 3 (2018). URL: https://www.mdpi.com/2306-5729/3/2/13. doi:10.3390/data3020013.

[24] P. Masek, M. Stusek, E. Svertoka, J. Pospisil, R. Burget, E. S. Lohan, I. Marghescu, J. Hosek, A. Ometov, Measurements of lorawan technology in urban scenarios: A data descriptor,

Data 6 (2021). URL: https://www.mdpi.com/2306-5729/6/6/62. doi:10.3390/data6060062.
[25] N. Saccomanno, A. Brunello, A. Montanari, What you sense is not where you are: on the relationships between fingerprints and spatial knowledge in indoor positioning, IEEE Sensors Journal 22 (2022) 4951–4961. doi:10.1109/JSEN.2021.3070098.
[26] P. Bahl, V. Padmanabhan, RADAR: an in-building RF-based user location and tracking system, in: Proceedings IEEE INFOCOM 2000, volume 2, 2000, pp. 775–784 vol.2. doi:10.1109/INFCOM.2000.832252.
[27] E. Fix, J. L. Hodges, Discriminatory analysis. nonparametric discrimination: Consistency properties, International Statistical Review 57 (1989) 238–247. URL: http://www.jstor.org/stable/1403797.
[28] C.-P. Lee, W.-S. Lin, Y.-M. Chen, B.-J. Kuo, Gene selection and sample classification on microarray data based on adaptive genetic algorithm/k-nearest neighbor method, Expert Systems with Applications 38 (2011) 4661–4667. URL: https://www.sciencedirect.com/science/article/pii/S0957417410006810. doi:https://doi.org/10.1016/j.eswa.2010.07.053.
[29] S. Tan, An effective refinement strategy for knn text classifier, Expert Systems with Applications 30 (2006) 290–298. URL: https://www.sciencedirect.com/science/article/pii/S0957417405001570. doi:https://doi.org/10.1016/j.eswa.2005.07.019.
[30] K. Huang, S. Li, X. Kang, L. Fang, Spectral–spatial hyperspectral image classification based on knn, Sensing and Imaging 17 (2015) 1. URL: https://doi.org/10.1007/s11220-015-0126-z. doi:10.1007/s11220-015-0126-z.
[31] M. N. A. H. Sha'abani, N. Fuad, N. Jamal, M. F. Ismail, kNN and SVM Classification for EEG: A Review, in: InECCE2019, Springer, Singapore, 2020, pp. 555–565.
[32] G. Caso, L. De Nardis, M.-G. Di Benedetto, A mixed approach to similarity metric selection in affinity propagation-based wifi fingerprinting indoor positioning, Sensors 15 (2015) 27692–27720.
[33] J. Torres-Sospedra, R. Montoliu, S. Trilles, Ó. Belmonte, J. Huerta, Comprehensive analysis of distance and similarity measures for wi-fi fingerprinting indoor positioning systems, Expert Systems with Applications 42 (2015) 9263–9278.
[34] G. Minaev, A. Visa, R. Piché, Comprehensive survey of similarity measures for ranked based location fingerprinting algorithm, in: Int. Conf. on Indoor Positioning and Indoor Navigation (IPIN), IEEE, 2017, pp. 1–4.
[35] A. Brunello, A. Montanari, N. Saccomanno, A genetic programming approach to wifi fingerprint meta-distance learning, Pervasive and Mobile Computing 85 (2022) 101681.
[36] ISO, Real time locating systems – Test and evaluation of localization and tracking systems (ISO1835), 2016. URL: https://www.iso.org/standard/62090.html.
[37] F. Potortì, A. Crivello, P. Barsocchi, F. Palumbo, Evaluation of indoor localisation systems: Comments on the ISO/IEC 18305 standard, in: Int. Conf. on Indoor Positioning and Indoor Navigation (IPIN), IEEE, 2018, pp. 1–7.
[38] F. Potortì, J. Torres-Sospedra, D. Quezada-Gaibor, A. R. Jiménez *et al.*, Off-line evaluation of indoor positioning systems in different scenarios: The experiences from ipin 2020 competition, IEEE Sensors Journal 22 (2022) 5011–5054. doi:10.1109/JSEN.2021.3083149.
[39] R. Kumar, J. Torres-Sospedra, V. Kumar Chaurasiya, Datasets for Indoor Positioning with Single-AP Wi-Fi Fingerprinting, Zenodo (2023). URL: https://doi.org/10.5281/zenodo.7949032.

A. List of acronyms

AP	Access Point
GNSS	Global Navigation Satellite System
k-**NN**	*k*-Nearest Neighbors
LBS	Location Based Services
LOS	Line-of-Sight
RNSS	Regional Navigation Satellite System
RP	Reference Point
RSS	Received Signal Strengh
S-AP	Single-AP
TP	Testing Point

B. Online Resources

The dataset and sources for generating the reports and figures are available in:

- Zenodo Package [39]