Jurnal Sosial dan Teknologi (SOSTECH)
Volume 6, Number 11, November 2026, pp. 3361-3377
e-ISSN 2774-5155
https://doi.org/10.59188/jurnalsostech.v6i11.33037

What A Random Partition Hides: Device Heterogeneity, Elapsed Time and The Deployment Accuracy of Wi-Fi Fingerprinting for Indoor Positioning

Sudrajat
Politeknik Siber Cerdika Internasional, Indonesia
Email: sudrajat@polteksci.ac.id

Keywords:
indoor positioning; Wi-Fi fingerprinting; device heterogeneity; evaluation protocol; received signal strength

This is an open access article under the CC BY-SA license.

Copyright © 2026 by Author. Published by
Publisher Publications

Abstract

Wi-Fi fingerprinting enables indoor positioning using existing wireless infrastructure; however, random data splits can overestimate positioning accuracy under real-world deployment conditions. This study examined how evaluation protocols, handset heterogeneity, and the time elapsed between surveys affect positioning accuracy and the relative performance of signal strength representations. Using the UJIIndoorLoc dataset, eight signal strength representations were compared using a weighted k-nearest neighbors (WKNN) estimator under three evaluation protocols: random partitioning, leave-one-device-out validation, and validation using data collected 91 days after the training survey ended. Familiar and unfamiliar handsets in the subsequent survey were compared after matching their distributions across buildings and floors. Random partitioning underestimated the mean planar positioning error by 4.82 m across representations because repeated measurements at identical locations allowed near-duplicate fingerprints to appear in both subsets. Raw signal strength measurements achieved the best performance under random partitioning, whereas exponential encoding performed best in the subsequent survey, yielding a mean planar positioning error of 6.45 m. After matching, unfamiliar handsets were associated with an additional mean positioning error of 2.59 m. Leave-one-device-out validation identified the same encoding as the subsequent survey and yielded a mean signed error difference of 0.05 m relative to the positioning errors observed with unfamiliar handsets. These findings support holding out entire handset models during validation to obtain more realistic estimates of positioning accuracy before deployment, particularly when data from only one survey campaign are available.

INTRODUCTION

An indoor positioning system based on Wi-Fi fingerprinting compares the signal strengths currently observed by a device with those recorded in a previously surveyed radio map and estimates the location whose recorded signal signature most closely matches the observation. This approach requires no additional infrastructure, which explains why it has remained viable despite numerous proposed alternatives and continues to underpin many deployed systems (Dai et al., 2023; Yaro et al., 2023; Zholamanov et al., 2025). Systematic reviews of smartphone-based localization and the evolution from conventional fingerprinting to machine learning-based estimators have reached similar conclusions (Hailu et al., 2024b; Martín-Frechina et al., 2025; Naser et al., 2023). Moreover, the field increasingly treats indoor

positioning as infrastructure to be specified and procured rather than merely demonstrated (Hailu et al., 2025; Perez-Navarro et al., 2022). Reported positioning accuracies are promising. Meter-level and submeter-level accuracy figures are frequently reported for public radio maps, and reviews of the field characterize the state of the art in these terms (Aziz & Koo, 2025; J. Li et al., 2024).

A building manager reviewing this literature might conclude that the positioning problem has largely been solved and that the remaining challenge is system integration. This study examines whether that conclusion holds under realistic deployment conditions and finds that it does not. The discrepancy does not arise from the estimators themselves, which are based on established methods. Instead, it becomes apparent when two conditions commonly left implicit in published evaluation protocols are explicitly considered: the device used for positioning may differ from those used to construct the radio map, and the radio environment encountered during positioning may differ from the environment surveyed three months earlier.

This study uses the UJIIndoorLoc dataset (Torres-Sospedra et al., 2014), selected because it is a widely used benchmark in indoor positioning research and, unusually, incorporates a temporal gap between training and validation data collection. The training dataset contains 19,861 fingerprints covering 465 access points, collected by 18 operators using 16 smartphone models over 21 days. The validation dataset contains 1,111 fingerprints collected 91 days after the training campaign ended, by a single operator using 11 smartphone models, nine of which were absent from the training dataset. Published studies that combine the two datasets and randomly partition the pooled observations eliminate this original separation. The central argument of this study is that such partitioning fundamentally changes what the evaluation measures.

Three strands of previous research are particularly relevant. The first concerns device heterogeneity: two smartphones placed side by side may report different received signal strength indicator (RSSI) values for the same access point because of differences in their radio-frequency front ends, antennas, and firmware. Consequently, a radio map constructed using one device may not accurately represent measurements obtained using another (Gidey, Guo, Li, & Zhang, 2022; Khan et al., 2023). A conventional approach to mitigating this problem is to use signal representations that reduce sensitivity to device-dependent gain differences, either by calculating differences between signal strength measurements (Hossain et al., 2013; Kjærgaard & Munk, 2008) or by discarding absolute magnitudes and retaining only the relative ranking of detected access points (Machaj et al., 2011). The second strand concerns signal representation independently of device heterogeneity. Monotonic transformations of signal strength values expressed in decibels can alter nearest-neighbor selection, and their effectiveness depends on the distance metric employed (C. Li et al., 2026; Torres-Sospedra et al., 2015). The third strand concerns the temporal stability of radio maps. Access points may be relocated, replaced, or reconfigured, causing positioning accuracy to deteriorate as the radio environment changes over time (Mendoza-Silva et al., 2018; Wan et al., 2022). These challenges have motivated the development of adaptive long-term positioning models and landmark-based approaches for updating radio maps as indoor environments evolve (Chia et al., 2025; Hailu et al., 2024a; Y. Zhu & Cheng, 2025).

To our knowledge, previous research has not systematically examined the effects of evaluation protocols and signal strength representations while holding the positioning estimator constant and distinguishing device-related degradation from temporal degradation within the same dataset. Addressing this gap constitutes the principal contribution of the present study, which comprises four components.

First, this study quantifies the extent to which random partitioning overestimates positioning accuracy. Eight signal strength representations are evaluated using three protocols: random partitioning of the training dataset, leave-one-device-out validation within the training dataset, and evaluation using the subsequent validation campaign. Random partitioning not only produces lower positioning errors but also reverses the relative ranking of signal representations, potentially leading practitioners to select a representation that performs poorly under deployment conditions.

Second, this study identifies the mechanism underlying this discrepancy. The training dataset contains only 933 distinct surveyed locations represented by 19,861 fingerprints. Consequently, random partitioning frequently places near-duplicate fingerprints from the same surveyed locations in both the reference and test subsets. Under these conditions, the evaluation primarily measures the retrieval of highly similar fingerprints rather than the ability to generalize positioning estimates to independent observations.

Third, this study distinguishes the effects of device novelty from those associated with elapsed time. Two smartphone models appear in both data collection campaigns. Validation fingerprints collected using these familiar models share the same temporal interval and operator as those collected using unfamiliar models, while differing in whether the smartphone model was represented during training. This structure provides a natural comparison for examining device-related differences under a common temporal setting. After matching the two groups by building and floor, unfamiliar devices exhibit an additional mean positioning error of 2.59 m.

Fourth, this study investigates whether deployment positioning errors can be estimated before implementation. Leave-one-device-out validation within a single survey campaign requires neither a subsequent data collection campaign nor an additional waiting period. The results indicate that this protocol estimates the positioning error subsequently observed on unfamiliar devices with a mean signed difference of 0.05 m, suggesting its potential value as a practical predeployment evaluation strategy.

The remainder of this paper is organized as follows. Section 2 describes the dataset, signal strength representations, positioning estimator, three evaluation protocols, and the approach used to distinguish device-related and temporal effects. Section 3 presents the results and discusses their implications for practical deployment and the interpretation of benchmark evaluations. Section 4 presents the conclusions.

METHOD

Database and its two campaigns

The UJIIndoorLoc dataset contained received signal strength measurements from up to 520 wireless access points for each fingerprint, along with building identifiers, floor levels, planar coordinates, operator identifiers, and handset identifiers (Torres-Sospedra et al., 2014).

Undetected access points were assigned a sentinel value, which was recoded to −104 dBm, one decibel below the weakest signal observed during the training campaign, to ensure that undetected signals were represented as weaker than detected signals.

Two preprocessing steps were performed using only the training data. First, access points that were not detected in any training fingerprint were excluded, removing 55 of the 520 columns and retaining 465. This prevented information from access points detected exclusively during the subsequent validation campaign from influencing distance calculations and introducing data leakage. Second, fingerprints without any detected access points were excluded because they contained no usable positioning information. This removed 76 training fingerprints, while all validation fingerprints were retained. Table 1 summarizes the resulting dataset. The construction and reuse of multi-building fingerprint datasets, including conventions for representing undetected access points, have been discussed in previous studies (S. Li et al., 2024; Nor Hisham et al., 2022).

Table 1. The Two Campaigns After Conditioning

Property	Training campaign	Validation campaign
Fingerprints	19,861	1,111
Access points retained	465	465
Operators	18	1
Phone models	16	11
Models also in training	–	2
Fingerprints on familiar models	–	397
Fingerprints on unfamiliar models	–	714
Collection window (days)	21	19
Access points per fingerprint	18.1	15.6
Distinct surveyed points	933	–

The validation campaign begins 91 days after the training campaign ends. Nine of its eleven phone models never appear in training.

Representations of signal strength

Let r be the reading for one access point in one fingerprint, in decibels relative to a milliwatt, and let the weakest reading anywhere in the training campaign be written as the floor. Eight representations are compared. The raw encoding passes the decibel values through unchanged. The positive encoding shifts them so that the floor becomes zero, which is the usual preparation for a Euclidean distance. The powed and exponential encodings apply a monotone rescaling on top of that, the former raising the shifted value to a fixed power and the latter passing it through an exponential, both normalised to the unit interval; their purpose is to stop a large number of weak readings dominating a distance that a few strong ones should decide (Torres-Sospedra et al., 2015). A parallel line of work instead selects or reweights the access points before any distance is taken (R. Jin et al., 2026; C. Li et al., 2026), or reconstructs the readings a fingerprint is missing (Wang et al., 2026); we hold the access-point set fixed so that the encoding is the only thing that varies.

The remaining four are calibration-free, meaning that a gain the receiving device applies to every access point alike cancels before the distance is taken. Two of them difference the readings against a statistic of the same fingerprint, either the mean of its detected readings or its strongest one, which is the signal-strength-difference idea (Hossain et al., 2013; Kjærgaard & Munk, 2008), an approach whose appeal is that it needs no per-device calibration campaign

(Gidey, Guo, Zhong, et al., 2022; Yang et al., 2022). One discards the magnitudes entirely and keeps the rank order of the strongest 12 access points (Machaj et al., 2011). The last keeps only which access points answered at all.

Each difference is then lifted by a constant so that every detected access point scores strictly above zero. This matters more than it appears: without the lift, the reference access point scores exactly zero and becomes indistinguishable from one that answered nothing, which penalises the two differencing representations for a coding accident rather than for anything about the signal. In a preliminary run without the lift their error was inflated by roughly nine metres, and we report that here because the same accident is easy to commit and invisible in a summary statistic.

Estimator

All representations are scored with the same estimator, a weighted k-nearest-neighbour rule over the radio map, which remains the reference method in this field and the one against which learned alternatives are measured (Park et al., 2024; Youssef & Agrawala, 2005). Neighbours are weighted by the reciprocal of their distance. The planar estimate is the weighted mean of the neighbours' coordinates; the building and floor estimates are the weighted modes of their labels. Holding the estimator fixed is deliberate. A comparison in which the representation and the learner vary together cannot attribute a difference to either, and the question here is what the protocol and the representation do, not which learner is strongest (Wu et al., 2024; H. Zhu et al., 2023). Ensembles, attention models, graph and variational formulations all report gains over this baseline (Esmaeili Gorjan & Gil Jiménez, 2024; S. Jin & Kim, 2024; Kim et al., 2025; Mao et al., 2025; Rana & Dulal, 2025), and nothing here contradicts them; the point is that whatever estimator is chosen must be scored under a protocol that reflects deployment.

Choosing configurations without looking at the deployment

Each representation needs a distance function and a neighbourhood size. Three distances are considered, Euclidean, Manhattan and Sørensen, the last restricted to the representations that are non-negative by construction because its denominator is unstable otherwise. The neighbourhood size is drawn from 1, 3, 5, 7, 11 and 15.

Both are chosen on a device-disjoint split inside the training campaign: 5 phone models are held out, the remaining 11 are used to fit, and the pair minimising mean planar error on the held-out models is retained. The choices are then frozen. Neither the later campaign nor the full leave-one-device-out procedure informs them, so the deployment figures reported below are not the best of many attempts.

Three evaluation protocols

P1, the random partition, draws a stratified twenty per cent of the training campaign as a query set and fits the radio map on the rest, repeated five times with different seeds. This mirrors what is commonly reported on this database. P2, leave-one-device-out, holds out every fingerprint taken with one phone model in turn and fits on the other fifteen, so the query device is always unfamiliar while the survey period is unchanged. P3 fits on the entire training campaign and queries with the later validation campaign, in which both the device population and the period differ. Only P3 asks the question an operator asks, and a minority of studies on this database do report it (Neupane et al., 2026).

Separating the device from the interval

P3 confounds two changes. The decomposition exploits the fact that 2 phone models appear in both campaigns. The 397 validation fingerprints taken with them share the 91 day interval and the single operator with the other 714, and differ from them only in that the querying hardware is already represented in the radio map. The difference between the two groups therefore isolates device novelty, and the difference between the familiar group and the random partition isolates everything else.

The two groups were not formed by randomisation, so their composition is checked rather than assumed. They are indistinguishable in their distribution across the three buildings but not across floors, and the unfamiliar group detects more access points per fingerprint (17.0 against 13.1) and reports a stronger best reading (-55.9 against -58.6 dBm). Both of those differences favour the unfamiliar group, so any penalty we attribute to it is conservative. To remove the floor imbalance we additionally report a post-stratified estimate in which each unfamiliar fingerprint is weighted so that the building-by-floor composition of its group matches the familiar group; 12 cells are common to both and 24 fingerprints fall outside them. Intervals throughout are percentile bootstrap intervals from 2,000 resamples.

What is reported

Mean planar error is the headline, but a positioning system is used through a floor plan, so three further quantities are reported: the proportion of queries placed in the correct building, the proportion placed on the correct floor of the correct building, and the proportion that are both on the correct floor and within 10 m, which we call a usable fix. The last of these is the one an application can act on, and it behaves differently from the mean because a floor error is not a large planar error but a categorical one (Schyga et al., 2022).

RESULTS AND DISCUSSION

The protocol decides the answer

Table 2 reports all three protocols for all eight representations, and Fig. 1 shows the same figures. The raw decibel encoding scores 2.13 m under the random partition and 8.84 m on the later campaign, a factor of 4.2. Across the eight representations the random partition understates the deployment error by 4.82 m on average. Reviews that aggregate reported accuracies across studies seldom record which protocol produced each figure (Aziz & Koo, 2025; Hailu et al., 2024b; J. Li et al., 2024), so a gap of this size propagates into the summary tables practitioners consult.

Table 2. Mean planar error in metres under the three protocols

Representation	Distance	k	P1 random	P2 device-out	P3 deployment	P3/P1
Raw dBm	euclidean	3	2.13	10.03	8.84	4.2
Positive	sorensen	15	3.03	7.77	7.22	2.4
Powed	sorensen	11	2.99	8.36	6.56	2.2
Exponential	sorensen	15	3.09	7.70	6.45	2.1
Rank	sorensen	5	3.49	7.95	7.09	2.0
SSD (mean)	sorensen	15	4.25	9.46	9.67	2.3
SSD (max)	sorensen	11	3.75	9.09	9.16	2.4
Binary visibility	sorensen	15	4.89	10.76	11.20	2.3

Distance and k were chosen on a device-disjoint split inside the training campaign and then frozen. P1 is averaged over five seeded partitions; its standard deviation across repeats never exceeds 0.1 m.

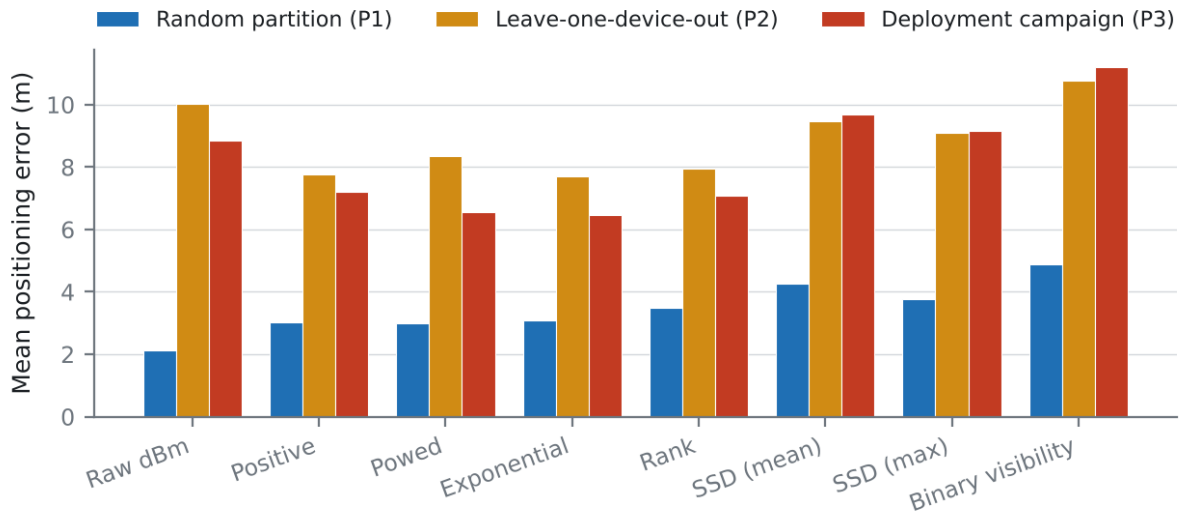

Fig. 1. Mean planar error under the three protocols. The random partition is not merely Optimistic: it orders the representations differently from the deployment campaign.

The more damaging finding is not the magnitude but the ordering. Under the random partition the best representation is raw dbm, by a clear margin. On the deployment campaign that representation is fifth of eight, and the best is exponential. A team that compared designs under a random partition would therefore not simply overestimate its accuracy; it would select the wrong design and then overestimate it. Leave-one-device-out, by contrast, selects the same representation the deployment campaign does.

Why the random partition is so generous

The mechanism is visible in the error distribution. Fig. 2 shows that under the random partition 56.5 per cent of queries with the raw encoding are placed within half a metre of the truth, and the median error is zero. On the deployment campaign that share falls to 3.5 per cent.

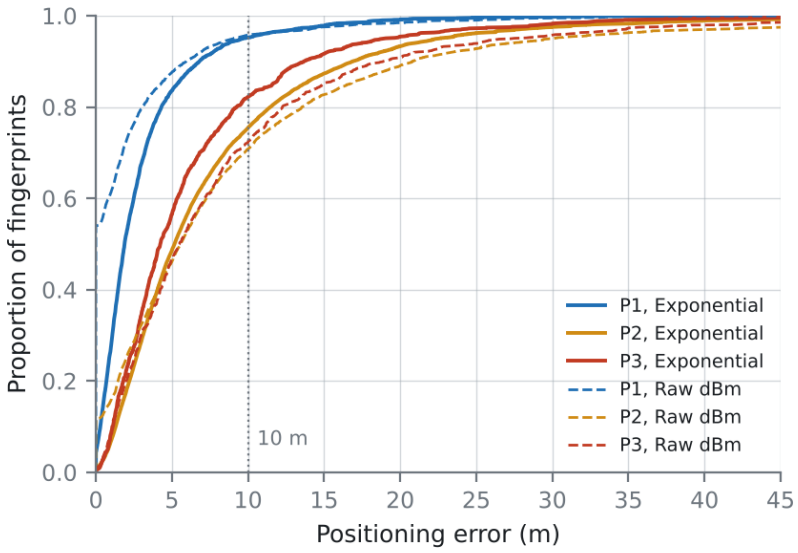

Fig. 2. Cumulative distribution of planar error for the exponential and raw encodings under each protocol. Under the random partition the raw encoding places most queries exactly on the truth, which is duplicate retrieval rather than positioning.

The explanation is in the survey design rather than in the models. The training campaign covers only 933 distinct surveyed points, so the 19,861 fingerprints average 21.3 per point and every one of them shares its coordinates with at least one other. A random partition therefore almost always leaves, in the reference set, a fingerprint taken at the same point minutes earlier with the same handset. Recovering it is a retrieval task, not a positioning task, and no amount of repetition or cross-validation within that partition will reveal the difference. This is a concrete instance of the general warning that evaluation partitions must respect the structure by which the data were generated (Schyga et al., 2022; Silva et al., 2022). How densely a campaign revisits each point is a decision of the survey protocol and differs between published databases (S. Li et al., 2024; Nor Hisham et al., 2022), so the size of this effect is not a constant of the method.

What the deployment campaign leaves

Table 3 reports the operational quantities on the deployment campaign. The best representation, exponential, places 99.9 per cent of queries in the correct building and 94.8 per cent on the correct floor of it, with a median planar error of 4.24 m and a mean of 6.45 m (95 per cent interval 6.04 to 6.90). The ninety-fifth percentile is 18.95 m. Usable fixes, meaning the correct floor and within 10 m, account for 79.7 per cent.

Table 3. Operational performance on the deployment campaign

Representation	Building (%)	Floor (%)	Median (m)	Mean (m)	95th pct (m)	Usable fix (%)
Raw dBm	99.7	89.6	5.49	8.84	27.39	68.1
Positive	99.9	92.4	4.78	7.22	22.01	74.3
Powed	100.0	95.2	4.41	6.56	19.86	78.3
Exponential	99.9	94.8	4.24	6.45	18.95	79.7
Rank	99.9	94.8	4.90	7.09	19.38	74.3
SSD (mean)	99.9	86.8	7.20	9.67	28.25	59.9
SSD (max)	99.9	88.4	6.71	9.16	26.13	62.7
Binary visibility	99.9	80.1	8.54	11.20	31.30	50.5

A usable fix is a query placed on the correct floor of the correct building and within 10 m. Percentages are of all 1,111 validation fingerprints.

The gap between the floor hit rate and the usable-fix rate is worth noting, because the two are often conflated. A system can place almost every query on the right floor and still fail to put a usable proportion of them near enough to a door to be acted on, and which of the two matters depends entirely on the application, a point that reviews of fielded systems make repeatedly (Hailu et al., 2025; Naser et al., 2023; Perez-Navarro et al., 2022).

The device and the interval, separated

Fig. 3 decomposes the deployment error for each representation into the random-partition error, the part added by the interval and the operator, and the part added by unfamiliar hardware; Table 4 gives the numbers. The two causes are not of equal size, and their balance depends on the representation. For the raw encoding almost all the degradation is attributable to the interval and the operator (6.39 m) and almost none to the device (0.50 m). For exponential

the device accounts for 2.47 m against 1.77 m for the interval. Across the eight, the device accounts for a median 44 per cent of the total.

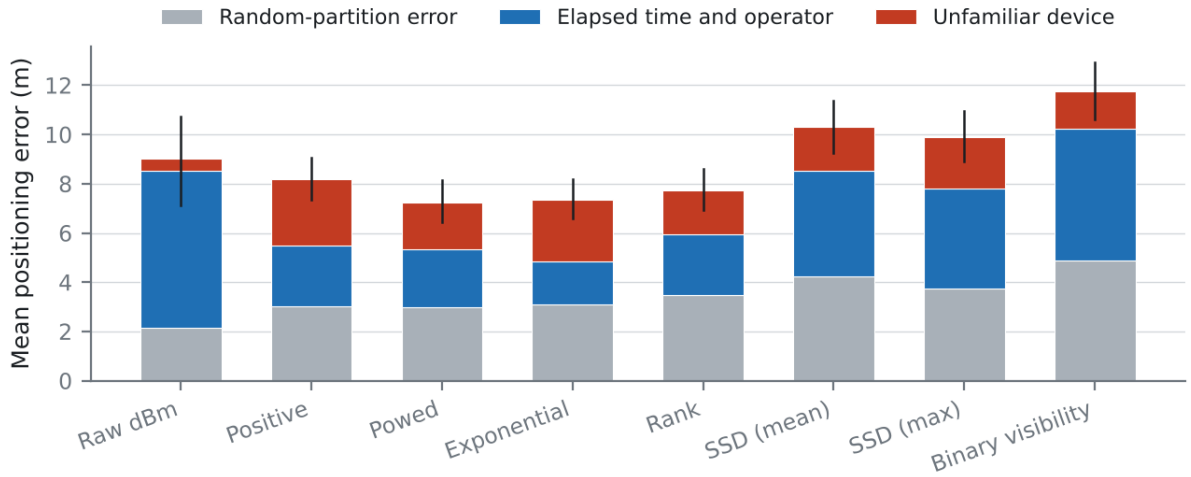

Fig. 3. Decomposition of deployment error. Each bar reaches the error observed on unfamiliar devices. Whiskers are bootstrap intervals on the device component.

Table 4. Device novelty and elapsed time, in metres of mean planar error

Representation	Calibration-free	P1 random	P3 familiar	P3 unfamiliar	Interval	Device [95% CI]
Raw dBm	no	2.13	8.52	9.02	6.39	0.50 [-1.41, 2.20]
Positive	no	3.03	5.49	8.18	2.47	2.68 [1.86, 3.55]
Powed	no	2.99	5.32	7.25	2.34	1.93 [1.12, 2.82]
Exponential	no	3.09	4.86	7.33	1.77	2.47 [1.72, 3.31]
Rank	yes	3.49	5.95	7.73	2.46	1.78 [0.97, 2.66]
SSD (mean)	yes	4.25	8.52	10.31	4.27	1.79 [0.74, 2.86]
SSD (max)	yes	3.75	7.82	9.91	4.06	2.09 [1.07, 3.14]
Binary visibility	yes	4.89	10.25	11.74	5.35	1.49 [0.35, 2.69]

Familiar means the query device also appears in the training campaign. Both validation groups share the interval and the operator, so their difference isolates device novelty.

Because the two validation groups were not randomised, the device column is recomputed with each unfamiliar fingerprint weighted so that the building-by-floor composition of its group matches the familiar group. Matching raises rather than lowers the estimate: the device penalty averages 2.59 m, ranging from 1.15 to 3.52 m, and the interval excludes zero for 7 of the eight representations. The exception is the raw encoding (1.15 m, -0.72 to 2.81), which is already so degraded by the interval that unfamiliar hardware adds little that is measurable. The magnitude is consistent with reports placing handset differences among the largest non-environmental error sources in fingerprinting (Gidey, Guo, Li, & Zhang, 2022; Khan et al., 2023).

This yields a result we did not expect. The four representations built to be calibration-free carry a device penalty of 2.55 m on average, against 2.62 m for the four that are not. On this database, in other words, differencing the readings or reducing them to ranks does not buy robustness to unfamiliar hardware, even though that is the property those representations were designed to provide (Hossain et al., 2013; Kjærgaard & Munk, 2008; Machaj et al., 2011). They also start from a worse position: the two differencing encodings and the binary one are the three

weakest in Table 2 under every protocol. We read this as a consequence of what the cancellation costs. Removing a per-fingerprint statistic removes real information about absolute signal level along with the device-specific part, and on a database where the surviving discriminative signal is modest, that trade is not worth making. A representation that is invariant to the device is of no use if it is also invariant to the location (Gidey, Guo, Zhong, et al., 2022; Yang et al., 2022).

The interval component, by contrast, is exactly what radio-map maintenance addresses: periodic resurveying, self-calibration, crowdsourced updating from landmarks, and estimators that adapt as the map ages (Chia et al., 2025; Hailu et al., 2024a; Wan et al., 2022; Y. Zhu & Cheng, 2025). The decomposition suggests where that effort pays. For an encoding whose degradation is mostly temporal, maintenance is the lever; for one whose degradation is mostly device-driven, resurveying will not help, and the remedy has to be sought in the representation or in broadening the device population of the survey. Databases with repeated campaigns spread over a longer period would let that distinction be traced directly rather than inferred from a single gap (Mendoza-Silva et al., 2018; Silva et al., 2022).

Estimating deployment error before deploying

The practical question a team faces is whether deployment accuracy can be estimated from a single survey campaign, before a second campaign exists. Fig. 4 answers it. Across the eight representations, leave-one-device-out error inside the training campaign tracks the error later observed on unfamiliar devices with a rank correlation of 0.76 and a product-moment correlation of 0.87, and it is close to unbiased: the mean signed gap is 0.05 m. Against the full deployment campaign, including its familiar devices, leave-one-device-out is mildly pessimistic, by 0.61 m. The random partition, by contrast, reaches a rank correlation of only 0.64. Reviews of this literature already note that evaluation practice is not standardised, which makes published accuracies difficult to compare across studies (Dai et al., 2023; Martín-Frechina et al., 2025; Yaro et al., 2023; Zholamanov et al., 2025); a protocol that is both cheap and close to unbiased is a practical answer to that complaint.

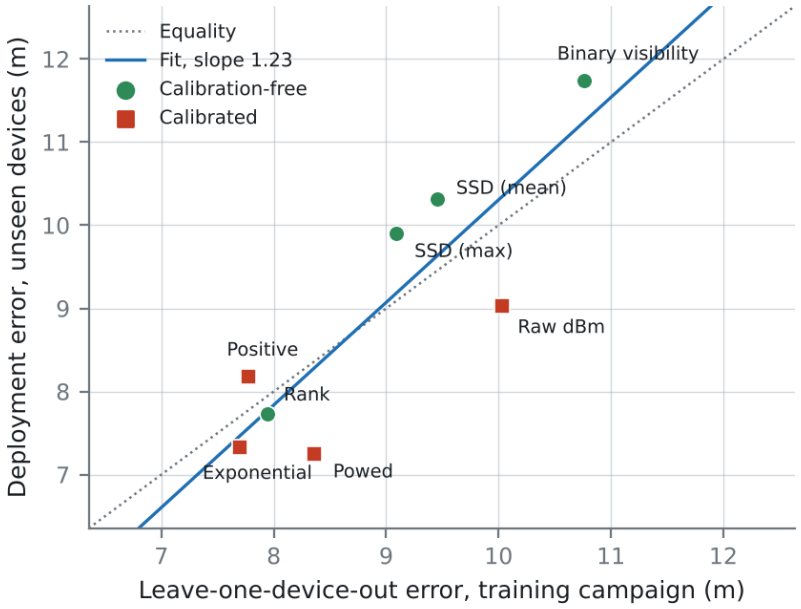

Fig. 4. Leave-one-device-out error inside the training campaign against the error observed on unfamiliar devices three months later. Points lie close to equality.

This is the most directly usable result in the paper. Leave-one-device-out costs nothing but bookkeeping: it requires that a survey campaign record which handset took each fingerprint, which UJIIndoorLoc does and most proprietary surveys could. It does not require a second campaign, a delay, or access to the devices that will eventually query the system. On this database it would have told a team, before deployment, both roughly what error to expect and which representation to choose.

Variation across devices

The device penalty is an average over a heterogeneous group. Table 5 breaks the deployment campaign down by handset for the best representation. Mean error ranges from 4.26 m to 9.76 m. A single figure for deployment accuracy therefore conceals a spread wide enough to decide whether an application works, and a system specified against the mean will disappoint a substantial minority of its users. Reporting the distribution across devices, not only its centre, should be routine (García et al., 2026).

Table 5. Deployment error by handset, best representation

Phone	Fingerprints	In training	Median (m)	Mean (m)
0	120	no	4.02	5.70
2	52	no	4.89	6.92
4	69	no	3.79	6.03
5	17	no	7.58	9.76
9	77	no	3.51	4.26
12	70	no	5.85	6.96
13	369	yes	3.13	4.50
14	28	yes	6.49	9.61
15	36	no	5.05	5.41
20	213	no	6.51	9.63
21	60	no	4.97	9.14

Phone identifiers are those of the database. Two models appear in both campaigns.

What each representation costs to compute

A representation is only useful if the device that needs it can compute it, which on a battery-powered tag or a microcontroller is not a given, and on such a target the energy and latency budget constrains the design as firmly as the accuracy requirement does (Alharthi et al., 2026; Echchidmi & Bouayad, 2026; Heydari & Mahmoud, 2025; Lee et al., 2026). Fingerprinting estimators that run entirely on the querying device have been demonstrated (Kargar-Barzi et al., 2024; Subakti & Jiang, 2026), which makes the arithmetic price of the encoding a live question. Table 6 reports the arithmetic cost of each transform, relative to the cheapest, together with its asymptotic cost. The spread is narrow. The costliest, rank, is 61.3 times the cheapest, and the representation that performs best on deployment costs 8.6 times the cheapest. On the scale of a single fingerprint these differences are immaterial against the nearest-neighbour search that follows them, which is linear in the size of the radio map. The practical conclusion is that the choice of representation should be made on accuracy under a device-disjoint protocol, not on computational grounds, since the cheap option is not meaningfully cheaper. That holds only while the estimator itself fits the device's budget; where it does not, the savings come from quantising or restructuring the estimator (Alharthi et al., 2026; Echchidmi & Bouayad, 2026; Heydari & Mahmoud, 2025; Lee et al., 2026), or from

arranging the search so that the device performs only part of it (Kargar-Barzi et al., 2024; Subakti & Jiang, 2026).

Table 6. Cost of each representation, relative to the cheapest

Representation	Order	Relative cost	Non-zeros per fingerprint	Calibration-free
Raw dBm	O(1)	1.00	465.0	no
Positive	O(1)	4.19	17.8	no
Powed	O(m)	38.74	17.8	no
Exponential	O(m)	8.56	17.8	no
Rank	O(d log d)	61.25	11.8	yes
SSD (mean)	O(d)	15.84	17.8	yes
SSD (max)	O(d)	9.23	17.8	yes
Binary visibility	O(1)	1.82	17.8	yes

d is the number of detected access points in a fingerprint and m the number of columns. Timings are medians over seven repeats on the host that ran the experiments; they order the transforms but say nothing about any particular embedded target.

Comparison with the literature and implications

Our deployment figures are well above the metre-level errors frequently reported for this database. We do not read that as a contradiction but as a difference in what is measured. A study that partitions the pooled file at random is reporting, on the evidence of Section 3.2, the recoverability of near-duplicate fingerprints; that quantity is well defined and reproducible, and it is not the quantity a building owner is buying. Reports that do use the official validation campaign arrive at errors of the same order as ours (S. Jin & Kim, 2024; Mao et al., 2025; Neupane et al., 2026), which is the comparison that matters.

Three recommendations follow. Evaluation should be device-disjoint whenever the survey recorded which handset took each fingerprint, because a random partition answers a question nobody asks and, worse, ranks designs wrongly. The device composition of both the survey and the expected user population should be reported, since the penalty for unfamiliar hardware is of the same order as the total error budget of many applications. And a representation should be defended by measurement rather than by the property it was designed to have: the calibration-free encodings tested here did not deliver the robustness that motivates them.

Threats to validity

Four limitations bound these conclusions. First, one database from one university campus cannot establish how large the device penalty is in general (García et al., 2026); it establishes that the penalty is measurable, that it is of the same order as the interval effect, and that a random partition conceals both. Second, the familiar-device group is small (397 fingerprints from 2 models), so the decomposition rests on a narrow base, and the post-stratified estimate discards 24 fingerprints that fall in building-by-floor cells the familiar group does not cover. Third, the interval and the change of operator are not separable in this database, because the validation campaign was taken by a single operator who does not appear in training; what we label the interval effect therefore includes any systematic difference in how that person carried the handset. Fourth, we fixed the estimator to isolate the effect of the representation, so nothing here establishes that a learned estimator would show the same pattern, although the mechanism identified in Section 3.2 is a property of the partition rather than of the model and would apply

to any estimator trained under it. That caution covers the weighted nearest-neighbour rule used here (Park et al., 2024; Youssef & Agrawala, 2005) as much as the learned alternatives usually benchmarked against it (Esmaeili Gorjan & Gil Jiménez, 2024; Kim et al., 2025; Rana & Dulal, 2025; Wu et al., 2024; H. Zhu et al., 2023).

CONCLUSION

This study evaluated Wi-Fi fingerprinting using a public radio map under realistic deployment conditions. Three principal findings emerged. First, random partitioning, commonly applied to this dataset, yielded a mean planar positioning error of 2.13 m using raw signal strength values expressed in decibels, compared with 8.84 m in the subsequent validation campaign. It also reversed the ranking of signal strength representations, favoring an approach that ranked fifth among eight representations under deployment conditions. Second, this discrepancy was primarily attributable to the survey design rather than the positioning estimator. The dataset contained 19,861 fingerprints collected at only 933 distinct locations, allowing random partitioning to favor near-duplicate fingerprint retrieval rather than independent positioning performance. Third, after matching device groups by building and floor, unfamiliar hardware was associated with an additional mean positioning error of 2.59 m. The remaining performance degradation was consistent with temporal changes in the radio environment, although the contribution of elapsed time could not be independently quantified from a single validation interval.

Signal strength representations designed to reduce sensitivity to device-specific calibration differences did not mitigate the additional positioning errors associated with unfamiliar hardware in this dataset. This finding challenged the assumption that calibration-independent representations necessarily improve cross-device positioning accuracy. In contrast, modifying the evaluation protocol proved more effective than changing the signal representation alone. Leave-one-device-out validation within a single survey campaign estimated the positioning error subsequently observed on unfamiliar devices with a mean signed difference of 0.05 m and identified the same optimal signal representation as the subsequent validation campaign. This approach required no additional survey campaign, provided that handset identifiers were recorded for each fingerprint.

These findings did not indicate that Wi-Fi fingerprinting was inherently unsuitable for indoor positioning. Rather, they demonstrated that conventional random partitioning could produce overly optimistic accuracy estimates that did not adequately reflect deployment conditions. Consequently, relying exclusively on randomly partitioned benchmark results could lead practitioners to overestimate positioning performance and select suboptimal signal representations. Future research should replicate this analysis using datasets collected across multiple survey campaigns over longer periods to distinguish temporal degradation from device-related effects more reliably. Further investigations should also determine whether machine learning-based positioning estimators can reduce cross-device errors more effectively than alternative signal representations. Additionally, access point selection and missing signal strength reconstruction should be investigated as potential strategies for improving positioning robustness across heterogeneous devices and changing radio environments (R. Jin et al., 2026; Wang et al., 2026).

AUTHOR CONTRIBUTION

The author conceived the study, designed and implemented the analysis, interpreted the results and wrote the manuscript.

Funding statement. This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

CONFLICT OF INTEREST

The author declares no conflict of interest

DATA AVAILABILITY

The UJIIndoorLoc database is publicly available from the UCI Machine Learning Repository (Torres-Sospedra et al., 2014). All code that produced the results, figures and tables in this paper, together with the archived experimental output, is available from the author on request.

DECLARATION ON THE USE OF ARTIFICIAL INTELLIGENCE

A large language model was used to assist with code implementation and manuscript drafting. The study design, the choice of protocols, the interpretation of results and all scientific claims are the author's, who has verified every number reported here against the archived experimental output and takes full responsibility for the content.

REFERENCE

Alharthi, S., Rashid, M., & Aljabri, M. (2026). TinyML in industrial IoT: A systematic review of applications, system components, and methodologies. Sensors, 26(8), Article 2550. https://doi.org/10.3390/s26082550

Aziz, T., & Koo, I. (2025). A comprehensive review of indoor localization techniques and applications in various sectors. Applied Sciences, 15(3), Article 1544. https://doi.org/10.3390/app15031544

Chia, Z. Y., Goh, P. Y., Ong, L. Y., & Tan, S. C. (2025). The challenge of dynamic environments in regard to RSSI-based indoor Wi-Fi positioning—a systematic review. Future Internet, 17(12), Article 540. https://doi.org/10.3390/fi17120540

Dai, J., Wang, M., Wu, B., Shen, J., & Wang, X. (2023). A survey of latest Wi-Fi assisted indoor positioning on different principles. Sensors, 23(18), Article 7961. https://doi.org/10.3390/s23187961

Echchidmi, M., & Bouayad, A. (2026). TinyML for sustainable edge intelligence: Practical optimization under extreme resource constraints. Technologies, 14(4), Article 215. https://doi.org/10.3390/technologies14040215

Esmaeili Gorjan, H., & Gil Jiménez, V. P. (2024). Improving indoor WiFi localization by using machine learning techniques. Sensors, 24(19), Article 6293. https://doi.org/10.3390/s24196293

García, A., Beltrán, J., Hernández, N., Parra, I., & Kim, E. (2026). Enhancing robustness to device heterogeneity in WiFi-based indoor localization. Journal of Sensor and Actuator Networks, 15(4), Article 49. https://doi.org/10.3390/jsan15040049

Gidey, H. T., Guo, X., Li, L., & Zhang, Y. (2022). Heterogeneous transfer learning for Wi-Fi indoor positioning based hybrid feature selection. Sensors, 22(15), Article 5840. https://doi.org/10.3390/s22155840

Gidey, H. T., Guo, X., Zhong, K., Li, L., & Zhang, Y. (2022). OHetTLAL: An online transfer learning method for fingerprint-based indoor positioning. Sensors, 22(23), Article 9044. https://doi.org/10.3390/s22239044

Hailu, T. G., Guo, X., & Si, H. (2025). Indoor positioning systems as critical infrastructure: An assessment for enhanced location-based services. Sensors, 25(16), Article 4914. https://doi.org/10.3390/s25164914

Hailu, T. G., Guo, X., Si, H., Li, L., & Zhang, Y. (2024a). Ada-LT IP: Functional discriminant analysis of feature extraction for adaptive long-term Wi-Fi indoor localization in evolving environments. Sensors, 24(17), Article 5665. https://doi.org/10.3390/s24175665

Hailu, T. G., Guo, X., Si, H., Li, L., & Zhang, Y. (2024b). Theories and methods for indoor positioning systems: A comparative analysis, challenges, and prospective measures. Sensors, 24(21), Article 6876. https://doi.org/10.3390/s24216876

Heydari, S., & Mahmoud, Q. H. (2025). Tiny machine learning and on-device inference: A survey of applications, challenges, and future directions. Sensors, 25(10), Article 3191. https://doi.org/10.3390/s25103191

Hossain, A. K. M. M., Jin, Y., Soh, W.-S., & Van, H. N. (2013). SSD: A robust RF location fingerprint addressing mobile devices' heterogeneity. IEEE Transactions on Mobile Computing, 12(1), 65–77. https://doi.org/10.1109/TMC.2011.243

Jin, R., Zhang, D., Tian, X., & Ma, J. (2026). Indoor localization algorithm based on information gain ratio and affinity propagation clustering. Sensors, 26(2), Article 664. https://doi.org/10.3390/s26020664

Jin, S., & Kim, D. (2024). WiFi fingerprint indoor localization employing Adaboost and probability-one access point selection for multi-floor campus buildings. Future Internet, 16(12), Article 466. https://doi.org/10.3390/fi16120466

Kargar-Barzi, A., Farahmand, E., Taheri Chatrudi, N., Mahani, A., & Shafique, M. (2024). An edge-based WiFi fingerprinting indoor localization using convolutional neural network and convolutional auto-encoder. IEEE Access, 12, 85050–85060. https://doi.org/10.1109/ACCESS.2024.3412676

Khan, I. M., Thompson, A., Al-Hourani, A., Sithamparanathan, K., & Rowe, W. S. T. (2023). RSSI and device pose fusion for fingerprinting-based indoor smartphone localization systems. Future Internet, 15(6), Article 220. https://doi.org/10.3390/fi15060220

Kim, D., Park, J.-H., & Suh, Y.-J. (2025). A Wi-Fi fingerprinting indoor localization framework using feature-level augmentation via variational graph auto-encoder. Electronics, 14(14), Article 2807. https://doi.org/10.3390/electronics14142807

Kjærgaard, M. B., & Munk, C. V. (2008). Hyperbolic location fingerprinting: A calibration-free solution for handling differences in signal strength. In Proceedings of the Sixth Annual IEEE International Conference on Pervasive Computing and Communications (PerCom) (pp. 110–116). https://doi.org/10.1109/PERCOM.2008.75

Lee, C.-K., Ohk, S.-R., & Kim, Y.-J. (2026). Revisiting the inference time optimization of TinyML for ARM-based microcontrollers. Electronics, 15(14), Article 2997. https://doi.org/10.3390/electronics15142997

Li, C., Xiao, H., Yu, S., & Xiao, J. (2026). Improved Wi-Fi fingerprinting positioning algorithm based on access point and neighboring point selection. Electronics, 15(18), Article 4277. https://doi.org/10.3390/electronics15184277

Li, J., Song, Y., Ma, Z., Liu, Y., & Chen, C. (2024). A review of indoor localization methods leveraging smartphone sensors and spatial context. Sensors, 24(21), Article 6956. https://doi.org/10.3390/s24216956

Li, S., Tang, Z., Kim, K. S., & Smith, J. S. (2024). On the use and construction of Wi-Fi fingerprint databases for large-scale multi-building and multi-floor indoor localization: A case study of the UJIIndoorLoc database. Sensors, 24(12), Article 3827. https://doi.org/10.3390/s24123827

Machaj, J., Brida, P., & Piché, R. (2011). Rank based fingerprinting algorithm for indoor positioning. In Proceedings of the 2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN) (pp. 1–6). https://doi.org/10.1109/IPIN.2011.6071929

Mao, D., Lin, H., & Lou, X. (2025). Fingerprinting indoor positioning based on improved sequential deep learning. Algorithms, 18(1), Article 17. https://doi.org/10.3390/a18010017

Martín-Frechina, S., Dura, E., Miralles, I., & Torres-Sospedra, J. (2025). From fingerprinting to advanced machine learning: A systematic review of Wi-Fi and BLE-based indoor positioning systems. Sensors, 25(22), Article 6946. https://doi.org/10.3390/s25226946

Mendoza-Silva, G. M., Richter, P., Torres-Sospedra, J., Lohan, E. S., & Huerta, J. (2018). Long-term WiFi fingerprinting dataset for research on robust indoor positioning. Data, 3(1), Article 3. https://doi.org/10.3390/data3010003

Naser, R. S., Lam, M. C., Qamar, F., & Zaidan, B. B. (2023). Smartphone-based indoor localization systems: A systematic literature review. Electronics, 12(8), Article 1814. https://doi.org/10.3390/electronics12081814

Neupane, I., Shahrestani, S., & Ruan, C. (2026). Wi-Fi RSS fingerprinting-based indoor localization in large multi-floor buildings. Electronics, 15(1), Article 183. https://doi.org/10.3390/electronics15010183

Nor Hisham, A. N., Ng, Y. H., Tan, C. K., & Chieng, D. (2022). Hybrid Wi-Fi and BLE fingerprinting dataset for multi-floor indoor environments with different layouts. Data, 7(11), Article 156. https://doi.org/10.3390/data7110156

Park, J.-H., Kim, D., & Suh, Y.-J. (2024). WKNN-based Wi-Fi fingerprinting with deep distance metric learning via siamese triplet network for indoor positioning. Electronics, 13(22), Article 4448. https://doi.org/10.3390/electronics13224448

Perez-Navarro, A., Montoliu, R., & Torres-Sospedra, J. (2022). Advances in indoor positioning and indoor navigation. Sensors, 22(19), Article 7375. https://doi.org/10.3390/s22197375

Rana, L., & Dulal, A. (2025). WiFi RSS and RTT indoor positioning with graph temporal convolution network. Sensors, 25(24), Article 7622. https://doi.org/10.3390/s25247622

Schyga, J., Hinckeldeyn, J., & Kreutzfeldt, J. (2022). Meaningful test and evaluation of indoor localization systems in semi-controlled environments. Sensors, 22(7), Article 2797. https://doi.org/10.3390/s22072797

Silva, I., Pendão, C., & Moreira, A. (2022). Collection of a continuous long-term dataset for the evaluation of Wi-Fi-fingerprinting-based indoor positioning systems. Sensors, 22(22), Article 8585. https://doi.org/10.3390/s22228585

Subakti, H., & Jiang, J.-R. (2026). A robust and sustainable machine learning framework for indoor localization in mobile IoT networks. Electronics, 15(17), Article 3994. https://doi.org/10.3390/electronics15173994

Torres-Sospedra, J., Montoliu, R., Martínez-Usó, A., Avariento, J. P., Arnau, T. J., Benedito-Bordonau, M., & Huerta, J. (2014). UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems. In Proceedings of the 2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN) (pp. 261–270). https://doi.org/10.1109/IPIN.2014.7275492

Torres-Sospedra, J., Montoliu, R., Trilles, S., Belmonte, Ó., & Huerta, J. (2015). Comprehensive analysis of distance and similarity measures for Wi-Fi fingerprinting indoor positioning systems. Expert Systems with Applications, 42(23), 9263–9278. https://doi.org/10.1016/j.eswa.2015.08.013

Wan, Q., Duan, X., Yu, Y., Chen, R., & Chen, L. (2022). Self-calibrated multi-floor localization based on Wi-Fi Ranging/Crowdsourced fingerprinting and low-cost sensors. Remote Sensing, 14(21), Article 5376. https://doi.org/10.3390/rs14215376

Wang, S., Zhang, L., & Huang, X. (2026). Indoor WiFi localization via robust fingerprint reconstruction and multi-mechanism adaptive PSO-LSSVM optimization. Applied Sciences, 16(2), Article 753. https://doi.org/10.3390/app16020753

Wu, Z., Hu, P., Liu, S., & Pang, T. (2024). Attention mechanism and LSTM network for fingerprint-based indoor location system. Sensors, 24(5), Article 1398. https://doi.org/10.3390/s24051398

Yang, J., Deng, S., Xu, L., & Zhang, W. (2022). Calibration-free 3D indoor positioning algorithms based on DNN and DIFF. Sensors, 22(15), Article 5891. https://doi.org/10.3390/s22155891

Yaro, A. S., Maly, F., & Prazak, P. (2023). A survey of the performance-limiting factors of a 2-dimensional RSS fingerprinting-based indoor wireless localization system. Sensors, 23(5), Article 2545. https://doi.org/10.3390/s23052545

Youssef, M., & Agrawala, A. (2005). The Horus WLAN location determination system. In Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services (MobiSys '05) (pp. 205–218).

Zholamanov, B., Saymbetov, A., Nurgaliyev, M., Bolatbek, A., Dosymbetova, G., Kuttybay, N., Orynbassar, S., Kapparova, A., Koshkarbay, N., & Beyca, Ö. F. (2025). RSSI fingerprint-based indoor localization solutions using machine learning algorithms: A comprehensive review. Smart Cities, 8(5), Article 153. https://doi.org/10.3390/smartcities8050153

Zhu, H., Cheng, L., Li, X., & Yuan, H. (2023). Neural-network-based localization method for Wi-Fi fingerprint indoor localization. Sensors, 23(15), Article 6992. https://doi.org/10.3390/s23156992

Zhu, Y., & Cheng, Y. (2025). Research on fingerprint map construction and real-time update method based on indoor landmark points. Sensors, 25(17), Article 5473. https://doi.org/10.3390/s25175473