

Article

Enhancing Robustness to Device Heterogeneity in WiFi-Based Indoor Localization

Adrián García ^{1,*} , Jorge Beltrán ^{2,†} , Noelia Hernández ¹ , Ignacio Parra ¹  and Euntai Kim ³¹ Computer Engineering Department, Universidad de Alcalá, 28801 Alcalá de Henares, Spain; noelia.hernandez@uah.es (N.H.); ignacio.parra@uah.es (I.P.)² Intelligent Robotics Lab, Universidad Rey Juan Carlos, 28942 Fuenlabrada, Spain; jorge.beltran@urjc.es³ School of Electrical and Electronic Engineering, Yonsei University, Seoul 03722, Republic of Korea; etkim@yonsei.ac.kr

* Correspondence: adrian.garcias@uah.es

† This work was partially developed during a research visit to the Computer Engineering Department at the Universidad de Alcalá, while the author was working at the Universidad Carlos III de Madrid.

Abstract

Indoor localization systems based on WiFi are gaining popularity due to their low implementation cost and the widespread availability of WiFi infrastructure. However, the wide variety of existing hardware poses a significant challenge in developing systems that maintain robust and consistent performance regardless of the device used. Recent research has addressed this issue of device heterogeneity by building datasets that include data from a diverse set of devices. In this paper, we tackle this challenge by presenting a novel, multi-device, WiFi Received Signal Strength dataset collected along unconstrained trajectories using nine Android devices over a three-month period with precise ground truth positions obtained using Simultaneous Localization And Mapping. We then study the effect of heterogeneity in the localization performance using an LSTM-based neural network that leverages the temporal nature of sequential WiFi scans, and introduce two mitigation strategies: per-device Received Signal Strength normalization and the incorporation of temporal features as additional input. Our results show that these methods significantly improve cross-device performance with a mean average localization error reduction of 56% and enable generalization to previously unseen hardware with a mean average localization error 8% higher for the unseen devices.

Keywords: dataset; fingerprinting; indoor localization; multi-device; LSTM; RSS; WiFi

Academic Editor: Claudio Savaglio

Received: 4 May 2026

Revised: 16 June 2026

Accepted: 23 June 2026

Published: 27 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Indoor localization plays a crucial role in many applications, such as autonomous robotics, navigation for visually impaired individuals, location-based services, and smart building management. From a broader Internet of Things (IoT) perspective, localization can be exposed as an IoT service that provides real-time spatial context to higher-level cyber-physical services and intelligent building management platforms. This paradigm is aligned with the concept of opportunistic cyber-physical services, where heterogeneous sensing resources are leveraged to dynamically provide context-aware functionalities [1]. Similar to localization services proposed for smart buildings [2], the presented system can be integrated into building automation platforms to support adaptive and personalized services based on user location.

Unlike outdoor environments, where Global Positioning System (GPS) technology provides reliable positioning, indoor settings require alternative methods, as GPS signals are often weak or unavailable. Various technologies have been explored for indoor localization, including Bluetooth, Ultra-Wideband (UWB), Visible Light Communication (VLC) [3], and WiFi. Among these, WiFi-based approaches stand out as the most used due to the widespread availability and the cost-effectiveness of leveraging existing infrastructure.

Localization approaches based on Wifi include trilateration, proximity, and fingerprinting, the latter being most popular because it maps Received Signal Strength (RSS) measurements to known locations without needing the exact positions of Access Points (APs). RSS-based methods excel in complex indoor environments by capturing signal variations and are robust to environmental changes like the rearrangement of furniture or the presence of people. In contrast, trilateration and proximity methods are less effective in adapting to these dynamic conditions.

Although RSS-based indoor localization offers several advantages, it faces significant challenges associated with the dynamic nature of APs in real-world environments [4,5]. Over time, new APs may emerge, some may disappear, and others may temporarily become unavailable, or change positions, leading to a non-static WiFi landscape. Additionally, the existence of multiple networks provided by a single physical AP and transient network configurations further increases the complexity of designing accurate localization systems.

Another major challenge in WiFi-based indoor localization is the impact of device heterogeneity on its performance [6,7]. Differences in hardware characteristics, such as variations in antenna quality, signal scanning duration, and measurement precision across different devices, can lead to inconsistent RSS readings. Such discrepancies introduce a domain gap between training and testing devices, ultimately compromising localization performance.

In this paper, we address the challenge of device heterogeneity in RSS-based indoor localization. We present a comprehensive analysis of the effects of using different mobile devices on localization accuracy, considering both the sampling rate and measured RSS variabilities. Our key contributions are as follows:

- We release a novel multi-device RSS fingerprint indoor localization dataset. Unlike prior works, which rely on fixed reference points or user annotations with limited accuracy, our dataset leverages Simultaneous Localization And Mapping (SLAM) techniques to provide precise ground-truth across unconstrained trajectories.
- We present a systematic, quantitative study of how differences in sampling rate and RSS bias impact localization error. This includes evaluations across matched and unmatched devices, as well as the effect of acquisition differences, rarely addressed in such detail in the existing literature.
- We propose a lightweight and effective training method that combines per-device normalization and the use of timestamp intervals as input features. Unlike existing approaches, our method requires no retraining or calibration for unseen devices.

The remainder of this paper is structured as follows. Section 2 reviews related work on WiFi-based indoor localization challenges. Section 3 details our methodology, including dataset acquisition and the proposed mitigation strategies. Section 4 presents the experimental evaluation of our normalization and temporal fusion methods. Section 5 discusses system limitations, and Section 6 concludes the paper with future research directions.

2. Related Works

Indoor localization has long been a prominent research domain, with WiFi-based systems emerging as a particularly effective approach due to the ubiquity of WiFi infrastructure and its cost efficiency. In particular, WiFi fingerprinting has become one

of the most extensively studied paradigms for indoor positioning, as highlighted by several recent surveys that analyze its methodologies, performance, and limitations in real-world deployments [8–10]. Fingerprinting involves mapping RSS measurements to predefined locations, providing robustness in dynamic environments characterized by furniture rearrangements or transient obstructions. WiFi availability and ease of deployment have contributed significantly to its popularity in both academic and industrial implementations [11].

However, RSS-based systems also face significant challenges, particularly concerning device heterogeneity, which affects the consistency and accuracy of RSS readings across different devices. Other well-known limitations include signal instability, multipath propagation, and the high cost of radio map construction and maintenance, which have been widely addressed in the literature [12].

2.1. Challenges of Device Heterogeneity

Device heterogeneity poses a critical challenge in WiFi-based indoor localization, as variations in hardware (e.g., antenna sensitivity, signal processing capabilities) across devices lead to discrepancies in RSS measurements. This inconsistency can result in reduced localization accuracy, especially when models trained on data from one device are applied to another. This problem has been consistently identified as one of the main bottlenecks preventing large-scale deployment of fingerprinting-based solutions, since it limits model transferability and requires costly recalibration processes [8,9]. Efforts to address this issue have largely focused on either inter-device calibration techniques or robust fingerprinting methods.

Inter-device calibration approaches aim to standardize RSS measurements by transforming or scaling the signals. Techniques such as using signal strength histograms to calibrate devices [13], performing multi-device joint localization through wireless data fusion [14], or employing linear regression models to map RSS values between devices [15] have demonstrated potential in minimizing device-induced biases. Other strategies, like particle filtering methods [14] or hidden Markov models that smooth RSS variations and reduce noise [16], have also been explored. Furthermore, methods like Gaussian Process modeling [17] and automatic linear transformation [18] provide device-agnostic normalization strategies to mitigate heterogeneity effects. While effective, these approaches often require extensive pre-profiling or simultaneous data collection from multiple devices, which may limit scalability. In parallel, recent research has explored machine learning and deep learning techniques to implicitly learn device-invariant representations, reducing the need for explicit calibration. These approaches, while promising, typically require large labeled datasets and may suffer from generalization issues in unseen environments [19].

Robust fingerprinting methods aim to create device-invariant features. For instance, some approaches focus on ranking RSS values rather than relying on absolute measurements, assuming that the relative order of RSS from different APs remains consistent across devices [20]. Others, such as Procrustes analysis, utilize statistical shape alignment to normalize RSS measurements into device-independent fingerprints [21]. Composite fingerprint sets that combine multiple features (e.g., RSS, signal strength differences, and hyperbolic fingerprints) have also been proposed, achieving strong performance in classification-based localization [22]. Ensemble methods combining multiple classifiers and diverse features have further enhanced the robustness of localization systems against device heterogeneity [23].

Beyond RSS-based approaches, alternative signal representations such as Channel State Information (CSI) have been investigated to improve robustness and sensitivity to environmental changes, although their practical deployment is still constrained by hardware limitations [24,25].

2.2. WiFi-Based Localization Datasets

To facilitate research on device-agnostic localization methods, numerous datasets have been developed, differing in scale, device diversity, and data collection methodologies. Notable examples include the UJIIndoorLoc dataset [26], which spans multiple buildings and incorporates data from various devices, and the JUIndoorLoc dataset [27], which proposes ensemble-based classifiers for device-heterogeneous environments. While these datasets provide valuable benchmarks, they often rely on predefined reference points or manual annotations, which can introduce granularity and precision limitations.

Crowdsourced datasets, such as those described in [28], offer an alternative by leveraging user-generated data for more organic collection processes. However, these datasets frequently suffer from lower coordinate accuracy due to manual annotations. In contrast, datasets employing advanced localization systems, such as Light Detection and Ranging (LiDAR) based SLAM for ground truth generation [29], ensure higher precision but are less common due to the complexity of setup and cost. Other noteworthy datasets include those that integrate multi-modal data (e.g., combining WiFi with geomagnetic or inertial sensors) [30,31], datasets focusing on long-term data collection to analyze temporal stability [32], and large-scale datasets spanning multiple floors and device configurations [33].

Recent works have emphasized the importance of standardized benchmarking and reproducibility, highlighting the need for datasets that capture realistic variability in user behavior, device diversity, and environmental dynamics. Despite these efforts, there is still a lack of universally accepted large-scale benchmarks that simultaneously address scalability, heterogeneity, and high-precision ground truth acquisition, which remains an open research challenge [34,35].

3. Materials and Methods

This section outlines the experimental pipeline and structural design of our device-robust indoor localization framework. We first describe the baseline materials, including the physical collection environment and the heterogeneous profiles of the hardware utilized. Following this, we detail our proposed algorithmic strategies—specifically, per-device RSS range normalization and temporal feature fusion—before defining the structural configuration of the sequence-based Long Short-Term Memory (LSTM) network.

3.1. Data Collection

The recorded dataset compiles data collected over 13 separate days between late March and early July 2023 (see Figure 1), at the Escuela Politecnica Superior of the University of Alcalá, Spain, a public building spanning 17,000 m², which hosts lectures for technical bachelor's and master's degree programs. The data was exclusively recorded on the 3rd floor, an area with lecturers' offices and laboratories, covering approximately 1639 m². In total, the dataset comprises approximately 5.5 h of recordings and over 13.0 km of traveled distance per device.

Data was collected using nine different devices, including seven smartphones and two tablets from various brands (see Table 1). All devices were Android-based and operated using a custom-designed application developed specifically for data collection. The devices were intentionally selected to represent a broad range of price categories, ensuring diversity in Wi-Fi hardware characteristics such as signal gain and sampling rate. This variability enhances the dataset's representativeness for real-world scenarios. Three of the devices were introduced on the final day of data acquisition and were used exclusively for testing purposes (see Figure 1).

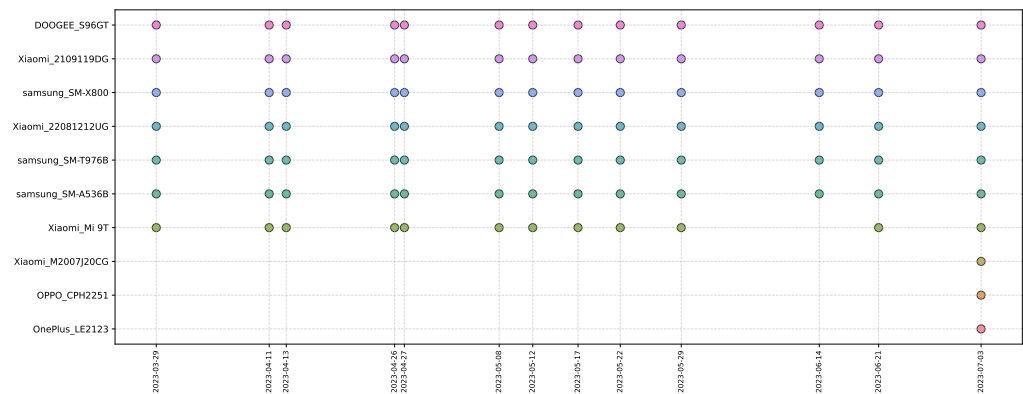


Figure 1. Devices used on the different collection days.

During the data acquisition, the devices were placed in fixed locations over a wheeled table that was pushed at an average speed of 0.7 m/s. All recordings were made on a single trajectory per day basis, with an average duration of approximately 20 min. The trajectories were random and aimed to cover the entire navigable area of the facility.

Table 1. Device type and acquisition time.

Device	Type	Acquisition Time (s)
Xiaomi 22081212UG ^a	Phone	1.6
Samsung SM-T976B ^b	Tablet	1.9
Doogee S96GT ^c	Phone	2.0
Xiaomi M2007J20CG ^a	Phone	2.2
OnePlus LE2123 ^d	Phone	2.4
Oppo CPH2251 ^e	Phone	2.5
Samsung SM-800 ^b	Tablet	3.0
Xiaomi 2109119DG ^a	Phone	3.1
Samsung SM-A536B ^b	Phone	4.1

^a Xiaomi Corporation, Beijing, China; ^b Samsung Electronics, Suwon, South Korea; ^c Doogee, Shenzhen, China;

^d OnePlus Technology Co., Ltd., Shenzhen, China; ^e OPPO Electronics Corp., Dongguan, China.

To calculate the ground truth position, a 3D map of the environment was previously created using RTAB-Map's mapping tool fed with images captured by an Intel RealSense D435i camera. When collecting the data, RTAB-Map [36] provided 1Hz 6-dof relative to the map positioning information. To address RTAB-Map localization errors caused by building symmetry and long corridors, the trajectory is segmented based on estimated speed to exclude anomalous tracking segments.

Regarding the format, the collected data for each device at time step k is as follows:

$$[RSS_k(m), t_k(s)] \quad (1)$$

where $RSS_k(m)$ is the RSS in dBm received from m APs, and t_k is the capture time from the device's clock. The ground truth consists of a 2D metric position relative to a map origin in meters:

$$[x_k(m), y_k(m), t_k(s)] \quad (2)$$

where t_k is the Unix timestamp extracted from the laptop that recorded the ground truth. The data and the ground truth are aligned using their respective time stamps.

3.2. Proposed Mitigation and Robustness Strategies

One of the main challenges in using WiFi for localization is designing a system that performs reliably across different devices. Changing the receiver often introduces variability in both signal acquisition timing and the RSS, which can significantly impact localization accuracy.

As an initial attempt to build a device-agnostic system, a model can be trained using data collected from all available devices. However, data from different devices cannot be interpreted uniformly. For instance, due to varying acquisition rates, sequences of the same length may correspond to substantially different durations in real time. Additionally, hardware-related differences, such as antenna design and receiver sensitivity, can lead to discrepancies in RSS values even when measured at the same physical location.

To address the challenges of device heterogeneity, we propose a two-step processing framework that operates at different layers of the telemetry stream. First, a device-specific calibration is applied to standardise absolute signal power distributions and remove antenna sensitivity biases. Second, a temporal feature fusion strategy integrates the dynamic sampling intervals into the input space to guide the recurrent state updates. Together, these complementary approaches transform raw, hardware-dependent measurements into device-agnostic representations before network ingestion.

3.2.1. Device-Specific RSS Normalization

To address discrepancies in RSS values across different devices, particular attention must be given to the normalization process. This step is crucial to ensure that low RSS values are not marginalized and that high values do not exert a disproportionate influence on the model. A common approach is global normalization, where fixed minimum and maximum values derived from the entire dataset are used to scale the data prior to inputting it into the neural network.

Although global normalization provides consistent scaling, it also introduces a critical limitation: the persistence of device-dependent biases in RSS measurements. Due to hardware variations, such as differences in antenna sensitivity or signal amplification, the same AP may produce substantially different RSS values across devices. Consequently, applying the same normalization parameters can compress or distort the signal distribution for certain devices, ultimately degrading model performance.

To enhance the robustness of the localization models to device-dependent variations, we propose a device-specific normalization strategy. In this approach, each device undergoes a calibration phase to determine its individual RSS range, which is then used for normalization during inference. While this method requires a brief calibration period prior to deployment, it significantly reduces both localization error and prediction variance, leading to more robust and consistent performance across heterogeneous devices.

3.2.2. Temporal Feature Fusion and Sampling Hypothesis

While device-specific normalization balances absolute signal amplitude discrepancies, hardware heterogeneity introduces a secondary, temporal domain gap caused by variable device-dependent scanning intervals. As detailed in Table 1, the raw telemetry acquisition rates vary significantly across hardware platforms, ranging from high-density sampling (1.6 s per scan) to sparse sampling intervals (4.1 s per scan).

We hypothesize that sequence-based recurrent networks, such as the utilized LSTM, implicitly encode spatial displacement as a function of the typical temporal delta observed between consecutive training steps. When a model optimized on high-density tracking data encounters sparse sampling intervals during inference, the physical distance traveled by the user between consecutive steps dramatically exceeds the spatial transitions learned

during the training phase. This mismatch disrupts the continuous propagation of the internal cell states (c_t) and hidden vectors (h_t), causing the network to miscalculate spatial motion dynamics and inflate trajectory tracking drift.

To validate this hypothesis and decouple the architecture from hardware clock biases, we propose an augmented Temporal Feature Fusion strategy. Let $\Delta t_t = t_t - t_{t-1}$ represent the real-time clock interval elapsed between consecutive WiFi scans. This raw 1D temporal feature is independently normalized to a uniform $[0, 1]$ range via global min-max scaling to prevent numeric dominance over the radio maps:

$$\Delta t_{norm} = \frac{\Delta t_t - \Delta t_{min}}{\Delta t_{max} - \Delta t_{min}} \quad (3)$$

The normalized time delta is then directly concatenated to the tail of the device-normalized RSS vector $\mathbf{x}_t \in \mathbb{R}^{556}$. This yields an augmented spatio-temporal input vector $\mathbf{z}_t \in \mathbb{R}^{557}$ before network ingestion:

$$\mathbf{z}_t = [\mathbf{x}_t \parallel \Delta t_{norm}] \quad (4)$$

By explicitly injecting the elapsed time interval into the feature space, we provide the network with the structural context necessary to dynamically adjust its internal gating mechanisms, scaling its mathematical coordinate updates relative to the real-world time step duration regardless of the testing device's hardware limitations.

3.3. Sequence-Based Model Architecture

The choice of model for the experiments presented in this paper is motivated by our previous studies [37,38], where Artificial Neural Networks (ANNs) and K-Nearest Neighbours (KNNs) were evaluated against LSTM networks [39]. The results consistently showed that LSTM models achieved superior localization performance, motivating their adoption in the present work. The final LSTM architecture and hyperparameter configuration were determined through an ablation study and are presented in Figure 2. The selected model consists of $L = 6$ stacked LSTM layers with a hidden size of $h = 32$ and a sequence length of 16 ($N = 15$).

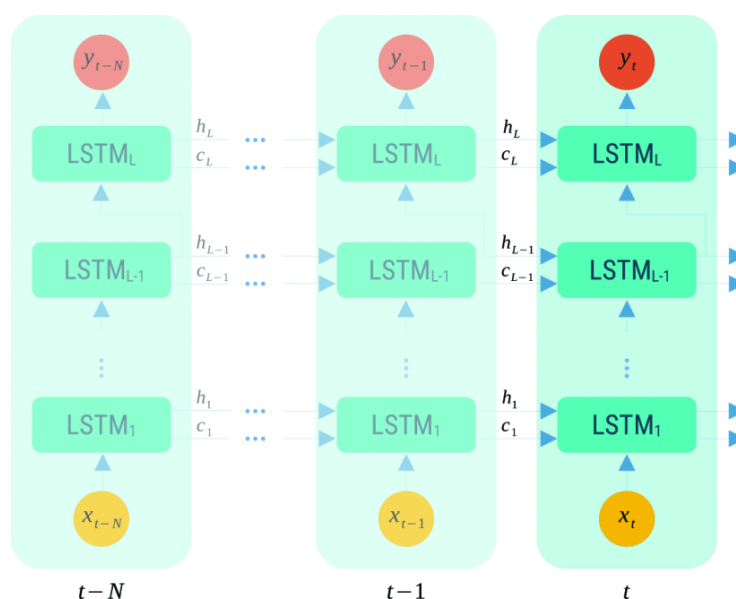


Figure 2. Structural diagram of the LSTM network used in the experiments. The models trained in this paper had $L = 6$ stacked LSTM layers with a hidden size $h = 32$ and were trained with a sequence length of 16 ($N = 15$).

4. Experimental Results

To systematically evaluate the impact of device heterogeneity on localization performance, the proposed LSTM architecture is subjected to various training strategies while keeping its network hyperparameters strictly uniform. In this section, we describe the experimental setup and the results of the different proposals to increase the robustness of cross-device Wi-Fi-based positioning.

4.1. Experimental Setup and Evaluation Metrics

To prevent look-ahead bias, the dataset was divided into three subsets: the first 8 days were used for training, the following 2 days for validation, and the final 3 days for testing (see Figure 1). Only APs observed during the training phase were considered, and any newly appearing APs in the validation or test sets were excluded. Out of a total of 760 unique APs observed in the dataset, only $K = 556$ were present in the training set and therefore used in the experiments.

RSS data from Wi-Fi scans is structured into vectors $x \in \mathbb{R}^K$, where each element represents the signal strength (in dB) from a predefined list of visible APs. Unseen APs are assigned a value $\epsilon = \min(AP_{RSS}) - 5$ to distinguish them from weak signals without affecting normalization. Both the input features and target positions are normalized to the $[0, 1]$ range using min–max scaling. Alternative normalization methods, including z-score normalization, were evaluated but did not improve model performance.

A separate LSTM-based model is trained for each device using sequences of RSS values over N time steps. At the start of each sequence, LSTM states are reset, and samples are fed sequentially. Although the model predicts coordinates at all N steps, only the final prediction (y_t) is used for evaluation; earlier outputs help build temporal context, requiring a warm-up period of N scans. During inference, LSTM states are initialized once per trajectory, and predictions are generated scan-by-scan to emulate real-time performance.

The model is trained by minimizing the Mean Squared Error (MSE) between the predicted and ground truth positions at each time step. Model performance is evaluated using the Mean Absolute Error (MAE) on the final position predictions.

4.2. Evaluation of Per-Device RSS Normalization

We evaluate the empirical performance shift when transitioning from a standard global scaling framework to our proposed device-specific RSS normalization strategy.

To isolate the baseline impact of this calibration method, Figure 3 compares the MAE obtained by the LSTM-based network under different normalization strategies. In both configurations, the model was trained on data from the six devices in the training set and evaluated individually on their corresponding test set data.

As shown, device-specific normalization leads to a reduction in both MAE and error variance. These results provide a direct quantitative measure of how shifting to localized signal ranges dampens hardware-induced sensitivity biases and stabilizes error variance across mismatched antenna components. However, the results also indicate that MAE tends to increase with the device's acquisition time, an effect that will be further analyzed in the following section.

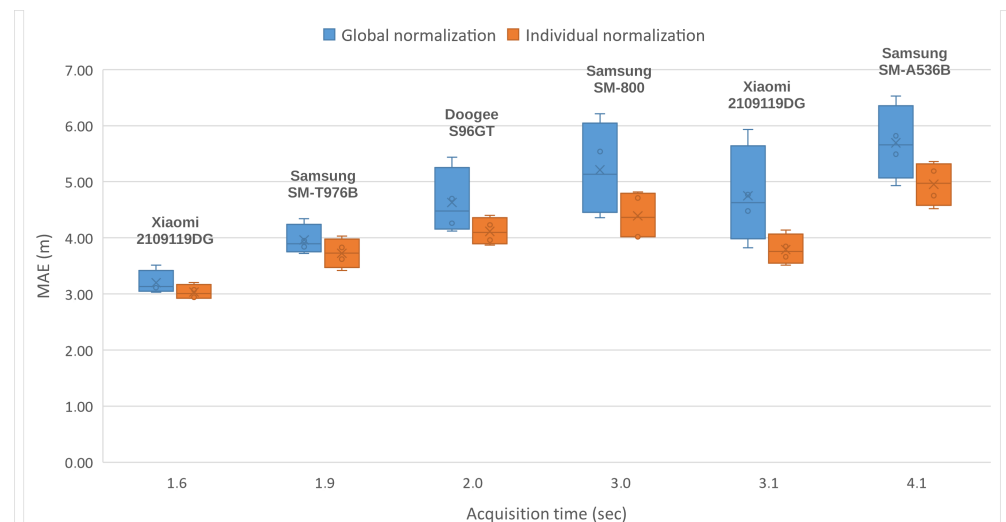


Figure 3. MAE on the test dataset for 21 June 2023. Models were trained on data from six devices in the training set and evaluated individually against the same devices in the test set. Results shown in blue use global normalization, while those in orange use device-specific RSS normalization.

4.3. Evaluation of Temporal Feature Fusion

While device-specific normalization successfully mitigates absolute signal amplitude biases, the localization error remains highly sensitive to variations in device-dependent scanning intervals. To systematically isolate and evaluate this temporal phenomenon, we studied the performance stability of the LSTM network when subjected to severe cross-device sampling discrepancies

Figure 4 presents a boxplot of the localization MAE in meters for models trained on data from a single device and tested on data from all other devices. The colors represent the device used for training, while the x-axis indicates the time offset in data acquisition between the training and testing devices. A value of 0 on the x-axis corresponds to the scenario where each of the six available devices is tested on its own data, i.e., the model is both trained and evaluated on the same device.

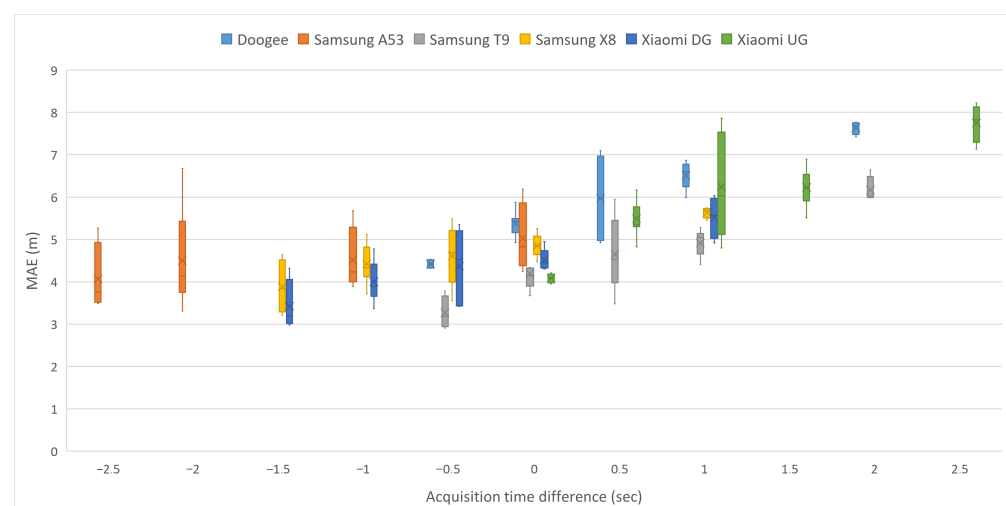


Figure 4. Boxplot of localization MAE in meters for the models trained using a single device and tested on all the other devices' data. As depicted in the legend, the colors indicate the device used to train the model. The x-axis represents the difference in the acquisition time between the training and the testing device. At 0 on the x-axis the results of training and testing the 6 different devices on their own test data are shown.

Looking at Figure 4 two different conclusions can be drawn. Firstly, using the same device for training and testing (0 in x-axis) produces similar errors to when the device used in test has faster acquisition times (negative times in x's), apparently, independently of the acquisition time difference. On the other hand, when the device used in test is slower (positive time difference in x's), the MAE increases proportionally to the acquisition time difference.

This effect is evident in the results of the Xiaomi UG device. Being the fastest, the MAE increases up to 80% when inferring with data from other devices compared to inferring with its own data. Conversely, with the slowest device, Samsung A53, since all other devices are faster, the network performs well and even reduces the error up to 30% with the fastest device.

Furthermore, using a different device in test leads to an increase in the variance of the error, as represented in Figure 4. This trend is evident in all cases, and is particularly pronounced the greater the difference in acquisition time. This empirical trend directly validates our sampling hypothesis. Because sequence-based recurrent models implicitly map spatial displacement relative to the consistent temporal deltas observed during training, hardware platforms characterized by significantly lower sampling rates introduce a severe temporal domain shift. The physical distance traversed by the user between these prolonged intervals vastly exceeds the spatial step size internalized by the LSTM cells. Consequently, the network fails to properly scale its internal state transitions, miscalculates the cumulative spatial motion dynamics, and introduces an artificial tracking drift that inflates the localization error.

To evaluate the empirical effectiveness of our proposed solution against these hardware sampling distortions, the LSTM network was evaluated using the augmented spatio-temporal input space presented in Section 3.2.2.

Figure 5 presents the localization MAE obtained across the test datasets when incorporating the normalized timestamp intervals (Δt) as an additional feature. As in previous experiments, the model was trained using data from the six devices in the training set, and evaluated individually on the corresponding test data for each device. Additionally, three devices not seen during training were included for evaluation: Xiaomi M2007J20CG (2.2 s), OnePlus LE2123 (2.4 s), and Oppo CPH2251 (2.5 s). As shown in Figure 5, incorporating temporal information enables the network to perform consistently across all devices.

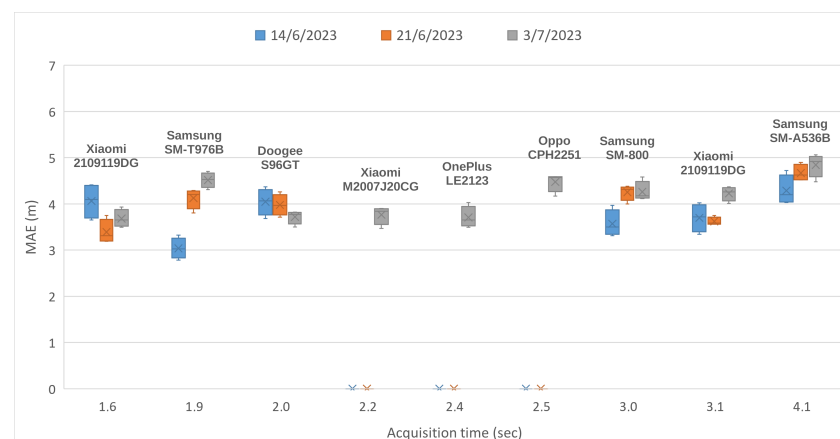


Figure 5. MAE on the test dataset for a model trained using the time difference between samples. The model was trained on the six devices available in the training set and evaluated individually on the same devices in the test set. Additionally, three previously unseen devices—available only on 3 July 2023—were included in the evaluation: Xiaomi M2007J20CG (2.2 s), OnePlus LE2123 (2.4 s), and Oppo CPH2251 (2.5 s).

Compared to the results shown in Figure 3, the mean error values follow a similar pattern: faster devices exhibit a slight performance drop, while slower devices show improvement. As a result, incorporating temporal information reduces performance variability across devices. This adjustment decreases the average inter-device discrepancy from approximately 35% to 25%, indicating improved consistency and fairness in the model's predictions.

4.4. Aggregated Model Performance and Unseen Hardware Generalization

To evaluate the effect of device data aggregation and the proposed device-specific RSS normalization, a new model, referred to as the Aggregated Model (AM), was trained using the combined RSS data from the six devices included in the training set and normalized as explained in Section 3.2.1. The model retained the same network architecture and hyperparameter configuration as the LSTM models described in Section 3.3. Unlike the previous single-device configurations, the AM is optimized by combining the telemetry of multiple distinct training devices into a unified spatio-temporal framework, leveraging both our device-specific normalization and temporal feature fusion strategies.

The RSS values from each device were independently normalized prior to training. During both validation and testing, RSS normalization was performed using the statistics specific to each individual device. In addition to RSS measurements, the time elapsed since the previous sample was incorporated into the input feature vector. To evaluate its performance, the AM was compared with models trained using data from individual devices (see Table 2).

Table 2. MAE on test (m) for different train models. Best performing model for each test is marked in bold. Cells marked with “-” indicate no results are available.

Train \ Test	Test									MAE for all Tests
	Doogee S96GT	Samsung SM-A536B	Samsung SM-T976B	Samsung SM-800	Xiaomi 2109119DG	Xiaomi 22081212UG	OnePlus LE2123	Oppo CPH2251	Xiaomi M2007J20CG	
Doogee S96GT	4.62	5.75	3.89	4.51	5.27	5.15	-	-	-	4.86
Samsung SM-A536B	20.06	4.42	12.81	9.47	8.43	8.89	-	-	-	10.68
Samsung SM-T976B	10.21	12.52	2.89	7.02	6.03	6.74	-	-	-	7.56
Samsung SM-800	5.10	5.40	3.34	3.69	3.99	4.50	-	-	-	4.33
Xiaomi 2109119DG	19.99	20.65	14.56	17.16	3.44	15.80	-	-	-	15.26
Xiaomi 22081212UG	8.40	10.62	4.50	5.84	5.95	5.15	-	-	-	6.74
AM (test-set)	3.82	4.39	3.08	3.71	3.68	3.04	-	-	-	3.62
AM (3 July 2013)	3.75	4.98	3.76	4.1	4.44	3.04	3.38	4.53	3.86	3.92

As shown, the AM achieves performance comparable to the best individual models (those trained and tested on the same device). The models trained on individual devices achieved an average test-set MAE of 8.24 m. In contrast, the AM achieved an average MAE of 3.62 m, corresponding to a 56% reduction in error. When evaluated separately for each device, the AM consistently outperformed the device-specific models, with the smallest improvement observed for the Doogee device, for which the MAE was reduced by 25%. The variability of the localization error can also be inferred from the cumulative distribution functions (CDFs) presented in Figure 6 as well as the percentile errors.

Furthermore, when tested on devices not seen during training or validation, the AM performs similarly to those it was trained on, demonstrating strong generalization to previously unseen hardware. Three new devices were used on the last test day and used to evaluate the robustness of the AM model to previously unseen hardware. As shown on Table 2 AM (3 July 2013), the AM model has similar MAE for the devices in the training dataset and for the three new ones. In all but one case, the lowest MAE was obtained when the training and testing data originated from the same device, highlighting the limited cross-device generalization of these models. In contrast, the performance of the AM degraded only marginally when evaluated on unseen devices, with the MAE increasing from 3.62 m to 3.92 m, corresponding to an 8% degradation. These results indicate that aggregating data from multiple devices during training substantially improves the model's robustness to device heterogeneity and enhances its ability to generalize to previously unseen hardware.

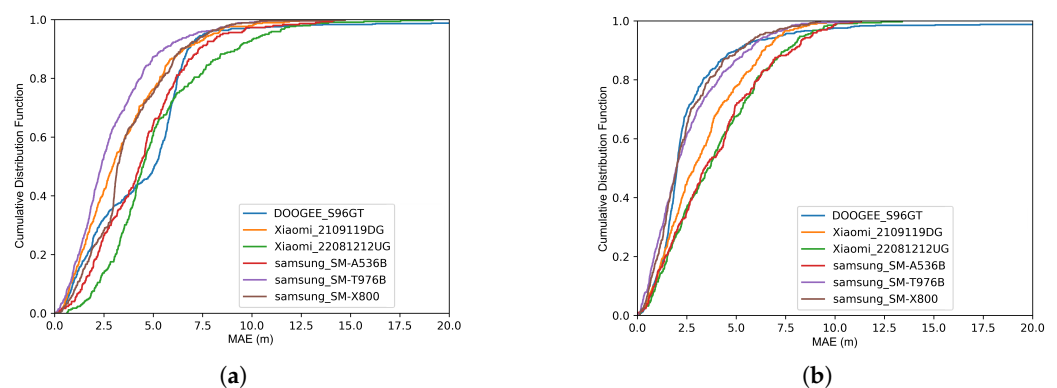


Figure 6. Comparison of CDFs for the results showed on Table 2. (a) Cumulative Distribution Function for the test-set with the models trained on a single device, (b) Cumulative Distribution Function for the test-set with the AM.

Also, qualitative results for the model can be observed on Figure 7 where the predicted trajectory and its error codified in colors are represented for two devices present in the training set (top row) and for two devices that were included only for testing and never seen on training or validation (bottom row). As can be seen, the results are very similar, showing the ability of the model to cope with devices never seen in training.

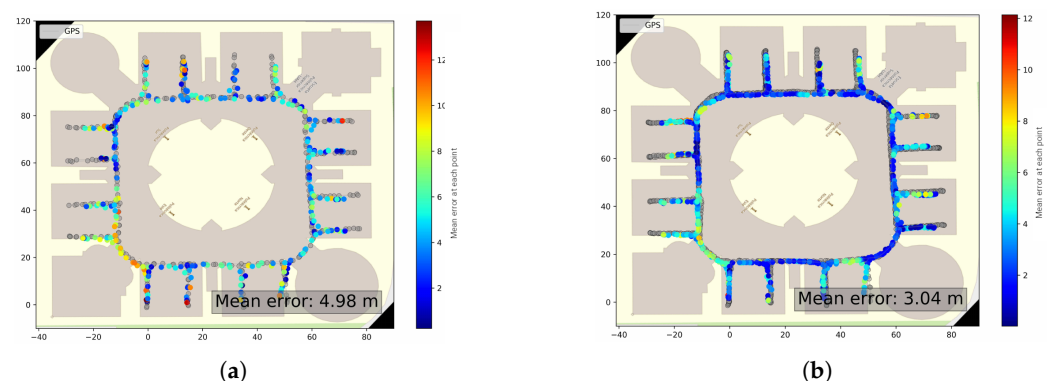


Figure 7. Cont.

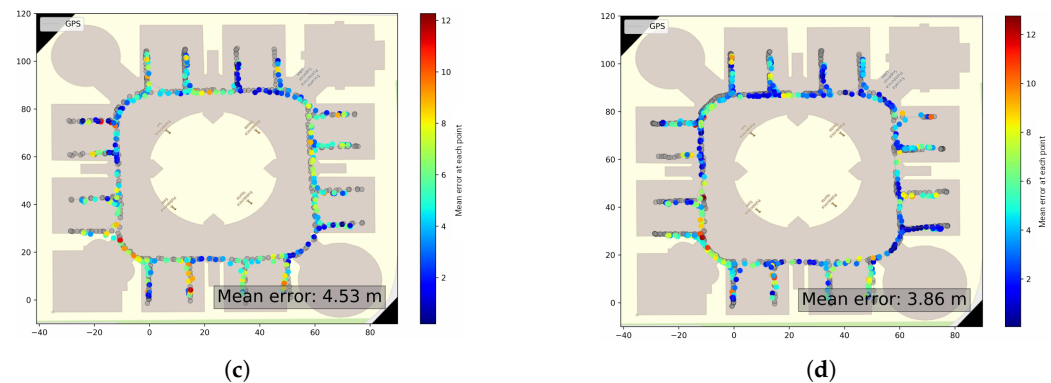


Figure 7. Localization results of the AM for 2 devices in the training set (**top row**) and 2 devices never seen (**bottom row**). Grey, the ground truth; jet color, the localization error. (a) Test results for Samsung SM-A536B device on 3 July, (b) test results for Xiaomi 22081212UG device on 3 July, (c) test results for Oppo CPH2251 device on 3 July, (d) test results for Xiaomi M2007J20CG device on 3 July.

5. Limitations

While the proposed framework demonstrates a significant improvement in cross-device robustness and generalization to unseen hardware, certain boundaries regarding our data collection and evaluation framework must be acknowledged. A transparent analysis of these tracking constraints, spatial bounds, and environmental dynamics is essential for contextualizing our current findings and defining future development paths.

5.1. Ground-Truth Positions Fidelity

Because the dataset's position coordinates are generated via a visual-inertial SLAM pipeline (RTAB-Map), the baseline annotations are inherently subject to minor tracking drift. This phenomenon is particularly common in indoor environments characterized by long corridors or structural symmetries, where repetitive visual patterns can challenge feature-matching algorithms.

To prevent these minor mapping inconsistencies from propagating into the deep network optimization phase, an automated velocity-based data filtering protocol was implemented during the preprocessing stage. By continuously monitoring the velocity profile of the data collection platform against the planned average push rate (0.7 m/s), segments containing anomalous tracking jumps or temporary tracking loss were systematically isolated and removed from the training matrices. This ensures that any localized SLAM drift artifacts are filtered out at the source, preventing them from skewing the LSTM network's downstream localization metrics.

5.2. Long-Term Environmental and Radio-Landscape Changes

Although our data collection spanned a three-month seasonal window, the indoor WiFi landscape is inherently non-static due to fluctuating human density and potential infrastructure modifications (e.g., changes in AP placement or transient channel configurations). While the sequence-based LSTM leverages historical trajectories to smooth out minor signal fades, an extreme long-term shift in the underlying AP infrastructure could degrade the fingerprinting baseline.

5.3. Spatial and Multi-Floor Constraints

A structural limitation of the current evaluation is its spatial scope, as data collection was restricted exclusively to a single 1639 m² floor of the university facility. While this area provided complex trajectories and unconstrained testing conditions, real-world deployment

scenarios—such as multi-story shopping centers or hospital complexes—require systems to seamlessly adapt to vertical tracking dynamics across stairs and elevators.

6. Conclusions

In this paper, we introduced a novel multi-device WiFi dataset with position coordinates obtained through visual SLAM techniques. Unlike many existing datasets that rely on fixed reference points or manual annotations, our dataset includes free-form trajectories and diverse device configurations, enabling robust evaluation of the effects of device heterogeneity.

Through extensive analysis, we identified two key sources of error caused by device diversity: inconsistencies in RSS signal strength and variation in sampling rates. To address these issues, we proposed two mitigation strategies: device-specific RSS normalization and the inclusion of temporal information as an additional input feature. Our experiments demonstrated that both approaches lead to improved localization accuracy and greater consistency across devices. To evaluate the model's performance across heterogeneous devices, we tested the model on previously unseen devices, confirming its ability to generalize.

Future work will expand this framework in three key directions to address current deployment boundaries. First, the dataset and LSTM network will be extended to support three-dimensional tracking across multi-floor environments, accounting for vertical transitions via stairs and elevators. Second, to handle long-term modifications in the indoor radio landscape, we will investigate life-long learning and autonomous self-calibration paradigms that adapt to moving or decommissioned APs without requiring manual profiling. Finally, future iterations will integrate advanced multi-session SLAM loop-closure optimizations to further reduce baseline tracking drift, ensuring stable trajectory annotation across sprawling ambient IoT infrastructures.

Author Contributions: Conceptualization: J.B., N.H., I.P., E.K.; Data curation: A.G., N.H.; Formal analysis: A.G.; Funding Acquisition: N.H., J.B.; Investigation: J.B., N.H., I.P., E.K.; Methodology: N.H.; Software: A.G., J.B.; Writing: J.B., N.H., I.P.; Review and editing: J.B., N.H., I.P., E.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by Research Grant PID2023-150854NB-I00 from the Spanish Ministry of Science, Innovation and Universities. Furthermore, Jorge Beltrán acknowledges the support of the Ministry of Universities and the Universidad Carlos III de Madrid's Call for Grants for the requalification of the Spanish University System for 2021–2023, based on Royal Decree 289/2021 of 20 April 2021, which regulates the direct granting of subsidies to public universities for the requalification of the Spanish university system.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Fortino, G.; Russo, W.; Savaglio, C.; Viroli, M.; Zhou, M. Opportunistic cyberphysical services: A novel paradigm for the future Internet of Things. In Proceedings of the 2018 IEEE 4th World Forum on Internet of Things (WF-IoT), Singapore, 5–8 February 2018; pp. 488–492. [\[CrossRef\]](#)
2. Varma, P.; Anand, V. Random Forest Learning Based Indoor Localization as an IoT Service for Smart Buildings. *Wirel. Pers. Commun.* **2017**, *117*, 3209–3227. [\[CrossRef\]](#)
3. Su, D.; Wang, X.; Liu, S.; Ding, W. Four-dimensional indoor visible light positioning: A deep-learning-based perspective. *J. Frankl. Inst.* **2023**, *360*, 4071–4090. [\[CrossRef\]](#)
4. Chen, X.; Kong, J.; Guo, Y.; Chen, X. An empirical study of indoor localization algorithms with densely deployed APs. In Proceedings of the 2014 IEEE Global Communications Conference, Austin, TX, USA, 8–12 December 2014; pp. 517–522. [\[CrossRef\]](#)

5. Gu, Y.; Chen, M.; Ren, F.; Li, J. HED: Handling environmental dynamics in indoor WiFi fingerprint localization. In Proceedings of the 2016 IEEE Wireless Communications and Networking Conference, Doha, Qatar, 3–6 April 2016; pp. 1–6. [\[CrossRef\]](#)
6. Luo, J.; Zhan, X. Characterization of smart phone received signal strength indication for WLAN indoor positioning accuracy improvement. *J. Netw.* **2014**, *9*, 739. [\[CrossRef\]](#)
7. Retscher, G.; Leb, A. Influence of the RSSI scan duration of smartphones in kinematic Wi-Fi fingerprinting. In Proceedings of the FIG Working Week 2019, Hanoi, Vietnam, 22–26 April 2019; p. 15.
8. Dai, J.; Wang, M.; Wu, B.; Shen, J.; Wang, X. A Survey of Latest Wi-Fi Assisted Indoor Positioning on Different Principles. *Sensors* **2023**, *23*, 7961. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Zholamanov, B.; Saymbetov, A.; Nurgaliyev, M.; Bolatbek, A.; Dosymbetova, G.; Kuttybay, N.; Orynbassar, S.; Kapparova, A.; Koshkarbay, N.; Beyca, Ö.F. RSSI Fingerprint-Based Indoor Localization Solutions Using Machine Learning Algorithms: A Comprehensive Review. *Smart Cities* **2025**, *8*, 153. [\[CrossRef\]](#)
10. Campos, R. A Survey of WiFi Indoor Positioning Techniques. *IOSR J. Eng.* **2022**, *12*, 16–25.
11. Ezhumalai, B.; Song, M.; Park, K. An Efficient Indoor Positioning Method Based on Wi-Fi RSS Fingerprint and Classification Algorithm. *Sensors* **2021**, *21*, 3418. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Neupane, I.; Shahrestani, S.; Ruan, C. Wi-Fi RSS Fingerprinting-Based Indoor Localization in Large Multi-Floor Buildings. *Electronics* **2026**, *15*, 183. [\[CrossRef\]](#)
13. Laoudias, C.; Piché, R.; Panayiotou, C.G. Device self-calibration in location systems using signal strength histograms. *J. Locat. Based Serv.* **2013**, *7*, 165–181. [\[CrossRef\]](#)
14. Sou, S.I.; Wu, F.J.; Wu, W.C. JoLo: Multi-Device Joint Localization Based on Wireless Data Fusion. *IEEE Trans. Mob. Comput.* **2023**, *22*, 4016–4031. [\[CrossRef\]](#)
15. Zhang, L.; Ma, L.; Xu, Y.; Li, C. Linear Regression Algorithm against Device Diversity for Indoor WLAN Localization System. In Proceedings of the GLOBECOM 2017–2017 IEEE Global Communications Conference, Singapore, 4–8 December 2017; pp. 1–6. [\[CrossRef\]](#)
16. Tiku, S.; Pasricha, S.; Notaros, B.; Han, Q. A Hidden Markov Model based smartphone heterogeneity resilient portable indoor localization framework. *J. Syst. Archit.* **2020**, *108*, 101806. [\[CrossRef\]](#)
17. Yen, H.C.; Wang, C.C. Cross-Device Wi-Fi Map Fusion with Gaussian Processes. *IEEE Trans. Mob. Comput.* **2017**, *16*, 44–57. [\[CrossRef\]](#)
18. Sadiq, S.J.; Valaee, S. Automatic Device-Transparent RSS-Based Indoor Localization. In Proceedings of the 2015 IEEE Global Communications Conference (GLOBECOM), San Diego, CA, USA, 6–10 December 2015; pp. 1–6. [\[CrossRef\]](#)
19. Alhomayani, F.; Mahoor, M.H. Deep learning methods for fingerprint-based indoor positioning: A review. *J. Locat. Based Serv.* **2020**, *14*, 129–200. [\[CrossRef\]](#)
20. Chen, L.H.; Wu, E.H.K.; Jin, M.H.; Chen, G.H. Homogeneous Features Utilization to Address the Device Heterogeneity Problem in Fingerprint Localization. *IEEE Sens. J.* **2014**, *14*, 998–1005. [\[CrossRef\]](#)
21. Zou, H.; Huang, B.; Lu, X.; Jiang, H.; Xie, L. A Robust Indoor Positioning System Based on the Procrustes Analysis and Weighted Extreme Learning Machine. *IEEE Trans. Wirel. Comms.* **2016**, *15*, 1252–1266. [\[CrossRef\]](#)
22. Yuan, Y.; Liu, X.; Liu, Z.; Xu, Z. MFMCF: A Novel Indoor Location Method Combining Multiple Fingerprints and Multiple Classifiers. In Proceedings of the 2019 3rd Intl. Symposium on Autonomous Systems (ISAS), Shanghai, China, 29–31 May 2019; pp. 216–221. [\[CrossRef\]](#)
23. Yuan, Y.; Liu, X.; Liu, Z.; He, Z.; Xu, Z. Indoor Localization With Wireless Heterogeneous Devices by Composite Fingerprint Sets and Hybrid Classification. *IEEE Trans. Veh. Technol.* **2022**, *71*, 12117–12127. [\[CrossRef\]](#)
24. Liu, W.; Cheng, Q.; Deng, Z.; Chen, H.; Fu, X.; Zheng, X.; Zheng, S.; Chen, C.; Wang, S. Survey on CSI-based Indoor Positioning Systems and Recent Advances. In Proceedings of the 2019 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Pisa, Italy, 30 September–3 October 2019; pp. 1–8. [\[CrossRef\]](#)
25. Han, Z.; Wang, Z.; Huang, H.; Zhao, L.; Su, C. WiFi-Based Indoor Positioning and Communication: Empirical Model and Theoretical Analysis. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 2364803. [\[CrossRef\]](#)
26. Torres-Sospedra, J.; Montoliu, R.; Martínez-Usó, A.; Avariento, J.P.; Arnau, T.J.; Benedito-Bordonau, M.; Huerta, J. 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), Busan, Republic of Korea, 27–30 October 2014; pp. 261–270. [\[CrossRef\]](#)
27. Roy, P.; Chowdhury, C.; Ghosh, D.; Bandyopadhyay, S. JUIndoorLoc: A Ubiquitous Framework for Smartphone-Based Indoor Localization Subject to Context and Device Heterogeneity. *Wirel. Pers. Comms.* **2019**, *106*, 739–762. [\[CrossRef\]](#)
28. Lohan, E.; Torres-Sospedra, J.; Leppäkoski, H.; Richter, P.; Peng, Z.; Huerta, J. Wi-Fi Crowdsourced Fingerprinting Dataset for Indoor Positioning. *Data* **2017**, *2*, 32. [\[CrossRef\]](#)

29. Zou, H.; Huang, B.; Lu, X.; Jiang, H.; Xie, L. Standardizing location fingerprints across heterogeneous mobile devices for indoor localization. In Proceedings of the 2016 IEEE Wireless Communications and Networking Conference, Doha, Qatar, 3–6 April 2016; pp. 1–6. [\[CrossRef\]](#)
30. Barsocchi, P.; Crivello, A.; La Rosa, D.; Palumbo, F. A multisource and multivariate dataset for indoor localization methods based on WLAN and geo-magnetic field fingerprinting. In Proceedings of the 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Alcalá de Henares, Spain, 4–7 October 2016; pp. 1–8. [\[CrossRef\]](#)
31. Moayeri, N.; Ergin, M.O.; Lemic, F.; Handziski, V.; Wolisz, A. PerfLoc (Part 1): An extensive data repository for development of smartphone indoor localization apps. In Proceedings of the 2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Valencia, Spain, 4–8 September 2016; pp. 1–7. [\[CrossRef\]](#)
32. Mendoza-Silva, G.; Richter, P.; Torres-Sospedra, J.; Lohan, E.; Huerta, J. Long-Term WiFi Fingerprinting Dataset for Research on Robust Indoor Positioning. *Data* **2018**, *3*, 3. [\[CrossRef\]](#)
33. Bi, J.; Wang, Y.; Yu, B.; Cao, H.; Shi, T.; Huang, L. Supplementary open dataset for WiFi indoor localization based on received signal strength. *Satell. Navig.* **2022**, *3*, 25. [\[CrossRef\]](#)
34. Li, S.; Tang, Z.; Kim, K.S.; Smith, J.S. 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* **2024**, *24*, 3827. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Ayub, A.; Abidin, Z.Z.; Alhammadi, A.; Soliman, N.F.; Khan, M.A.; Algarni, A.D. Comparative study of indoor positioning datasets focusing on localization accuracy success rate and floor classification. *Sci. Rep.* **2025**, *15*, 34514. [\[CrossRef\]](#)
36. Labbé, M.; Michaud, F. Multi-Session Visual SLAM for Illumination-Invariant Re-Localization in Indoor Environments. *Front. Robot. Ai* **2022**, *9*, 801886. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Gómez, C.G.; Sánchez, A.G.; Parra, N.H.; Alonso, I.P.; Kim, E. WiFi-based Localization for Fail-Aware Autonomous Driving in Urban Scenarios. In Proceedings of the 2023 IEEE Intelligent Vehicles Symposium (IV), Anchorage, AK, USA, 4–7 June 2023; pp. 1–6. [\[CrossRef\]](#)
38. Montalvo, L.; Hernández, N.; Parra, I. A Comparison of Deep Learning Architectures for WiFi-based Urban Localisation. In Proceedings of the 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), Indianapolis, IN, USA, 19–22 September 2021; pp. 122–127. [\[CrossRef\]](#)
39. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [\[CrossRef\]](#) [\[PubMed\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.