Internet of Things 34 (2025) 101753

Contents lists available at ScienceDirect

Internet of Things

journal homepage: www.elsevier.com/locate/iot

Survey of smartphone-based datasets for indoor localization: A machine learning perspective

Gaetano Carmelo La Delfa [a],*,[1], Javier Prieto [a], Salvatore Monteleone [b], Hamaad Rafique [c], Maurizio Palesi [c], Davide Patti [c]

[a] BISITE Research Group, University of Salamanca, Calle Espejo, s/n, Salamanca, 37007, Spain
[b] Engineering Department, Niccolò Cusano University, Via Don Carlo Gnocchi, 3, Rome, 00166, Italy
[c] Department of Electrical, Electronics and Computer Engineering (DIEEI), University of Catania, Via Santa Sofia, 64, Catania, 95123, Italy

ARTICLE INFO

Keywords:
Indoor localization
Datasets
Machine learning
Smartphone sensors
Indoor positioning
Deep learning
Indoor navigation

ABSTRACT

Indoor localization has gained significant attention in recent years due to its applications across sectors such as healthcare, logistics, manufacturing, and retail. However, while outdoor localization has been effectively addressed with GPS, indoor localization remains challenging despite significant research progress. Many studies have explored the capabilities of modern smartphones, equipped with a variety of sensors, to develop machine-learning methods for indoor localization, ranging from classical fingerprinting to deep sequence models and transformers. Nevertheless, most rely on small, proprietary datasets that are not publicly available. Large, high-quality public datasets are essential for researchers to efficiently test, refine, and validate algorithms, enable comparisons between different approaches and develop robust and accurate localization solutions. To reduce data collection time and costs and help researchers find the most appropriate datasets for their needs, this paper surveys 20 publicly available high-quality indoor localization datasets suitable for Machine Learning, released between 2014 and 2024, that cover various sensing technologies. The survey reveals a shift toward multi-sensor data collection, extending beyond Wi-Fi and Bluetooth signals to include inertial sensors such as accelerometers and gyroscopes, as well as magnetic fields. It also highlights that while over 75% of datasets cover multi-floor structures or multiple buildings, there is a scarcity of datasets covering diverse types of indoor environments, with most focused on office or academic settings. Moreover, the temporal dimension, crucial in dynamic indoor scenarios, remains largely underrepresented, limiting the development of ML models for tracking dynamic trajectories or adapting to evolving signal patterns.

1. Introduction

Over the past decades, outdoor localization has seen significant advancements primarily due to the ubiquity and meter-level precision of Global Positioning System (GPS) technologies [1]. Indoors, however, building structures attenuate or block satellite signals, rendering GPS practically unusable. This limitation has driven the exploration of alternative approaches for indoor localization, each presenting its own unique set of challenges [2,3].

* Corresponding author.
E-mail addresses: gaetano.ladelfa@usal.es (G.C. La Delfa), javierp@usal.es (J. Prieto), salvatore.monteleone@unicusano.it (S. Monteleone), hamaad.rafique@phd.unict.it (H. Rafique), Maurizio.palesi@unict.it (M. Palesi), davide.patti@unict.it (D. Patti).

[1] Marie curie postdoctoral fellow.

https://doi.org/10.1016/j.iot.2025.101753
Received 12 June 2025; Received in revised form 19 August 2025; Accepted 26 August 2025

Available online 30 August 2025

Indoor localization is defined as the process of determining the precise position of an object or individual within an indoor environment. Several key factors contribute to the complexity of indoor localization:

- *Multipath propagation and signal attenuation*: indoor environments contain multiple obstacles such as walls, furniture, and other structures which cause radio frequency (RF) signals to reflect, refract, and diffract, leading to multipath propagation [4]. This phenomenon affects the accuracy of indoor localization systems that rely on Wi-Fi or Bluetooth signals. Moreover, building materials, especially those used in walls and floors, absorb and weaken RF signals [5]. This attenuation varies based on the material, further complicating the indoor localization process.
- *Diverse indoor structures*: from large open spaces like malls to multi-story buildings with intricate layouts, the diversity in indoor structures demands versatile localization solutions that can adapt to different settings [6].
- *Dynamic environments*: indoor settings are dynamic, with people moving around, doors opening and closing, and changes in the environment due to factors like heating or air conditioning. A myriad of electronic devices, from Wi-Fi routers to microwaves, coexist in close proximity, adding further complexity. This dynamism contrasts with the relatively stable outdoor environment and influences the propagation and strength of signals used for localization [7].
- *Temporal magnetic field variations and ambient light interference*: factors such as electrical appliances or moving metallic objects can alter the magnetic field (MF) over time, posing challenges for magnetic fingerprinting-based indoor localization approaches. These systems require periodic re-mapping to maintain accuracy [8]. Furthermore, ambient light sources, such as sunlight or artificial lighting, can interfere with the modulated light signals used in Visible Light Communication (VLC)-based indoor localization systems, thereby degrading their performance, especially in environments with variable lighting conditions [9]. Systems that are adaptive and resilient to such environmental changes are needed.

1.1. Motivation and contribution

The massive adoption of smartphones has created new opportunities for indoor localization research, as modern smartphones are equipped with a rich array of sensors, including accelerometers, gyroscopes, magnetometers, and cameras, making them ideal platforms for developing innovative localization solutions. Additionally, the increasing market demand for precise indoor localization systems across sectors such as retail, healthcare, and logistics, has driven significant research interest in this field: according to recent industry reports [10], the indoor positioning and navigation market was valued at over USD 11.9 billion in 2024 and is expected to exceed USD 31.4 billion by 2029, reflecting a compound annual growth rate (CAGR) of more than 21% over the forecast period. On the other hand, Machine Learning (ML) approaches have been shown to be particularly efficient in addressing the complex challenges of indoor localization. These data-driven techniques can automatically adapt to dynamic indoor environments and handle the hidden relationships between various sensor inputs, leading to more robust and accurate solutions than traditional algorithmic approaches.

Numerous studies have been conducted in the field of indoor localization, along with several high-quality surveys [11]. However, to the best of the authors' knowledge, there is a lack of contributions that provide a detailed survey of high-quality datasets, with existing reviews primarily addressing localization methods or sensor technologies rather than dataset-specific attributes for ML benchmarking. This survey tries to address this gap by identifying, analyzing, and categorizing publicly available datasets that can serve as benchmarks for indoor localization research, with a particular focus on ML. In our context quality is defined by comprehensive documentation, sufficient spatial coverage, temporal granularity and sensor diversity. The primary objective of the paper is to offer a novel resource that enables direct comparisons and highlights trends in dataset evolution, assisting researchers in selecting the most appropriate datasets to support their work and advance the field. The paper surveys 20 publicly available datasets collected between 2014 and 2024. Following the principles of Systematic Literature Review (SLR) to ensure a rigorous and reproducible analysis, comprehensive searches across three digital libraries were conducted, and each dataset was evaluated against inclusion criteria such as public accessibility, sufficient documentation, adequate environmental and temporal coverage.

The remainder of this paper is organized as follows: Section 2 presents related works in indoor localization, with particular attention to ML-based and smartphone-based approaches. Section 3 discusses the role of ML in indoor localization and the importance of high-quality datasets. Section 4 provides the complete methodological framework adopted for selecting the datasets considered in this study. Sections 5 and 6 provide an in-depth evaluation of open-source datasets designed for indoor localization research, while Section 7 carries out a structured comparative analysis of the selected datasets. Finally, Section 8 concludes the paper with recommendations for future dataset development and research directions.

2. Related works

Given the aforementioned challenges, various positioning techniques have been proposed in the scientific literature. Many of these leverage the ubiquity of smartphones — with their embedded sensors and wireless connectivity — combined with state-of-the-art ML algorithms to achieve precise indoor localization, enabling a wide range of location-based services. This section reviews the existing techniques, focusing on how the different smartphone sensors commonly included in recent devices and ML approaches are used to address indoor localization challenges, categorized by sensor type as better shown in Fig. 1.

This review of the state-of-the-art will serve as a foundation for better understanding the subsequent analysis of datasets. For readers seeking a deeper understanding of the various approaches employed for each sensor type, this work refers to dedicated surveys available in the literature.

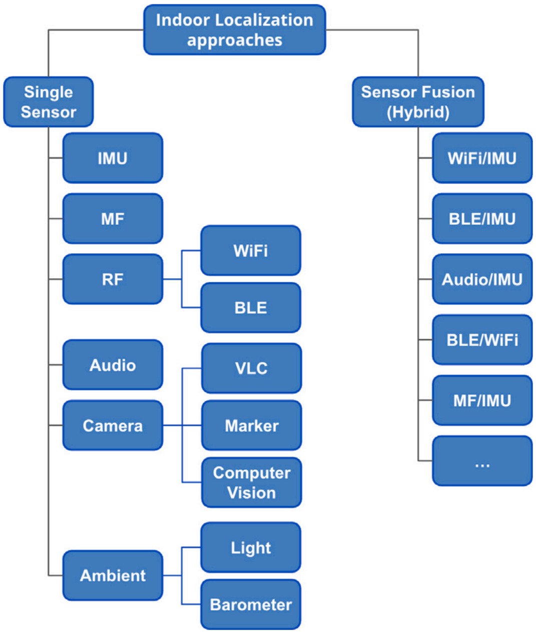

Fig. 1. Taxonomy of indoor-localization approaches. Single-sensor methods: IMU (Inertial Measurement Unit), MF, RF (WiFi, BLE [Bluetooth Low Energy]), audio, camera (VLC, markers, computer vision), ambient (light, barometer). Sensor-fusion (hybrid) configurations: combinations of two or more sensor types.

2.1. Accelerometers and gyroscopes

These motion-detecting sensors are the central part of inertial navigation systems embedded in smartphones. They facilitate the tracking of a user's movement and orientation, enabling functionalities such as path tracing and step counting. Through advanced algorithms, these sensors can calculate the user's speed and direction, significantly contributing to the enhancement of indoor localization accuracy.

In [12] Poulose et al. presented novel Kalman filter-based sensor fusion techniques for pitch and roll estimation and heading estimation, a pitch-based step detection algorithm and a position estimation algorithm, achieving significant improvements in position accuracy compared to traditional methods. Antsfeld et al. in [13] proposed a deep learning-based indoor localization framework that combines pedestrian dead reckoning (PDR) with barometer data, Wi-Fi signals, and physical environment constraints to predict a user's trajectory using smartphone sensors. The system uses a Kalman Filter for drift correction and a map-free projection method to align the predicted path with walkable paths. Convolutional and Recurrent Neural Networks (RNNs) are exploited in the pre-processing phase to learn a model of user's local displacement. In [14] Wang et al. enhanced PDR by proposing a particle swarm optimization particle filter (PSO-IPF) algorithm to improve step frequency detection, step size estimation, and direction estimation. Specifically, their approach employs accelerometer data for detecting step frequency, constructs a model for step estimation, and fuses gyroscope and magnetometer data using a Kalman filter for accurate direction estimation. The adaptation of the fitness function within the PSO process improves particle diversity and enhances the accuracy of position estimation, addressing the limitations of traditional PDR approaches in complex indoor environments.

A deeper review of the state of the art for inertial navigation approaches is presented in surveys by Panja et al. [15], and Wang et al. [16]. These surveys provide comprehensive analyses of inertial navigation systems for pedestrian localization using smartphones, discussing various algorithms, sensor fusion techniques, integration of ML approaches and challenges associated with inertial sensor-based indoor localization. The surveys also identify future research directions, providing useful guides in better understanding the research domain.

2.2. Magnetometers

Primarily serving as digital compasses, magnetometers detect fluctuations in the Earth's MF, providing essential data for orientation and outdoor navigation. This capability can also be leveraged to develop indoor positioning systems because buildings often exhibit distinct magnetic anomalies, resulting from the composition of construction materials and electrical infrastructure. Such variations in MFs create unique magnetic fingerprints that can be used to determine the user's precise position within a building in areas where traditional GPS signals fail. In the study by Yeh et al. [17] the researchers aimed to understand how environmental factors affect MF measurements from magnetometers in mobile devices. They suggested a novel approach that uses a weighted MF component combined with the k-nearest neighbors (k-NN) algorithm to enhance the accuracy of indoor positioning systems.

In [18] Ye et al. investigated the statistical behavior of MF readings in indoor settings across various smartphone models. They demonstrated that, without external interference, these readings exhibit a Gaussian distribution characterized by both time stability and spatial distinctiveness. The authors introduced Robust Locally Weighted Scatterplot Smoothing (RLOWESS) smoothing filter to eliminate the effects of noise, distortion, and outliers in the raw MF measurements. Additionally, they performed magnetometer calibration to standardize measurements across different devices. They evaluated indoor positioning performance using ML techniques and summarized the feasibility and limitations of using only MF measurement for indoor positioning. To address the low discernibility of geomagnetic features and the real-time processing requirement for large indoor areas, Zhang et al. [19] described a novel approach based on long short-term memory (LSTM) networks. The authors proposed a solution that formulates magnetic indoor localization as a recursive function approximation problem and introduced a double sliding window-based dimension expansion technique to enhance the feature dataset for LSTM analysis. Their method improved localization accuracy, making it suitable for real-time applications in complex indoor environments. An LSTM-based approach is adopted also by Vinciguerra et al. [20]. Rafique et al. in [21,22] introduced a novel Characteristics-Based Least Common Multiple (LCM) Clustering Algorithm to optimize MF-based indoor localization. The proposed method overcomes the limitations of traditional clustering techniques, such as K-Means and agglomerative clustering, which require predefined parameters and struggle with irregular cluster shapes. It autonomously determines the number and shape of clusters while accurately classifying misclassified points based on characteristic similarities. The same authors, in [23] proposed a multi-objective optimization approach to improve magnetic fingerprint-based indoor localization by optimizing the trade-off between sampling time and localization error rate on resource-constrained mobile devices, employing clustering techniques and ML classifiers. Moreover, in [24] they explored the MF measurements for indoor positioning and proposed a mini-batch magnetometer calibration technique.

Surveys [8,25] provide an in-depth analysis of the state-of-the-art MF-based indoor positioning systems, describing both the advantages and limitations of using smartphone magnetometers. These studies highlight the potential of MF data for localization while also addressing the challenges in achieving high accuracy.

2.3. Environmental sensors

Ambient light sensors and barometers, though less commonly used, provide unique inputs for indoor localization. These sensors are often employed in hybrid approaches to enhance accuracy and perform floor detection.

The variability in ambient light across different indoor environments can be detected by ambient light sensors and exploited for distinguishing between indoor locations. In the study by Sugimoto et al. [26] the authors introduced a smartphone-based indoor positioning system that integrates visible-light positioning and PDR. The system utilizes the smartphone's ambient light sensor (ALS) and inertial sensors to achieve accurate tracking by correcting position estimates through ambient light sensing. It compensates challenges such as low sampling rates of ALS, sampling rate drift, and missing samples due to interruptions in data collection by employing advanced signal processing techniques. Similarly, Sato et al. [27] proposed a novel system called ALiSA (A Visible-Light Positioning System Using the Ambient Light Sensor Assembly in a Smartphone). This system exploits the light sensors in a smartphone — typically used for automatically adjusting screen brightness — to receive signals from LED-based room lighting. By performing a Discrete Fourier Transform (DFT) analysis of the received signals, it is able to identify the source LEDs. Subsequently, the system determines the smartphone's position using trilateration based on the Received Signal Strength (RSS) values of the signals coming from the various LEDs, which are identified by their unique frequency signatures.

Barometers, which measure atmospheric pressure, can be exploited for floor detection in multi-story buildings, enhancing vertical localization. In their work, Ye et al. [28] introduced B-Loc, a novel approach that leverages the widespread availability of smartphone barometers and the relatively stable atmospheric pressure difference between floors. Using crowdsourcing techniques, B-Loc constructs a detailed barometric fingerprint map for each floor level. This system does not require any user-specific training and is designed to be energy-efficient, significantly reducing power consumption compared to floor localization methods that rely heavily on other sensors. It is able to achieve over 98% accuracy in a 10-floor building scenario. To overcome the limitations of traditional floor positioning technologies — which often depend heavily on environmental factors and have low availability — Yu et al. [29] proposed a three-step algorithm that relies on smartphones and a small number of Wi-Fi signal sources. In the first step, an "entering" detection algorithm predicts the initial floor location. Next, fluctuations in air pressure are analyzed to identify when the user is going upstairs or downstairs. Finally, vertical distances and floor changes during stairwell activity are calculated for accurate floor positioning. Potential inconsistencies between different mobile phone barometers are addressed by incorporating a neighborhood-based pressure sensor relative error calibration method into the algorithm. The survey in [30] provides a detailed discussion of advancements in VLC systems for accurate indoor positioning, highlighting key challenges and exploring future research directions in this domain.

2.4. Audio sensors

Microphones in smartphones, mainly designed for audio capture, can facilitate acoustic-based indoor localization by detecting unique soundscapes within complex environments. In the study by Rishabh et al. [31] the authors presented a non-intrusive method that leverages existing audio infrastructures, such as loudspeakers in public places or sound conditioning systems in offices, for device localization (e.g., smartphones or laptops). This is achieved by emitting low-energy, barely audible sounds from the audio infrastructure, which are then recorded and analyzed by the devices based on the time of arrival and the delays in the sound signals from different speakers. The authors utilized windowed cross-correlation techniques to enhance signal-to-noise ratio and describe two prototypical implementations: the request-play-record localizer, that requires a deliberate action from the user; and the continuous tracker, that updates device positions autonomously, without manual triggers. To overcome the low accuracy caused by the accumulated error experienced in several types of indoor navigation systems, a common technique is to recalibrate the position (hence reset the error) once the device is close to some type of beacons whose position is known. In this context, the research presented in [32] proposes using the unique noise signatures of different indoor spaces to pinpoint the room or corridor the user is in, without any additional on-site hardware, using low-cost Android devices. After recording ambient sounds of several rooms, isolating the noise and preprocessing it by performing audio frame segmentation and feature extraction, the authors trained several ML classification algorithms, including J48, KNN, Multi-layer perceptron (MLP), Naive Bayes, Support Vector Machine, AdaBoost, Random Forest, Bagging, two Stacking, and two Voting. They stated they were able to determine which rooms, hallways, entries, and meeting spaces the audio sample belongs to, achieving over 90% accuracy with several of these models.

The work by Liu et al. [33] provides a deeper review of the state of the art for acoustic-based indoor localization methods, discussing various techniques employed and the challenges they pose in indoor environments.

2.5. Radio frequency sensors

The ubiquity of Wi-Fi and Bluetooth technologies in smartphones supports RF-based localization techniques, which leverages the signal strength and proximity of known Wireless Access Points (WAPs) and Bluetooth beacons for indoor localization purposes. Fingerprinting is one of the most used approaches. It consists of creating a detailed map of signal strengths from various WAPs or Bluetooth beacons at different locations within the indoor environment during an offline phase. During the live localization phase, the measured signal strengths are compared to the recorded values to determine the user's location with a high degree of accuracy. The approach requires initial data collection and frequent updates to account for changes in the environment. Trilateration is another well-known approach, which calculates the position of a device by measuring the distances from multiple Wi-Fi access points or BLE beacons based on the received RSS values. By knowing the exact positions of the access points or beacons, the device's location can be triangulated. The accuracy of this technique is affected by obstacles and signal reflections, making it less commonly used than fingerprinting in indoor environments.

In the study by Homayounvala et al. [34], the authors proposed to leverage smartphones, ML, and Wi-Fi fingerprinting techniques for indoor localization. In the first phase, they measured and collected RSS values from Wi-Fi access points at several known reference points (RPs) using a specifically developed Android App. After a preprocessing step to standardize data, the collected dataset was used in the second phase to train a positioning model using Gaussian Process Regression (GPR), with model parameters estimated using a gradient-based algorithm. The study also includes a comparative analysis of different kernel functions used in GPR, identifying the Matérn kernel as the most effective in terms of positioning accuracy.

The paper by Khattak et al. [35] addressed the challenges of achieving accurate Wi-Fi-based indoor localization in complex environments that suffer from signal attenuation and multipath effects due to the presence of obstacles. The authors proposed a novel ML framework that adopts a preprocessing Bag-of-Features (BoF) strategy combined with K-Means clustering to transform raw RSS data into robust features. These features are then used to train a ML model. A simple KNN classifier is employed to convert the extracted test features into geographical coordinates. The proposed method is validated through both simulated environments and real-time experiments.

The research presented by Hsieh et al. [36], explored the use of RNNs and LSTM networks to enhance indoor positioning systems using Wi-Fi fingerprinting. The authors used a publicly available Long-Term Wi-Fi Fingerprinting dataset, which contains data from sensors deployed on two floors, providing coordinates based on RSS values. Multi-layer RNNs are employed for regression tasks to predict the x, y coordinates of sensors and for classification tasks to determine the floor on which each sensor is located. The study also implemented LSTM networks to mitigate long-term dependency issues common in standard RNNs, and investigated the performance of LSTM with various configurations. Experiments conducted using TensorFlow on a GPU, demonstrated that both RNN and LSTM models achieve high accuracy in indoor positioning tasks. Specifically, a 5-layer LSTM achieved 99.7% accuracy in floor classification, and distance errors for predicting the x, y coordinates were around 2.5 to 2.7 m. The study highlighted the importance of further optimization to enhance the robustness and accuracy of the system.

The paper [37] by Tsuchida et al. proposed a ML approach that utilizes the RSSI fingerprints of both Wi-Fi and Bluetooth Low Energy (BLE) signals to improve accuracy in indoor environments. In the initial site-survey phase, beacons are transmitted from mobile terminals and received by multiple receivers mounted under the ceiling. Unified RSSI fingerprints are created by capturing signal strengths at various locations, forming a radio map that exploits the strengths of both technologies. These fingerprints can then be matched with real-time measurements for positioning. In the subsequent phase, MLP neural network processes these unified RSSI fingerprints. Results showed that the unified RSSI fingerprints perform better than using Wi-Fi or BLE alone. In [38] Ventura et al. extending their previous work [39], described a BLE-based location manager block as part of GOOSE, a goal-oriented architecture that exploits semantic and location awareness to autonomously carry out complex tasks in a cloud-based ecosystem.

Surveys by Tong et al. [40] and Roy et al. [41,42] underscore the role of machine ML in enhancing RF-based indoor localization techniques and emphasize the necessity of high-quality, publicly available datasets which cover different environments to train and evaluate ML models effectively.

2.6. Cameras

Over the last decade, research has explored the potential of the high performance smartphone cameras for indoor localization purposes, and several approaches have been proposed. *VLC approaches* utilizes the built-in camera to detect modulated light signals emitted by strategically placed LEDs within the environment to determine the user's position. Such systems can achieve high localization accuracy due to the directional nature of light, low interference (as light does not penetrate walls), energy-efficiency (since LEDs are used as transmitters) and can exploit existing LED lighting infrastructure. However, they also have several disadvantages. VLC requires a direct line of sight between the transmitter (LEDs) and the receiver (smartphone camera), and obstacles blocking this line of sight can degrade performance. Additionally, their efficiency can vary with ambient lighting conditions, and the range of VLC is generally limited to the illumination area, which can restrict the scope of localization compared to RF-based systems. Finally, Setting up a VLC system for localization can be complex on both the transmitter and receiver sides.

In the study [43] Kuo et al. presented a method for indoor positioning using smartphones and slightly modified commercial LED luminaires. The system utilizes the visible light spectrum to transmit data through rapid (and human-imperceptible) on-off keying of the LEDs, which is decoded by a smartphone's camera to determine the device's position and orientation relative to the light sources. This approach leverages the rolling shutter effect [44] of CMOS cameras to capture and decode the modulated light signals in a single frame capture. The method provides the advantage of using existing smartphone hardware and, as stated by the authors, achieves decimeter-level accuracy in positioning and orientation determination. Similarly, the paper [45] by Rahman et al. introduced a VLC-based indoor positioning system that exploits the smartphone's CMOS image sensor and ML techniques to overcome the common assumption that the receiver device is parallel to the ground. The proposed system uses Linear Support Vector Machines to recognize the light-emitting diode (LED) identifiers based on a set of extracted features, providing a more efficient method than traditional coding and decoding techniques. It also integrates the smartphone's accelerometer, gyroscope, and magnetometer to further refine position accuracy. A review of the state of the art on VLC-based indoor localization approaches is presented in [9,46,47].

Visual marker-based approaches rely on strategically placed markers within an indoor environment, such as QR codes, ArUco markers [48], or AprilTag [49,50]. Typically, the smartphone's camera detects these markers, decodes the embedded information, and links it with specific coordinates or areas through a mapping database, allowing the system to pinpoint the user's position. Such systems are particularly effective in environments like airports, shopping malls, museums and warehouses, offering simplicity and versatility without the need for complex infrastructure. Despite their advantages, these systems are susceptible to lighting conditions, obstructions, and camera resolution limitations, as well as the need for periodic marker maintenance and mapping database updates.

In the study [51,52], La Delfa et al. compared the performance of three visual markers, Vuforia, ArUco [48], and AprilTag [49] suitable for real-time applications. They focused on a specific case study involving visual markers placed onto floor tiles. The authors evaluated factors such as detection speed, accuracy, and robustness under varying lighting conditions and angles, providing useful insights for each marker type. Similarly, in [53] Khan et al. presented an indoor navigation and pathfinding system using ARToolKit markers placed on the ceiling of a building and tracked by smartphone cameras. The system aims to be low-cost, simple, and easy to install, primarily designed to assist blind people, tourists, and newcomers in navigating complex indoor environments. By utilizing ceiling-mounted markers, the system reduces occlusion issues and leverages the camera of smartphones for continuous tracking. The paper [54] by Nowicki et al. proposed an indoor navigation system based on the fusion of PDR data and Wi-Fi fingerprints. It uses QR codes as landmarks to correct the estimated path based on constraints to known locations. The system utilized a factor graph where nodes represent user positions and edges represent constraints, continually adjusting user position estimation through least-squares optimization to minimize the error between the observed and expected values of the user's location derived from both QR codes and Wi-Fi signals. These and other studies [55] demonstrate the effectiveness of visual marker-based approaches in indoor localization and highlight the importance of selecting appropriate markers.

Computer vision-based approaches leverage the smartphone's camera to capture rich visual data of the surrounding environment. By comparing real-time images with a pre-built database of labeled locations, computer vision algorithms can identify distinctive features and landmarks, enabling the system to estimate the user's location. These methods can achieve high accuracy by using visual signals that are robust to changes in environmental conditions.

In a demonstration paper, Mpeis et al. [56] introduced AnyplaceCV, an indoor localization system based on computer vision technology. The system eliminates the need for external infrastructure, such as Wi-Fi networks or BLE beacons, relying instead on fingerprints generated from video recordings of indoor spaces. AnyplaceCV employed deep learning to train a model offline using video recordings of static objects, which is then used to generate a fingerprint database. A dedicated smartphone application enables users to determine their location or be continuously tracked by leveraging these fingerprints. The system, designed to enhance accuracy in challenging environments such as vessels and industrial sites, is integrated into the open-source Anyplace architecture.

DeepMoVIPS (Deep Mobile Visual Indoor Positioning System) [57] by Werner et al. describes a room-level indoor positioning system that uses the image classification capabilities of deep convolutional neural networks (CNNs). Its approach is based on AlexNet [58], pre-trained on the ImageNet dataset. Features extracted from AlexNet's fully connected layer are used to train classifiers such as Random Forests [59] and Naive Bayes [60]. The system was evaluated in an office building environment, proving its effectiveness in real-world scenarios. Another notable work by Li et al. [61] introduced a precise indoor visual positioning method utilizing smartphone cameras. The method begins by capturing a sequence of RGB images from the indoor environment, which are then used to construct a fingerprint database. This database is created through a classical SURF point-matching strategy combined with multi-image spatial forward intersection techniques. Subsequently, the relationships between SURF feature points and 3D object points are established using a novel matching error elimination approach based on Hough transform voting. Finally, the intrinsic and extrinsic parameters of the positioning images are computed using Efficient Perspective-n-Point (EPnP) and Bundle Adjustment (BA) methods, resulting in a precise determination of the smartphone's location. The SLR presented in [62] examines 68 studies on vision-based indoor navigation, categorizing them based on key aspects and summarizing their findings. The review also outlines promising research directions for future investigation.

Hybrid approaches:. previous subsections have provided an overview of key advancements in indoor localization, categorized by sensor type. However, while single-sensor approaches have shown promising results in specific contexts, the complexity of indoor environments often needs the fusion of multiple sensor modalities. Consequently, numerous studies focus on the development of hybrid approaches [63–67] that integrate multiple data sources to leverage the complementary strengths of various sensor modalities such as inertial, radio frequency, magnetic field, visual, and environmental data. These approaches, often applied in complex and dynamic indoor environments, address the limitations of individual systems, enhancing both accuracy and overall robustness. The increasing adoption of hybrid approaches has significant implications for dataset collection and curation. It highlights the need for datasets that capture synchronized readings from multiple sensors, enabling researchers to develop and evaluate fusion algorithms effectively. The subsequent sections first examine how ML techniques leverage these diverse data sources. They then introduce a framework for systematically evaluating the quality and suitability of publicly available datasets for indoor localization, demonstrating that algorithmic performance is strongly linked to dataset quality. For a more comprehensive analysis of hybrid methods and their applications, we direct the reader to dedicated surveys [68–71].

3. Machine learning for indoor localization

Over the past decade, machine learning (ML) has made significant progress in various domains, ranging from natural language processing to computer vision, speech recognition, recommender systems, and robotics [72–74]. The application of modern ML techniques has emerged as a promising solution for addressing the indoor localization problem. The complexities of indoor environments — including dynamic conditions, various types of noise, and signal interference — pose significant challenges for traditional algorithmic approaches. By leveraging data-driven techniques, ML models can learn the complex patterns and characteristics of indoor spaces, enabling precise location predictions based on sensor readings and environmental data [75–78]. Moreover, such techniques cope better with distortions, adapt to device heterogeneity, and enable continual re-training as the environment evolves. ML approaches range from classical fingerprinting baselines (e.g., k-nearest neighbors, Random Forests, Gradient Boosting) [79–84] to deep architectures that learn features (e.g., Convolutional Neural Networks, [85,86]), sequence models that exploit trajectories and timestamps (e.g., RNN/LSTM or temporal convolutional networks [87]). More recently, transformer-based architectures [88,89] use self-attention to capture long-range temporal and cross-sensor dependencies. Despite the advances, key challenges such as cross-domain generalization across buildings and devices, radio-map aging over time, label noise in ground truth, asynchronous sampling and missing data across sensors and privacy constraints that limit data sharing remain.

3.1. The role of high-quality datasets

The efficacy of any ML-based indoor localization solution is closely tied to the quality of the data it is trained on [90,91]. Datasets form the foundation on which these models are built, trained, validated, and tested. High-quality datasets offer several advantages:

1. *Robustness of the trained models*: the diversity and richness of high-quality datasets enable the training of models that do not simply memorize the training data, but instead generalize effectively to real-world contexts and make accurate predictions in unseen environments [92].
2. *Benchmarking and comparison*: common benchmark datasets allow researchers to evaluate and compare positioning algorithms under controlled conditions, facilitating reproducible research and enabling validation and incremental improvements [93, 94]. Benchmark datasets must be representative and balanced to serve as reliable performance baselines [95].
3. *Feature engineering*: high-quality datasets, especially those with rich multimodal sensor readings, facilitate effective feature engineering, feature selection, and dimensionality reduction — essential steps in enhancing model performance. Redundant or noisy features can be identified more reliably when evaluated across diverse samples and conditions. Analyzing such datasets enables researchers to determine the most informative features and uncover subtle signal patterns useful for localization.
4. *Understanding environmental dynamics*: indoor environments are inherently complex and dynamic. Datasets that capture these dynamics, such as human mobility patterns or variations in the electromagnetic landscape, provide valuable insights into localization challenges and help in designing resilient algorithms [96]. Datasets that span extended time periods or capture multiple sessions under varying conditions allow for the study of temporal variations and their impact on localization accuracy.
5. *Transfer Learning*: in scenarios where data collection is challenging or not feasible, existing datasets from diverse environments can be used to pre-train models, which are then fine-tuned on smaller, task-specific datasets. This approach has been shown to significantly improve accuracy when data in the target domain are scarce [97–99].

Creating and curating high-quality indoor localization datasets is a challenging task. Achieving fine spatial granularity and broad coverage requires extensive surveying of indoor spaces — often involving the use of precise measurement tools (e.g., laser rangefinders or total stations) to obtain accurate ground truth coordinates. Labeling each fingerprint with the correct location is time-consuming and error-prone if not carefully managed. When multiple sensor modalities are involved, synchronization and calibration become critical: sensor timestamps must be aligned, and outputs may require calibration to a shared reference frame. Heterogeneous device data (as in crowdsourced approaches) further introduce variability in sensor biases and noise, which—while beneficial for robustness—complicates the dataset collection and cleaning processes.

Ensuring repeatability represents another difficulty: a dataset collected in one building at one point in time may become outdated due to environmental changes, reducing its relevance. This necessitates either continuous dataset updates or the development of long-term datasets capturing multiple time epochs, both of which demand considerable effort. Additionally, there is a trade-off between dataset size and practicality: surveying every meter of a large facility using multiple devices to achieve high density and device diversity may be too expensive. Despite these challenges, the investment in better datasets pays off in the form of more reliable and generalizable indoor localization systems.

Given the central role of datasets in indoor localization research, the following sections present the methodology adopted for selecting the 20 publicly available datasets included in this survey. This is followed by an in-depth description of these datasets, analyzing their key characteristics, data collection procedures, and technical specifications, and discussing their respective strengths and limitations. The goal is to provide researchers with a systematic framework for dataset selection and utilization.

4. Methodology

An SLR requires the prior definition of a clear protocol that specifies the inclusion and exclusion criteria. It must ensure a comprehensive and transparent search across multiple scientific databases and be thoroughly documented to allow replication and validation by other researchers. Following SLR principles [100], this section outlines the methodological framework adopted to identify, select, and classify high-quality, publicly available indoor localization datasets.

4.1. Objective and research questions

The process began with the formulation of the research questions that guided the review and defined the main objectives of the survey:

- **RQ1**: What are the most significant publicly available indoor localization datasets released between 2014 and 2024 that meet established quality standards?
- **RQ2**: How do the identified datasets vary in terms of data collection methodology, scale, environment type, spatial and temporal coverage, sensor and device diversity?
- **RQ3**: What recurring limitations emerge across datasets, and what mitigation strategies have been proposed in the literature?

RQ1 defines the overall objective of the survey. RQ2 aim to analyze the datasets from multiple perspectives, providing a level of detail sufficient for researchers to use this survey as a practical guide when selecting datasets for their specific needs. Finally, RQ3 highlights the strengths and weaknesses of each dataset, offering insights into recurring challenges and mitigation strategies reported in the literature.

4.2. Keywords and logic search

Following the formulation of the research questions, a structured and replicable search strategy was developed to identify relevant indoor localization datasets published over the past ten years. We started with the definition of an initial set of candidate keywords based on domain knowledge (see Table 1). These keywords were iteratively refined through initial searches and analysis, leading to the **selected keywords**: *indoor localization, indoor positioning, indoor location, dataset, database, open-source*.

The search was conducted across three popular academic digital libraries (IEEE Xplore,[2] SpringerLink,[3] MDPI[4]) using boolean operators to ensure completeness and reproducibility. When datasets were mentioned in scientific papers but not directly accessible, additional investigation was carried out to determine their availability. Regarding language, only works published in English were considered.

4.3. Inclusion and exclusion criteria

Datasets were selected based on well-defined inclusion and exclusion criteria. A dataset was included in the survey if it met the following **inclusion criteria**:

1. Released between January 1, 2014, and December 31, 2024, ensuring recency and coverage.
2. Publicly downloadable or obtainable on request with a clear license (e.g., CC-BY/CC-BY-NC/custom) and documentation (readme/metadata), to ensure reproducibility and legal clarity.
3. Explicitly designed for indoor localization research.

[2] https://ieeexplore.ieee.org/Xplore/home.jsp.
[3] https://link.springer.com/.
[4] https://www.mdpi.com/.

Table 1
Possible keywords considered during the initial brain-storming phase.

Concept block	Candidate terms considered
Indoor localization	"indoor localization", "indoor positioning", "indoor navigation", "indoor tracking", "indoor location"
Data	"dataset", "database", "benchmark", "open data", "public data set", "open-source", "data"
Device scope	"smartphone", "mobile phone", "hand-held", "consumer device"
Machine-learning scope	"machine learning", "deep learning", "neural network", "artificial intelligence", "Training model", "prediction".
Sensors and modalities	"Wi-Fi", "BLE", "Bluetooth", "magnetic", "magnetometer", "IMU", "inertial", "accelerometer", "gyroscope", "UWB", "Vision", "VLC", "Light", "fingerprint", "IoT"
Exclusion terms (noise)	"outdoor", "satellite", "GNSS", "UAV", "robot-only", "RFID tag"

4. Providing sufficient spatial coverage, density of reference points, and environmental diversity to ensure statistical relevance and device/environment robustness.

The following types of datasets were excluded based on the following **exclusion criteria**:

1. Datasets that are not publicly accessible or proprietary datasets.
2. Datasets lacking clear documentation, metadata, methodological transparency or licensing information.
3. Datasets not primarily designed for indoor localization purposes.

Additionally, datasets already known to the authors from previous research activities or personal use were also considered for inclusion, as long as they fulfilled the defined inclusion criteria. After the identification and selection of datasets, a detailed analysis was carried out to assess their characteristics and suitability for indoor localization research. The analysis focused on key attributes such as sensor modalities, area coverage, data collection methodologies (e.g., static vs. time-series approaches), granularity of reference points, structure of the dataset, and a comparative evaluation was then performed.

4.4. Systematic literature review

Conducting a systematic review on datasets for indoor localization was challenging. One of the main difficulties involved the need for rigorous screening, as the selected keywords were frequently found in articles that did not actually propose or release a dataset. This led to a high volume of non-relevant results, making the identification of suitable datasets a non-trivial and time-consuming task.

4.4.1. IEEE Xplore

The following logical search string was applied to the IEEE Xplore Digital Library, targeting both titles and abstracts:

`(''indoor localization'' OR ''indoor positioning'' OR ''indoor location'') AND (dataset OR database OR ''open data'' OR ''open-source'')`

This query initially returned 1100 results. After filtering to retain only documents published in journals or conference proceedings, the number was reduced to 1077. Restricting the publication date range to the period from 2014 to 2024 further narrowed the results to 948 documents. A manual screening was then performed. Articles not written in English, short papers, and those whose primary focus was not the presentation of a dataset were excluded, in accordance with the inclusion and exclusion criteria defined above. As a result, eleven relevant datasets were identified from the IEEE repository.

4.4.2. SpringerLink

The same logical search string used for IEEE Xplore was applied to the SpringerLink repository in the *"Keywords"* field, and the following query was applied to the *"Title"* field: `(''indoor localization'' OR ''indoor positioning'' OR ''indoor location'')`

This combined search returned a total of 585 results. After applying a publication date filter to retain documents published between 2014 and 2024, the number of results was reduced to 482. Subsequent manual screening was performed following the same procedure described above. After removing duplicates and applying the defined inclusion and exclusion criteria, two datasets were ultimately selected from the SpringerLink repository.

4.4.3. MDPI

A search was conducted on the MDPI repository using the search string `(''indoor localization'' OR ''indoor positioning'' OR ''indoor location'')` in both the *Title* and *Abstract* fields. This initial query returned 1310 results. After applying a publication date filter to include only articles and conference proceedings published between 2014 and 2024, the number of results was reduced to 1235.

Due to query syntax limitations in MDPI's search interface, logical `AND` operations with dataset-related terms were applied separately. Specifically, individual refined searches were conducted by combining the base query with the following additional keywords, applied to the *Title* field: `dataset` (10 results), `database` (10 results), and `open-source` (2 results). Among these, two datasets were considered relevant and selected from these results. Additional searches applying the same logic in the *Abstract*

Table 2
Field descriptions of the UJIIndoorLoc dataset.

Fields	Content
Fields from 001 to 520	RSSI Levels of WAPs.
Fields from 521 to 523	Real-world coordinates of the sample points.
Field 524	Building ID.
Field 525	SpaceID.
Field 526	Relative position to SpaceID.
Field 527	UserID.
Field 528	PhoneID.
Field 529	Timestamp.

field returned 122, 112, and 17 results respectively. However, these did not yield any additional datasets that met the inclusion criteria, or they referred to datasets already identified. The SLR process conducted across three digital libraries (IEEE Xplore, SpringerLink, and MDPI) identified 15 datasets, supplemented by 5 datasets already known to the authors from prior research, all of which met the defined inclusion criteria, for a total of 20. This number ensures a balance between comprehensiveness—under the stated criteria at the time of the search—and focus, thereby avoiding a loss of analytical rigor due to the inclusion of lower-quality or redundant datasets. We prioritized recency (with 65% of datasets published from 2019 onward to ensure coverage of emerging trends), as well as dataset size (with 60% covering an area greater than 2000 square meters) and diversity (e.g., varied sensor modalities, environments, and device heterogeneity), since these factors make datasets more suitable for developing ML models that can generalize beyond the specific conditions of data collection. This selection strategy ensures that the survey remains both practical for researchers and reflective of evolving benchmarks.

5. Datasets for indoor localization

In the following subsections, we present the selected datasets. Each dataset is analyzed in detail, with particular attention to its technical characteristics, data collection methodology, sensor modalities, strengths, and limitations, to facilitate dataset selection for specific research needs.

5.1. UJIIndoorLoc

In [101] Torres-Sospedra et al. introduced UJIIndoorLoc, a publicly available multi-floor, multi-building, Wi-Fi-based dataset [102] released under a Creative Commons Attribution 4.0 International (CC BY 4.0) license. Created in 2014, it addressed the lack of publicly accessible datasets for comparing indoor localization methodologies, which had limited the research community's ability to evaluate and benchmark new algorithms.

UJIIndoorLoc covers an area of 108,703 square meters across three buildings, each with four or five floors depending on the building. The dataset includes 933 distinct reference points. A total of 21,049 sampled data points were captured, split into 19,938 samples for training and 1111 samples for validation and testing, the latter collected four months later. This temporal gap between the training and validation phases ensures that the validation data is independent and are not influenced by the conditions under which the training data were collected. The dataset was designed with a focus on realistic data collection. More than 20 users and 25 different devices contributed to generating the dataset. Each training reference point was assigned to at least two different users to enhance diversity. Data collection was conducted with human users physically holding the devices so as to reflect real-world scenarios more accurately — since the human body can partially block radio wave communication. This approach contrasts with other datasets where devices are often left stationary to take multiple samples. The number of different WAPs in the dataset is 520, of which 465 appear in the training set and 367 in the validation set. The dataset has been widely utilized by researchers, as demonstrated in studies such as [103–106]. Table 2 details the information included in each record of the dataset, which contains a total of 529 fields.

RSSI Levels of WAPs are the most critical part of the dataset, representing approximately 98% of the data in each record. Each RSSI level is associated with a corresponding WAP identifier (MAC Address). Since not all the WAPs are detected in each scan, the dataset assigns an artificial value of +100 dBm for undetected WAPs, ensuring that each record has a complete set of data points to maintain consistency in the dataset. The real-world coordinates of the sample points include latitude, longitude, and building's floor number. Additional field content and preprocessing details are more thoroughly described in the paper. Table 3 presents technical details of the UJIIndoorLoc dataset, while Table 4 summarizes its advantages and limitations.

5.2. SODIndoorLoc

Bi et al. [107] introduced the SODIndoorLoc dataset to overcome the limitations of existing Wi-Fi-based indoor localization datasets, which, as a result, fail to adequately capture the complexity and characteristics of modern indoor Wi-Fi environments. They identified the following common limitations:

Table 3
Technical details of the UJIIndoorLoc dataset.

Technical aspect	Details
Total area covered	108,703 m^2, encompassing 3 buildings with 4 or 5 floors each
Reference points	933
Sampled points	21,049 (19,938 for training/learning, 1111 for validation/testing)
WAPs	520
Maximum number of detected WAPs	51
Users involved in the dataset creation	20
Devices involved in the dataset creation	20 (25 considering the Android Versions)
RSSI range	[−104...0] dBm for training, [−102...−34] dBm for validation

Table 4
Advantages and limitations of the UJIIndoorLoc dataset.

Advantages	Limitations
Researchers covered an extensive area and multiple buildings, ensuring a diverse dataset.	The dataset's primary focus is on WLAN fingerprinting, hence it does not cover other aspects of indoor localization.
Data collection was performed by multiple users and devices, increasing generalizability of findings.	While the dataset spans several buildings, it is entirely gathered within a single university campus, which may limit its applicability to environments with different structural or architectural characteristics.
Data were gathered using handheld devices carried by users, thus capturing a realistic range of motion and expanding the dataset's potential applications.	

1. *Insufficient density and coverage of reference points (RPs)*: it limits the granularity of spatial information required for high-accuracy indoor positioning, thereby reducing the precision of localization algorithms.
2. *Lack of detailed information about the physical environment*: existing datasets often omit precise locations and channel information of WAPs or fail to include detailed floor plans, making algorithm development difficult.
3. *Outdated technological adaptations*: many datasets were created before significant advancements in Wi-Fi technology, such as the widespread adoption of dual-band WAPs, making them less representative of modern indoor environments.

The SODIndoorLoc dataset was collected in three distinct buildings with multiple floors, covering diverse indoor environments such as corridors, office rooms, and meeting rooms, for a total area of approximately 8000 m^2. It includes detailed information about 105 pre-installed WAPs, specifying their locations and channel information. Of these, 56 are single-band WAPs, while 49 are modern dual-band types. The dataset includes a dense distribution of 1630 RPs, with an average distance between two adjacent RPs being less than 1.2 m, ensuring fine-grained spatial resolution. Additionally, 272 testing points (TPs) were considered. The precise locations of WAPs and reference points were determined with millimeter precision using a highly accurate electronic surveying instrument. During data collection, ten participants moved sequentially from one point to another, holding smartphones directed upwards to ensure uniform data gathering. A custom-developed app recorded data at a frequency of 1 Hz, mimicking real-world pedestrian movement. Nine different Android phones were used, and anonymized data on participant heights and the height at which the phone was held during collection were provided. In total, the authors collected 23,925 samples, with 21,205 allocated for training and 2720 for testing. For training data, 30 samples were collected at each RP for single-floor buildings, while for the three-story building, a single average vector per RP was used to reflect the varying indoor environments. Testing data included ten samples for each TP. SODIndoorLoc was designed as a supplement to existing datasets such as UJIIndoorLoc. While maintaining a consistent format with its predecessor, SODIndoorLoc introduces a denser array of RPs and more detailed AP information across diverse indoor environments. This granularity allows for more precise indoor positioning and provides valuable data for algorithms that rely on a detailed understanding of Wi-Fi environments. Despite its strengths, SODIndoorLoc dataset may introduce complexity, particularly for researchers seeking simpler datasets for preliminary testing. Moreover, while it minimizes the impact of smartphone handling variations and environmental conditions on data quality, these factors can still introduce noise, potentially affecting localization accuracy. The dataset is suitable for clustering, classification, and regression analyses, offering a robust basis for evaluating the performance of indoor positioning applications that utilize Wi-Fi fingerprinting [108].

SODIndoorLoc is organized into nine training sheets and five testing sheets to cover different scenarios and requirements and facilitate a wide range of indoor positioning experiments; refer to the original paper [107] and the official repository [109] for further details. Although it is freely downloadable from the public repository, no license is provided. Table 5 presents the key features included in each datasheet, while Table 6 provides a sample to illustrate the structure.

5.3. JUIndoorLoc

To address the lack of publicly available datasets offering detailed granularity of location points across varied conditions — critical for developing and evaluating state-of-the-art indoor localization algorithms — the authors in [110] proposed the JUIndoorLoc dataset. It is accessible via a private google drive folder link, however no license is provided. A disclaimer specifies that the dataset can only be used for academic purposes. This dataset captures RSSI values of Wi-Fi signals collected from the 3rd,

Table 5
Technical details of the SODIndoorLoc dataset.

Feature	Description
MAC addresses and RSS values	Each Wi-Fi fingerprint is identified by detected MAC addresses, with corresponding RSS values in dBm.
Local coordinates (ECoord, NCoord)	Eastward and northward coordinate values for the location where the Wi-Fi fingerprint was captured, measured in meters.
Space identifiers (FloorID, BuildingID, SceneID)	Identifiers for spatial context including floor, building, and scene type.
User identifier (UserID)	Identifier for the user who collected the Wi-Fi fingerprint, associated with user-specific information like height and phone height.
Phone identifier (PhoneID)	Identifier for the smartphone used to collect the Wi-Fi fingerprint, important for understanding device-specific signal variations.
Sample times (SampleTimes)	Indicates the number of Wi-Fi samples collected at each location, ranging from 1 to 30 for training data and 1 to 10 for testing data.

Table 6
The header of a sheet from the SODIndoorLoc dataset along with a sample from the training set.

MAC1	...	MACn	ECoord	NCoord	FloorID	BuildingID	SceneID	UserID	PhoneID	SampleTimes
−44	...	100	47.4	18	1	1	1	4	3	1

4th, and 5th floors of a five-story building at Jadavpur University. Data were collected in diverse environments, including rooms, laboratories, corridors, and stairs, to ensure a rich diversity of RSSI patterns. Each floor, covering 882 m^2, was segmented into 1 m × 1 m cells, serving as unique location points identified by floor and coordinate. A total of 1000 unique location points were distributed across three floors. Four heterogeneous Android devices with different configurations were used to simulate smartphone heterogeneity in real-world scenarios. Training data were collected over a period of 31 days, while test data were gathered over five days. This distinction in duration aims to capture a broad range of data under diverse conditions for training, while a shorter, more focused period is used for testing the models. A two-month gap between the collection of training and test data accounts for the dynamic nature of indoor environments, where WAPs may be added or removed. This gap also helps identify stable APs that remain consistent over time, enhancing the reliability of the localization model. A mobile application was used to measure the RSSI of available APs from specific location points, and this data was sent to a server for analysis. The dataset includes not only the RSSI values but also the collection time, location point identifiers, device identifiers, and detailed AP features such as Basic Service Set Identifiers (BSSIDs) and Service Set Identifiers (SSIDs).

The authors observed that RSSI varies significantly, not only with different hardware but also over time at the same location, and due to various environmental factors such as obstacles and the presence of humans or other devices. Hence, collecting data from multiple perspectives is crucial to ensure the dataset's robustness. JUIndoorLoc comprises a total of 25,364 samples, with 23,904 used for training and 1460 designated for testing. The dataset is structured around 177 distinct attributes, 172 of which are RSSI values, which offer a detailed view of signal strength across various locations and conditions. It also includes a unique identifier for each device involved in data collection, and attributes indicating door status (open or closed), signal strength, the presence or absence of humans in the environment, and the specific time each data point was recorded. The captured RSSI signals range from −11 dBm to −100 dBm (for very poor signal strength), with −110 dBm used to represent missing values in the dataset. As the data is captured in raw form, preprocessing is required to remove redundant values, fill unobserved AP entries, and eliminate Wi-Fi hotspots. Numerous researchers, such as those cited in Refs. [111–113], have utilized this dataset, demonstrating its broad applicability and significance in the field. Table 7 illustrates the dataset record structure, while Table 8 presents its technical details. For further information and a deeper understanding of the JUIndoorLoc dataset, please refer to the original paper [110].

5.4. Wi-Fine

The majority of Wi-Fi-based indoor navigation approaches in scientific literature treat the localization as a discrete classification task, not taking into account the temporal and sequential nature of human movement through indoor spaces. Consequently, most existing Wi-Fi-based datasets are collected in a discrete manner, making them unsuitable for ML approaches that aim to capture the real-world dynamics of indoor navigation. To address these limitations Khassanov et al. [114] introduce a novel, publicly available (under MIT License) Wi-Fi-based dataset named *Wi-Fine*. This dataset captures sequential RSS values along 290 trajectories in the fourth, fifth, and sixth floors of a building at Nazarbayev University, covering a total area of over 9000 m^2, with fine-grained spatial and temporal resolution. As of 2021 Wi-Fine is distinguished as the largest sequential dataset in the scientific literature by the number of trajectories and collected samples. The dataset was assembled using 654 ArUco fiducial markers [48] placed along the trajectories to ensure high spatial accuracy. Operators recorded the stream of RSS values while moving along predefined continuous trajectories using a cart equipped with an HP ProBook 650 G4 notebook featuring an Intel(R) Dual Band Wireless-AC 8265 Wi-Fi module which scanned surrounding WAPs every five seconds. An external Logitech C920 HD web camera was used for marker detection. This setup facilitated precise mapping of signal strength to specific locations within the building. Operators simulated natural movement patterns, including pauses, floor changes via elevators, and occasional rotations of 90 or 180 degrees.

Table 7
Field descriptions of the JUIndoorLoc dataset.

Fields	Content
1	A unique identifier for each cell.
Fields from 2 to 173	RSSI values from Wi-Fi APs, used to estimate the device's location. Each of these 172 attributes corresponds to a unique AP.
174	Status of doors (open/closed) near each data collection point, crucial for understanding how physical barriers affect signal propagation.
175	Presence/absence of humans in the environment during data collection, which can affect signal strength due to absorption and reflection.
176	A unique identifier for the device used to collect the dataset, addressing device heterogeneity.
177	Specific timestamp when each data point was recorded, useful for analyzing temporal variations in Wi-Fi signal strength.

Table 8
Technical details of the JUIndoorLoc dataset.

Technical aspect	Details
Total area covered	882 m^2, one building, 3 out of 5 floors
Number of different WAPs	172
Total samples	25,364, with 23,904 for training and 1460 for testing.
Attributes per sample	177 attributes: 172 RSSI values, device ID, door status, signal strength, human presence, and recording time.
Number of cells for data collection	1000
Cell size	1 m × 1 m
Devices used for collection and operating system versions	Samsung Galaxy Tab 10 (Android version 4.0), Samsung Galaxy Tab E (Android version 5.0), Samsung Galaxy Tab 2 (Android version 4.1.1) and Motorola Moto E (Android version 5.1).
RSSI value range	From −11 dBm to −100 dBm, with −110 dBm for missing values
Data collection duration	120 seconds per data collection point.

Table 9
Field descriptions of the Wi-Fine dataset.

Fields	Content
Fields from 1 to 436	RSSI Levels of WAPs
Field 437	Number of attainable WAPs
Field 438	Timestamp of the recorded sample
Fields from 439 to 441	(x, y, z) position coordinates
Field 442	Floor ID

The dataset was divided into training, validation, and test sets. The training set consists of a single trajectory path per floor, covering all corridors in both directions, recorded 40 times at different days and times, resulting in 120 trajectories collected over two weeks. A total of 436 distinct WAPs were identified, with a minimum of 14 and a maximum of 95 WAPs detected per recorded sample, averaging 50. The test and validation sets include 170 trajectories, equally divided, with paths manually designed to mimic realistic walking patterns (e.g. operators were allowed to change floors via elevators). These trajectories are distinct from those in the training set.

The public availability of Wi-Fine has facilitated its adoption by researchers, as demonstrated in studies such as [115]. The dataset's record structure, comprising 442 fields, is outlined in Table 9. Tables 10 and 11 summarize key characteristics of the environment where the data was collected.

5.5. Indoor Location Competition 2.0 Dataset

The Indoor Location Competition 2.0 Dataset [116], released in 2021, is a large-scale benchmark dataset containing multi-sensor data, including Wi-Fi, BLE, Inertial Measurement Unit (IMU) readings, and gMF strengths. It also includes ground truth locations marked by accurate waypoints tied to specific timestamps, providing accurate position references. The dataset, designed to advance research in indoor localization and navigation domains, was collected using Android smartphones by surveyors who carried the devices flat in front of their bodies while a dedicated logging application recorded data in the background.

Data collection took place across numerous buildings located in various Chinese cities, covering environments such as shopping malls with multiple levels, elevators, escalators, hallways of varying dimensions, and adjacent outdoor areas.

Indoor Location Competition 2.0 Dataset includes approximately 30,000 traces from over 200 buildings, resulting in a total volume of 60 GB of data. Each trace file represents a specific path walked by the surveyor within a building and includes detailed metadata, such as collection start and end times, site and floor identifiers, and floor plan details. The trace files also document sensor-specific information, such as type, name, version, vendor, resolution, power consumption, and measuring range. It was showcased during an online indoor localization competition hosted on the Kaggle platform, attracting 1446 contestants from over 60 countries,

Table 10
Buildings, floors, WAPs and total area for the Wi-Fine dataset.

Buildings	Floors	WAPs	Area (m^2)
1	3	436	9564

Table 11
Description of train, validation and test sets for the Wi-Fine dataset.

Sets	Trajectories	Reference Points (RP)	Samples	Total length (m)	Total duration (s)
Train	120	18,975	18,975	≈31,814	112,321
Validation	85	3034	3034	≈4789	17,280
Test	85	3781	3781	≈6162	22,116
Total	290	25,790	25,790	≈42,765	151,717

Table 12
Description of fingerprint data contained in Indoor Location Competition 2.0 Dataset trace files.

Data type	Description
TYPE_ACCELEROMETER	Measures (X, Y, Z) acceleration including gravity, with a fourth value referring to the accuracy level of the accelerometer.
TYPE_ACCELEROMETER_UNCALIBRATED	Provides uncalibrated (X, Y, Z) acceleration data plus the corresponding deviation in (X, Y, Z) and the accuracy level.
TYPE_BEACON	Details on BLE beacons: UUID, major/minor IDs, transmitted and received strength, estimated distance, MAC address, last seen timestamp.
TYPE_GYROSCOPE	Captures angular velocity in X, Y, Z directions, with a fourth value referring to the accuracy level of the gyroscope.
TYPE_GYROSCOPE_UNCALIBRATED	Provides uncalibrated (X, Y, Z) angular velocity data plus the corresponding deviation in (X, Y, Z) and the accuracy level.
TYPE_ROTATION_VECTOR	Provides more accurate and reliable orientation data based on sensor fusion.
TYPE_MAGNETIC_FIELD	Records MF in X, Y, Z directions, with a fourth value referring to the accuracy level of the magnetometer.
TYPE_MAGNETIC_FIELD_UNCALIBRATED	Provides uncalibrated (X, Y, Z) MF measurements plus the corresponding deviation in (X, Y, Z) and accuracy level.
TYPE_WAYPOINT	Represents ground truth position (X, Y, Z) of the collector.
TYPE_Wi-Fi	Includes Wi-Fi information: SSID, BSSID, signal strength, frequency, and last seen timestamp.

organized into 1170 teams. The dataset has been utilized in several scientific studies such as [117,118] demonstrating its value for developing and refining localization algorithms in real-world scenarios. Released under MIT License, it constitutes a valuable resource for researchers, particularly due to its large scale, diversity of environments, extensive metadata, and comprehensive multi-modal sensor data. Table 12 illustrates the types of fingerprint data contained in each trace file.

5.6. IMUWi-Fine dataset

The IMUWi-Fine dataset [119] is designed to address the need for high-resolution, sequential data in indoor localization research, particularly for algorithms leveraging the complementary capabilities of Wi-Fi and inertial sensing. Data collection was conducted within the multi-story C4 building at Nazarbayev University — a realistic office and academic environment — by two individuals using the same commercial off-the-shelf OPPO A5 smartphone, equipped with a custom Android application developed with OpenCV. The dataset spans an aggregate distance of 14.2 km and covers the fourth, fifth, and sixth floors of a 9000 m^2 area. Ground truth positions were established utilizing 654 strategically placed ArUco fiducial markers, calibrated to a global coordinate system using a Leica TS06 Plus total station, ensuring centimeter-level accuracy. During data acquisition, users traversed pre-defined trajectories at natural walking paces while holding the smartphone, which synchronized recordings of Wi-Fi RSSI from 220 distinct WAPs and IMU readings (accelerometer, gyroscope, magnetometer) at a consistent 150 Hz sampling rate. Preprocessing steps included upsampling and linear interpolation to harmonize the sampling rates of Wi-Fi, IMU, and ground truth data. Additionally, the RSSI values underwent a specifically developed transformation, while IMU data were normalized using min–max scaling to enhance feature representation. The dataset is partitioned into training (60 trajectories), validation (30 trajectories), and test (30 trajectories) sets, comprising approximately 5.4 million samples and 1.2 GB of data. Validation and test sets were deliberately collected on different days to simulate real-world deployment scenarios and assess model generalization. IMUWi-Fine provides detailed spatiotemporal information, including (x, y, z) positions, timestamps and floor identifiers, making it a valuable benchmark for evaluating and developing advanced sequential indoor localization techniques. It is released under Creative Commons Attribution 4.0 International (CC-BY-4.0) license, and it can be downloaded in .csv format by simply contacting the authors. Table 13 summarizes the technical details of the dataset, while Table 14 provides a description of its fields. Table 15 outlines its key advantages and disadvantages for indoor localization research.

Table 13
Technical details of the IMUWi-Fine dataset.

Technical aspect	Details
Location	Nazarbayev University, Republic of Kazakhstan
Building type	Office and academic building
N° of covered floors	3
Total area	Over 9000 m^2
Trajectories	120 total trajectories. Train (60), Validation (30), Test (30) trajectories.
Total length of trajectories	Over 14 km
Sampling rate	150 Hz (synchronized for IMU, Wi-Fi RSSI, and position)
Number of APs	220 distinct APs
RSSI range	−93 dBm (weakest signal) to −34 dBm (strongest), −100 dBm (undetected signal)
Number of total samples	5.4 Million
Collection period	Approximately two weeks, different days and times
Total duration of recordings	Approximately 10.4 h
Ground truth acquisition	ArUco fiducial markers (654), Leica TS06 Plus Total Station calibrated
Smartphone	OPPO A5
ArUco markers size	14 cm × 14 cm, ArUco original dictionary for indoor localization
ArUco detection library	OpenCV ArUco library integrated into Custom-developed Android app
Data format	CSV files, structured format
Dataset size	Approximately 1.2 GB
License	Creative Commons Attribution 4.0 International License

Table 14
Data fields descriptions of the IMUWi-Fine dataset.

Fields	Content
From 1 to 220	RSSI values from 220 distinct WAPs (dBm).
221–223	Gyroscope readings along x, y, and z axes (rad/s).
224–226	Accelerometer readings along x, y, and z axes (m/s^2).
227–229	Magnetometer readings along x, y, and z axes (μT).
230	Timestamp of a recorded sample (s).
231–233	Ground truth position coordinates (m).
234	Identifier of the floor (floor_id) where the data was collected (0, 1, or 2).

Table 15
Advantages and limitations of the IMUWi-Fine dataset.

Advantages	Limitations
Captures temporal dependencies crucial for sequence learning models like RNNs.	Relies on specific hardware (OPPO A5 smartphone), which may introduce device-dependent biases.
High spatiotemporal resolution with a 150 Hz sampling rate, enabling precise tracking.	Limited to a single building and specific floors, reducing generalizability to other environments.
Combines Wi-Fi RSSI and IMU data for hybrid localization, leveraging complementary sensor modalities.	Data collection requires manual placement and calibration of fiducial markers, adding overhead.
Extensive dataset with 120 trajectories and over 14 km of walking distance.	The environment was relatively static during data collection; the dataset may not fully capture the challenges of highly dynamic environments (furniture rearrangement, people movement).

5.7. MagWi

The MagWi dataset [120], freely available at [121] (no license is provided), is a hybrid, high quality benchmark dataset designed to support indoor positioning research using sensor fusion techniques. It provides data collected over a period of approximately five years, specifically Wi-Fi signal strength, MF measurements, IMU (accelerometer, gyroscope) and barometer data from smartphones. The dataset includes recordings from five smartphones to ensure device heterogeneity (Samsung Galaxy S8, S9+, A8, LG G6, and G7), used by four different individuals (males and females) across five distinct buildings, with various smartphone orientations such as navigation, call listening, and swinging. The environments comprises multiple floors and simple and complex layouts, including corridors, halls, staircases, and uneven floor structures. This diversity allows researchers to evaluate indoor positioning techniques in dynamic environments and under varying conditions, including different levels of human mobility, multi-orientation scenarios, multi-floor navigation and spatial complexities.

The MF data are separated into static data (collected at fixed points with the surveyor standing still) and continuous data (recorded while walking along predefined paths). Static data were collected at points spaced 1 meter apart, with each location being sampled 125–150 times at a frequency of 10 Hz. Continuous data, on the other hand, were gathered at the same frequency but during user movement, allowing researchers to analyze the impact of varying walking speeds on positioning accuracy. The MagWi dataset is meticulously organized to facilitate ease of use. It is divided into two main folders: one for MF data and another for Wi-Fi

Table 16
Wi-Fi data fields descriptions of the MagWi dataset.

Fields	Content
Time	Timestamp of the data collection in the format yyyy.mm.dd hh.mm.ss.
X_pos, Y_pos	Local x-coordinate and y-coordinate of the location where the data were collected.
SSID	Service Set Identifier, representing the name of the detected Wi-Fi access point.
BSSID	Basic Service Set Identifier, a unique address for each detected access point.
RSS	Received Signal Strength for a given access point, measured in dBm.

Table 17
Magnetic data fields descriptions of the MagWi dataset.

Fields	Content
Time	Timestamp of the data collection in the format yyyy.mm.dd hh.mm.ss.
X_pos, Y_pos	Local x-coordinate and y-coordinate of the location where the data were collected (for static data).
Mag_x, Mag_y, Mag_z	MF intensity along the x-axis (North component), y-axis (East component), z-axis (Vertical component).
Acc_x, Acc_y, Acc_z	Acceleration along the x-axis, y-axis, z-axis in m/s^2.
Gyro_x, Gyro_y, Gyro_z	Angular velocity along the x-axis, y-axis, z-axis in rad/s.
Orn_x, Orn_y, Orn_z	Orientation of the device along the x-axis, y-axis, z-axis.
Pressure	Atmospheric pressure measured by the barometer in hPa.

Table 18
Field descriptions of the Multisource and Multivariate dataset.

Fields	Content
PlaceID	Unique identifier for each reference point.
AccelerationX, AccelerationY, AccelerationZ	Acceleration data along the x, y, and z axes (m/s^2).
MagneticFieldX, MagneticFieldY, MagneticFieldZ	MF data along the x, y, and z axes (μT).
Z-AxisAgle(Azimuth), X-AxisAngle(Pitch), Y-AxisAngle(Roll)	Orientation angles in degrees.
GyroX, GyroY, GyroZ	Rotational velocity around the x, y, and z axes (rad/s).
Timestamp	Time of data capture.
Wi-Fi AP Name	Service Set Identifier (SSID) of detected Wi-Fi networks.
Wi-Fi AP MAC Address	Media Access Control address of WAPs.
RSSI	127 (Campaign 1), 132 (Campaign 2).
Synchronization Method	RSSI (dBm); −100 dBm represents missing values.

data. Each folder is further structured into sub-folders categorized by building, data collection scenario, smartphone orientation, and user. Each data file is saved in an easy-to-process Excel format, with file names providing detailed metadata about the collection process. The dataset also includes data collected on staircases, enabling researchers to test algorithms for events like "stairs up" and "stairs down", as well as multi-floor transitions. Table 16 summarizes fields and descriptions for the Wi-Fi data records in the MagWi dataset, while Table 17 do the same for the MF data records (for continuous data collected while the user is walking, local x-coordinate and y-coordinate are not recorded).

The following key challenges are addressed by MagWi:

1. *Device Heterogeneity:* different models and manufacturers often result in variations in sensor sensitivity and noise levels. By including data from five distinct smartphone models, MagWi ensure robustness of positioning algorithms in real-world applications.
2. *Multi-Orientation Scenarios:* MagWi includes data collected under three common smartphone orientations: navigation mode (held in front of the user), call listening mode (held near the ear), and swinging mode (hanging freely in the user's hand). These scenarios introduce realistic variability, allowing researchers to study the impact of device orientation on positioning accuracy.
3. *Dynamic Environments:* Data were collected in both static and dynamic conditions, including locations with medium and high levels of human activity, such as a busy exhibition hall. This makes it possible to analyze the effects of human mobility, signal shadowing, and dynamic environmental changes on Wi-Fi and MF data.
4. *Multi-Floor and Uneven Structures:* the dataset provides data from multi-floor buildings and uneven floor structures, which are challenging for most positioning algorithms. By including these scenarios, MagWi supports the development and testing of floor identification and multi-level navigation algorithms.
5. *Sensor Fusion:* MagWi combines MF, Wi-Fi, and IMU data, enabling researchers to explore sensor fusion techniques for improved accuracy. The dataset can be used to combine coarse Wi-Fi-based localization with fine-grained MF positioning, leveraging the complementary nature of the two technologies.

Table 19
Technical details of the Multisource and Multivariate dataset.

Technical aspects	Details
Total area covered	185.12 m^2.
Number of measurement campaigns	2
Sampling frequency	10 Hz.
Number of reference points	325
Samples per campaign	36,795
Device used for data collection	Sony Xperia M2 smartphone and LG W110G Watch R smartwatch.
Sensor data types	Accelerometer, Gyroscope, MF, Wi-Fi RSSI.
Total number of Wi-Fi access point observations recorded during each campaign	10,850 (Campaign 1), 10,945 (Campaign 2).
Number of unique detected WAPs	127 (Campaign 1), 132 (Campaign 2).
Synchronization method	Timestamped data synchronized across smartphone and smartwatch.

Table 20
Advantages and limitations of the Multisource and Multivariate dataset.

Advantages	Limitations
Multi-sensor data facilitates hybrid indoor localization approaches which combine RSSI, geo-magnetic, and inertial data.	The dataset was collected in a single building, limiting its applicability to other architectural settings and layouts.
The dataset's uniformly spaced reference points, set at intervals of 0.6 m, provide a high granularity for testing indoor localization strategies.	Data acquisition was performed using only two devices (a smartphone and a smartwatch), which do not represent a wide range of hardware configurations.
The synchronization of smartphone and smartwatch data ensures precise temporal alignment, enabling accurate correlation of sensor readings and multi-source data analysis.	The dataset's size is modest compared to larger datasets like UJIIndoorLoc, which span multiple buildings and floors.

5.8. A multisource and multivariate dataset

To address the lack of datasets for indoor localization that integrate multiple sensor modalities, Barsocchi et al. in [122] introduced a publicly available (MIT-licensed) dataset which integrates data from multiple sources. It was collected during two measurement campaigns (each lasting one hour) conducted in a controlled 185.12 m^2 indoor environment at the National Research Council (CNR) facilities in Pisa, Italy. The campaigns employed both Sony Xperia M2 smartphones and LG W110G Watch R smartwatches running Android operating system and a dedicated app, worn simultaneously by two users, with smartphones kept at chest level and synchronized data recording. For each campaign the dataset comprises approximately 36,795 continuous samples collected at a frequency of 10 Hz and is mapped to 325 predefined reference points spaced uniformly at 60 cm intervals to ensure uniform spatial coverage and granularity.

At each point, collected data includes accelerometer readings (x, y, z axes in m/s^2), orientation angles (azimuth, pitch, roll in degrees), MF values (x, y, z axes in μT), gyroscope outputs (angular velocity in rad/s). Additionally, the dataset records Wi-Fi RSS values (in dBm), essential for fingerprinting-based indoor localization methods, and metadata (SSID, MAC address) for each WAP observed at each location. RSS values for non-detected WAPs are set to −100 dBm. During the first campaign, a total of 10,850 AP-MAC addresses were recorded, corresponding to RSSI measurements from 127 WAPs across the 325 reference points, while in the second campaign, 10,945 AP-MAC addresses were recorded, corresponding to 132 unique WAPs. On average, 10.17 unique WAPs were visible at each reference point, with the number of detectable WAPs ranging between 0 and 16, depending on proximity and environmental conditions. The RSSI data collection was synchronized with inertial sensor readings and MF measurements using timestamps and, differently from existing datasets, data from the smartphone and smartwatch are collected simultaneously. These synchronizations allow researchers to correlate RSSI variations with other sensor data and explore multi-source data fusion methodologies for indoor localization. Additionally, the indoor space includes a combination of straight corridors, connected hallways, and rooms, enabling the evaluation of algorithms in diverse architectural scenarios. Table 18 shows the structure of the dataset records as described in the repository. Table 19 resume the technical details of the dataset, while Table 20 highlight its advantages and limitations.

5.9. BLE RSSI dataset for indoor localization and navigation

The dataset presented by Mohammadi et al. [123], aimed at facilitating indoor localization and navigation in real-world environments, contains RSSI signals collected from a network of 13 BLE iBeacons separated by a distance of approximately 9.15–12.2 m. It was collected in the first floor of Waldo Library at Western Michigan University using an iPhone 6S during active library hours. It is divided into a labeled subset with 1420 records (820 for training, 600 for testing) and an unlabeled subset with 5191 records. The labeled dataset includes specific location labels, timestamps, and RSSI values from the 13 iBeacons. RSSI values, represented as negative integers, indicate proximity to iBeacons, with a lesser negative value denoting closer proximity (e.g., −65 is closer than −85, while −200 signifies an out-of-range iBeacon). Locations are precisely identified using a letter-number format (column-letter and row-number) as illustrated in Fig. 2. Their density (each cell size is about 3 × 3 m^2) ensure good spatial

Table 21
Field descriptions of the BLE RSSI dataset.

Fields	Content
Location	Denotes the reception point of RSSIs from iBeacons b3001 to b3013, using symbolic values to represent the location's column and row on a map (e.g., A01 indicates column A, row 1). See Fig. 2.
Date	Timestamps providing the specific date and time when each RSSI reading was captured, formatted as 'd-m-yyyy hh:mm:ss'.
iBeacon b3001 to iBeacon b3013	RSSI readings linked to the corresponding iBeacon, expressed as integers

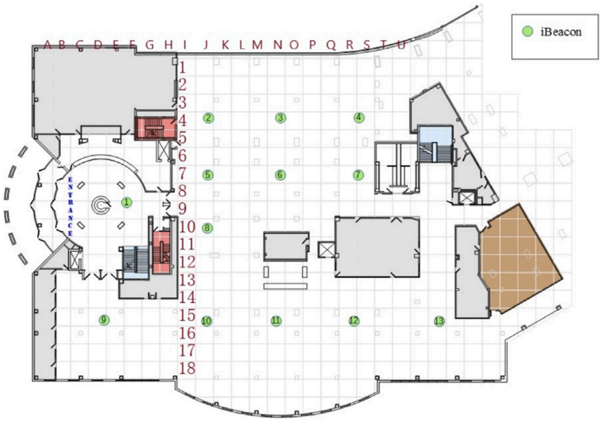

Fig. 2. iBeacons deployed inside the first floor of Waldo Library at Western Michigan University and format specifications for BLE RSSI Dataset.

Table 22
Sample of RSSI readings from multiple iBeacons at different locations and timestamps for labeled dataset.

Location	Date	b3001	b3002	b3003	b3004	b3005	b3006	...	b3012	b3013
O02	10-18-2016 11:15:21	−200	−200	−200	−200	−200	−78	...	−200	−200
P01	10-18-2016 11:15:19	−200	−200	−200	−200	−200	−78	...	−200	−200
P01	10-18-2016 11:15:17	−200	−200	−200	−200	−200	−77	...	−200	−200
P01	10-18-2016 11:15:15	−200	−200	−200	−200	−200	−77	...	−200	−200
P01	10-18-2016 11:15:13	−200	−200	−200	−200	−200	−77	...	−200	−200

granularity, which is essential for precise indoor localization within indoor settings. Additionally, the inclusion of timestamps with each reading provides insights into the temporal dynamics of indoor movement and can support the analysis of patterns over time. The use of a specific value for out-of-range iBeacons standardizes the dataset, simplifying filtering and analysis based on signal presence.

The unlabeled dataset lacks the location field but includes precise timestamps and RSSI values, enabling the study of signal strength variability over time and the inference of potential movement or usage patterns within the indoor environment. Analyzing RSSI values can reveal the distribution and consistency of iBeacon signals across the surveyed area, identifying zones of strong and weak coverage. Techniques such as k-means or hierarchical clustering can uncover natural groupings in the data, which may correlate with physical locations or environmental features. Table 21 presents the labeled dataset's record structure, and Table 22 shows the first five labeled records.

Even though the dataset is not particularly large, it represents a valuable resource for developing and testing indoor localization systems using BLE technology. In particular, the unlabeled dataset offers opportunities to advance indoor localization technologies by analyzing BLE iBeacon signal behavior in real-world settings. This fosters the development of localization solutions that do not rely on labeled data, opening the way for novel approaches in the field. The dataset is freely available at [124], licensed under the Creative Commons Attribution 4.0 International (CC BY 4.0).

5.10. UJIIndoorLoc-Mag

The UJIIndoorLoc-Mag dataset [125], introduced by Torres-Sospedra et al. was designed to address the lack of publicly available datasets for evaluating MF-based indoor localization methods. Data collection was conducted within a 260-square-meter laboratory located on the fifth floor of the Espaitec-2 building, part of the Universitat Jaume I campus, Spain. The laboratory was segmented into 8 distinct corridors and 19 intersections, with office bookcases and desktop tables delineating the corridors, creating a complex indoor environment for MF data collection. The dataset, freely available from the UCI Machine Learning Repository [126] under

Table 23
Field descriptions of the UJIIndoorLoc-Mag dataset.

Fields	Content
Timestamp	The exact time at which the sample was taken
Magnetometer X	MF value on the X-axis
Magnetometer Y	MF value on the Y-axis
Magnetometer Z	MF value on the Z-axis
Accelerometer X	Acceleration value on the X-axis, excluding the effect of gravity
Accelerometer Y	Acceleration value on the Y-axis, excluding the effect of gravity
Accelerometer Z	Acceleration value on the Z-axis, excluding the effect of gravity
Orientation sensor X	Orientation value on the X-axis
Orientation sensor Y	Orientation value on the Y-axis
Orientation sensor Z	Orientation value on the Z-axis

Table 24
Generic structure of a trajectory data file in the UJIIndoorLoc-Mag dataset. The parameters *n* and *m* represent the number of samples and segments in the trajectory, respectively.

ts_1	mx_1	my_1	mz_1	ax_1	ay_1	az_1	ox_1	oy_1	oz_1
$\vdots$									
ts_n	mx_n	my_n	mz_n	ax_n	ay_n	az_n	ox_n	oy_n	oz_n
$\langle m \rangle$									
lat_1	lon_1	lat_2	lon_2	FS_1	LS_1				
$\vdots$									
lat_m	lon_m	lat_{m+1}	lon_{m+1}	FS_m	LS_m				

a Creative Commons Attribution 4.0 International (CC BY 4.0) license, comprises 281 continuous samples (270 for training and 11 for testing). Each sample consists of discrete captures recorded along the laboratory's corridors and intersections. Data were collected at 0.1-s intervals (10 captures per second) using magnetometers, accelerometers, and orientation sensors embedded in mobile devices. The authors utilized a custom app and a Google Nexus 4 running Android 5.0.1 for training data collection. For testing, they employed both a Google Nexus 4 with Android 5.0.1 and an LG G3 Smartphone with Android 5.0, to introduce diversity. Data recording was initiated by users at the start of a path, with "turning points" marked at intersections to capture directional changes, resulting in a rich dataset comprising both linear movements along corridors and more complex movements. The training dataset consists of 54 unique paths, each sampled five times, resulting in a total of 35,779 discrete captures, while 4380 discrete captures were collected for testing. The structure of the dataset records for each trajectory is illustrated in Table 23. At the end of the dataset, the coordinates (latitude and longitude in decimal degrees) of the initial, intermediate (intersections), and final points are included, along with information about the first and last sample in the full sequence of collected samples. The scheme on Table 24 illustrates the structure of a trajectory data file. In this scheme, $\langle m \rangle$ represents the number of segments of the trajectory, *lat_i* and *lon_i* represent the latitude and longitude of the initial and final points of each segment in the trajectory, while *FSi* and *LSi* denote the first and last samples of the *i*th trajectory in the full sequence of collected data during trajectory mapping. Each sample includes a timestamp (*ts*) along with measurements from the magnetometer, accelerometer, and orientation sensors in three axes (denoted as *mx*, *my*, *mz* for the magnetometer, *ax*, *ay*, *az* for the accelerometer, and *ox*, *oy*, *oz* for the orientation sensor). To determine the position for the training data in the UJIIndoorLoc-Mag dataset, two baseline methods were developed:

- *Discrete method*: This approach is based on the premise that the positions of discrete test points, derived from continuous test samples, can be precisely estimated using the known coordinates of initial and final path points, timestamps (both recorded through the app), and the user's near-constant velocity.
- *Continuous method*: this approach segments both training and testing data into 5-s intervals, providing a more detailed representation of the user's movement and orientation over time (details of this method are less explicitly described in the paper compared to the discrete method).

The creation of the UJIIndoorLoc-Mag dataset aims to establish a standard reference for the research community, enabling fair comparisons among different indoor localization techniques. This initiative addresses the challenges of indoor localization, where GPS is unavailable, by leveraging variations in the Earth's MF caused by indoor structures and objects. The dataset has been utilized in several scientific studies, including [127], and [128].

5.11. MagPie

In [129], Hanley et al. introduced MagPIE (Magnetic Positioning Indoor Estimation), a publicly available dataset (no license is provided) designed for evaluating indoor positioning algorithms using magnetic anomalies and inertial measurements. The dataset provides magnetometer, IMU, and ground truth position and orientation measurements with centimeter-level accuracy. Data collection was conducted across three laboratory buildings using both handheld and robot-mounted sensing devices. Over 13 h of data across 51 km of total distance traveled and a total area of more than 960 square meters were collected during 723 different

Table 25
Technical details of the MagPIE dataset over all trials.

Technical aspect (Over All Trials)	Details
Data collection location	Three buildings at the University of Illinois, Urbana-Champaign campus
Total data collection time	13.96 h
Total distance covered	51.35 km
Total area covered	Over 960 square meters across three buildings
Average time of trials	90.80 s
Minimum time of trials	21.10 s
Maximum time of trials	245.66 s
Average distance of trials	101.54 m
Minimum distance of trials	23.01 m
Maximum distance of trials	297.98 m
Device used for sensing	Motorola Moto Z Play
Device used for ground truth measurements	Lenovo Phab 2
Sampling rate for magnetometer data	50 Hz
Sampling rate for accelerometer and gyroscope data	200 Hz

Table 26
Format specifications of the MAGPIE collected data.

Data type	Axis	Unit	Description
Timestamp	–	s	Timestamp for each sensor reading, expressed in seconds.
Accelerometer	x, y, z	m/s^2	Tri-axial acceleration indicating movement changes.
Gyroscope	x, y, z	rad/s	Angular velocity around each axis to track orientation changes.
Magnetometer	x, y, z	μT	MF intensity sensitive to magnetic anomalies within the buildings.
Position (Ground truth)	x, y, z	m	Precise location coordinates.
Orientation (Ground truth)	quat_x, quat_y, quat_z, quat_w	–	Quaternion values for detailed orientation information relative to the Earth's frame.

Table 27
Advantages and limitations of the MAGPie dataset.

Advantages	Limitations
Integration of 'Live Magnetic Loads' in the experimental design, accounting for temporary magnetic anomalies generated by moving objects in the test environment.	Incorporating data with 'live magnetic loads' poses a challenge in standardizing and accurately interpreting magnetic anomalies, due to the temporal and spatial variability of magnetic anomalies.
Implementation of data collection across an extensive and heterogeneous testing area, to overcome the spatial limitations of other existing datasets, often restricted to small areas in individual buildings.	The variation in data collection methodologies (handheld versus robot-mounted sensors) and environments (multiple buildings with different layouts) introduces additional complexity in data handling and algorithm development.
Data collection conducted using devices that were both handheld and mounted on a wheeled robot, allowing for diverse motion patterns and expanding the dataset's potential applications across various use cases.	

trials. Individual trials typically lasted approximately 90 s and covered 101 m. The minimum trial duration was 21 s, while the shortest distance covered was 23 m. Conversely, the most extended trial lasted about 245 s and the longest distance covered was 297 m. Data was collected in various scenarios, including office spaces and other indoor settings, with and without changes in the placement of objects affecting magnetometer readings (e.g., computer carts, industrial fans, refrigerators, microwaves, and toolboxes, collectively referred to as "live loads"). The sensors used for data collection operated at distinct sampling rates: magnetometer data at approximately 50 Hz, accelerometer and gyroscope data at approximately 200 Hz. The authors also detailed preprocessing steps, including magnetometer calibration and initial device orientation estimation.

The MagPIE dataset serves as a benchmark for multiple applications, such as magnetometer map surveying, indoor localization systems, SLAM (Simultaneous Localization and Mapping) and loop closure approaches, and improving PDR. It has been extensively used by the research community in several studies such as in [130–132]. To facilitate further analysis, Table 25 summarize the technical details of the dataset, providing a convenient reference for researchers. Each measurement modality in the MagPIE dataset is temporally synchronized via corresponding timestamps, ensuring precise correlation across all sensor types. The Table 26 present the format specifications of the collected data, while Table 27 provides an overview of dataset's advantages and limitations.

5.12. NAVER LABS datasets

Lee et al. in [133] introduce five large-scale indoor localization datasets — freely available at [134], under Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) license — that were captured in challenging real-world environments

Table 28
General information about the NAVER LABS datasets.

Dataset	Environment	Mapping images	Query images	Size (m^2)
Dept. B1	Store	22,726	10,757	8513
Dept. 1F	Store	21,600	5323	10,046
Dept. 4F	Store	12,421	7515	8348
Metro St. B1	Store & Ticket gate	23,795	17,638	20,879
Metro St. B2	Platform	6768	8240	5250

Table 29
Advantages and limitations of the NAVER LABS dataset.

Advantages	Limitations
Datasets cover large indoor spaces (up to 20,879 m^2) with over 130k images.	Environment specificity (shopping malls and metro stations), may not generalize to other indoor spaces (e.g., offices).
Robust 6DoF camera poses via automated LiDAR SLAM + SFM pipeline for ccurate ground truth.	Highly symmetric areas (e.g., metro platforms) can confuse feature-matching methods.
Includes occlusions, repetitive structures, and human presence, making it highly relevant for Real-world applications.	High crowdedness (up to 12.9% human pixels in Metro St. B1) can introduce noise and may hinder generalization.
Database and query sets captured up to 128 days apart. Such temporal variation reflects real-world scene changes.	

in Seoul, South Korea. They were collected with a dedicated mapping platform equipped with ten cameras and two 3D LiDAR scanners, enabling dense image sampling and high-precision 3D model construction. The primary objective of the datasets is to address key challenges in visual indoor localization algorithms, such as occlusions, textureless or repetitive surfaces, and varying lighting conditions. Each dataset comprises sequences captured over a four-month period across five different floors: three in a large department store and two in a metro station. Both database images (for mapping) and query images (for test and validation) are included, covering a surface area ranging from 5250 m^2 to 20,879 m^2 depending on the floor. The datasets are densely sampled, containing over 130,000 images with ground-truth camera poses obtained through a combination of LiDAR-based SLAM (to estimate initial platform trajectories and generate dense point clouds) and a Structure-from-Motion (SfM) pipeline (to refine pose accuracy by integrating feature matching with LiDAR-derived poses). The resulting 3D reconstructions contain between 1.3 million and 3.0 million 3D points, which are triangulated from 4.8 million to 22.7 million local feature correspondences, yielding a detailed and precise spatial representation. A summary of each of the five datasets, highlighting their characteristics and challenges:

- *Dept. B1*: Collected in one of the largest shopping malls in South Korea, with restaurants, supermarkets, cafés and numerous reflective surfaces, under low-light conditions.
- *Dept. 1F and Dept. 4F*: Captured on the first and fourth floors of a department store, with areas for cosmetics, jewelry, and clothing. These floors presents good lighting conditions but include a significant number of mirrors and reflective textures.
- *Metro St. B1 and Metro St. B2:* Recorded in a highly crowded metro station, encompassing dynamic elements such as moving people, trolleys, and large digital screens. The B2 level, in particular, presents strong symmetry, repetitive structures, and frequently changing advertisements, making accurate localization especially challenging.

The datasets were collected using a custom mapping platform featuring 6 industrial cameras (Basler acA2500-20gc, 2592 × 2048 resolution, 2.5 Hz, global shutter), 4 smartphone cameras (Samsung Galaxy S9, 2160 × 2880 resolution, 1 Hz, rolling shutter), 2 LiDAR scanners (Velodyne VLP-16, 360° × 30° FoV, 10 Hz rotation rate), Wheel encoders for odometry measurements. This comprehensive setup provides rich multimodal data for benchmarking indoor localization methods in diverse, real-world conditions. Table 28 shows the details of each dataset, while Table 29 summarizes their advantages and limitations. These datasets have been used in several scientific papers, such as [135,136].

6. Minor datasets

There are other datasets in the literature which are less known, less frequently used, or smaller in size. They are included here for the sake of completeness but, due to space constraints, they will be only described in a summarized way.

6.1. XJTLUIndoorLoc

XJTLUIndoorLoc [137] is a publicly available dataset collected at Xi'an Jiaotong-Liverpool University. It integrates Wi-Fi RSS and MF data, collected using two Android smartphones (Huawei P9 and MiX2) equipped with a specially developed app. The dataset was designed to support large-scale fingerprint-based localization and time-series trajectory modeling with Deep Neural Networks (DNNs), leveraging Convolutional Neural Networks (CNNs) for localization and LSTM networks for trajectory estimation. Measurements were conducted on the 4th and 5th floors of the International Business School Suzhou building, covering approximately 300 square meters. A total of 969 reference points, spaced at 60 cm intervals, were selected, and data was collected at each point in multiple device orientations. In total, 515 different WAPs were detected. To enhance the resolution of the MF maps, each interval was further subdivided into six 10-cm interval and a Clough–Tocher 2D interpolation [138] was applied, increasing the data granularity to over 31,000 virtual points. While the dataset is robust and allows for combining Wi-Fi RSS with MF data — enabling researchers to explore fingerprinting approaches using diverse data types — it also has limitations. These include discrepancies in MF values across devices, differences in manual measurements due to user height variability, and variations in the number of RSS measurements recorded by different devices. The XJTLUIndoorLoc dataset is available under MIT License on GitHub [139].

6.2. Intel Open Wi-Fi RTT dataset

The Intel Open Wi-Fi RTT Dataset [140] is a collection of nearly 30,000 Wi-Fi Round Trip Time (RTT) raw channel measurements captured in a real-world office environment using client devices and WAPs, ensuring a realistic representation of Wi-Fi RTT behavior. The dataset enables accurate distance estimation between devices and access points, facilitating precise indoor positioning and navigation. It can be used for Time of Arrival (ToA) analysis, essential for precise distance estimation in indoor settings, as well as for testing ML models and conducting other types of research in Wi-Fi indoor location field. Intel Open Wi-Fi RTT dataset is publicly available through the IEEE Dataport (Copyright (C) 2018 Intel Corporation, SPDX-License-Identifier: BSD-3-Clause) and includes MATLAB functions to aid in data visualization and range estimation. However, the dataset has certain limitations. It was collected in a specific office environment, which may limit its applicability to other settings with different layouts, building materials, or sources of interference. Additionally, data collection was performed using specific client devices and WAPs, so results may vary when using different hardware with varying RTT capabilities. Furthermore, the dataset primarily focuses on RTT measurements and does not include other potentially relevant information that could be beneficial for certain applications.

6.3. Tampere dataset

The Tampere Dataset [141] is a Wi-Fi crowdsourced fingerprinting dataset collected at Tampere University of Technology in Finland, consisting of 4648 Wi-Fi fingerprints gathered using 21 different Android devices (in different orientations) within a five-floor university building spanning 22,570 m^2. These fingerprints were collected by volunteers who moved randomly and with no main indications throughout the building, capturing diverse signal propagation patterns across different floors and environments. Each fingerprint includes local coordinates within the building, the RSS levels from all detected WAPs, the timestamp of the measurement, and the unique identifier of the Android device used. The RSS data integrates signals from multiple frequency bands (2.4 GHz and 5 GHz) without differentiation, reflecting realistic scenarios where WAP locations and frequency bands are typically unavailable. A total of 991 unique MAC addresses were recorded during the data collection campaign. The dataset is publicly available at [142] under CC Attribution 4.0 International (CC BY) license in CSV format and has been divided into training and test sets to facilitate the development and evaluation of indoor positioning algorithms. The training set comprises 15% of the fingerprints, while the test set contains the remaining 85%.

In addition to the dataset, supporting software for data reading, visualization, preprocessing, and basic benchmarking, implemented in MATLAB/Octave and Python, is provided to assist researchers in processing and analyzing the data. Its size, diversity of devices, and availability of supporting software make the Tampere dataset a valuable resource for the indoor positioning research community.

6.4. Miskolc IIS Hybrid IPS

The Miskolc IIS Hybrid IPS Dataset [143] is a multi-sensor indoor positioning dataset collected in the three-story Institute of Information Science (IIS) building at the University of Miskolc. It integrates measurements from Wi-Fi, Bluetooth, and magnetometer sensors, enabling the evaluation and development of hybrid indoor positioning systems. Data were recorded using Samsung Galaxy Young GT-S5360 devices running Android 4.4.4, and specific modules of the ILONA system [144]. The dataset covers approximately 50% of the building's 4275 m^2 area, focusing primarily on publicly accessible corridors and laboratories, where symbolic (e.g., room names) and absolute (x, y, z coordinates) positions are provided for each measurement. It includes 1571 records with 65 attributes stored in SQL and CSV formats, making it suitable for benchmarking, symbolic positioning, and ML-based algorithm development. During the measurements, 31 Wi-Fi access points and 22 Bluetooth-enabled devices were sensed. This dataset's combination of technologies and dual representation (absolute and symbolic positioning) distinguishes it from others like UJIIndoorLoc, which are limited to single-sensor data. Despite its strengths, limitations such as restricted building coverage and a static measurement environment must be considered when deploying or testing positioning algorithms. The Miskolc IIS Hybrid IPS dataset can be freely downloaded from UCI Machine Learning Repository [145]. It is licensed under the Creative Commons Attribution 4.0 International (CC BY 4.0).

Table 30
Dataset comparison. M-F = Multi-Floor (dataset covers more than one floor or building), M-S = Multi-Sensor (dataset provides more than one sensor modality), L-S = Large-Scale (dataset covers an area greater than 20,000 m^2), E-T = Extended Time (data are collected over weeks, months or years), C-D = Continuous Data (dataset includes time-series measurements along trajectories), H-D-C = Heterogeneous Data Collection (data are collected with different smartphones, devices, or users).

Dataset	M-F	M-S	L-S	E-T	C-D	H-D-C
UJIIndoorLoc	✓	×	✓	×	×	✓
SODIndoorLoc	✓	×	×	×	×	✓
JUIndoorLoc	✓	×	×	✓	×	✓
Indoor Loc. Comp. 2.0	✓	✓	✓	×	✓	✓
Wi-Fine	✓	×	×	✓	✓	×
IMUWi-Fine	✓	✓	×	✓	✓	✓
MagWi	✓	✓	–	✓	✓	✓
BLE RSSI	×	×	×	×	×	×
UJIIndoorLoc-Mag	×	✓	×	×	✓	✓
MagPIE	✓	✓	×	×	×	✓
NAVER LABS	✓	×	✓	✓	×	×
Multisource & Multivariate	×	✓	×	×	✓	✓
XJTLUIndoorLoc	✓	✓	×	×	×	✓
Intel Wi-Fi RTT	×	×	–	×	×	×
Tampere	✓	×	✓	×	×	✓
Miskolc IIS	✓	✓	×	×	×	×
OpenCSI	×	×	×	×	×	×
InLoc	✓	×	–	×	×	×
RISEdb	✓	✓	×	✓	✓	×
Continuous Long-Term	×	×	–	✓	✓	×

6.5. OpenCSI

The openCSI dataset [146] is one of the first publicly available collections of high-density Channel State Information (CSI) measurements captured in a realistic indoor environment. The dataset, released under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, was collected using a software-defined radio (SDR) mounted on a wheeled robot (to automate the radio map acquisition), in a multi-room office area (within a space of approximately 3.5 m by 5 m) at the Swisscom Digital Lab. The acquisition was done during working hours (8:00 AM to 6:00 PM) while the offices were unoccupied. The robot traversed pre-defined paths, stopping at 1 cm intervals. At each RP, an average of 1000 consecutive CSI measurements was taken, resulting in an extremely dense radio map suitable for high-resolution localization experiments. Specifically, the dataset consists of CSI extracted from the communication between the SDR and an LTE eNodeB (an Ericsson Radio Dot System) mounted onto the ceilings, which broadcasts LTE signals across four distinct antenna ports on a 20 MHz channel. Supplementary auxiliary metrics such as RSRP (Reference Signal Received Power), RSSI, RSRQ (Reference Signal Received Quality), SNR (Signal-to-Noise Ratio), CFO (Carrier Frequency Offset), noise estimate, and connection details are also collected. By making this dataset open-source, the authors provide a benchmark for researchers to investigate and compare different algorithms in fingerprint-based indoor localization without the overhead of collecting large-scale CSI maps from scratch. However, the dataset is limited by its relatively small coverage area and the fact that it was collected under controlled conditions in an unoccupied environment.

6.6. InLoc

The InLoc dataset [147] is a large database of RGB-D images that are geometrically registered to the floor maps, acquired via a high-end 3D Scanner and complemented by a separate set of RGB query photos. Specifically, the database contains approximately 10,000 perspective RGB-D images generated from 277 panoramic scans across two multi-floor buildings at Washington University in St. Louis. To simulate real-world usage, 356 additional query images were taken, using an iPhone 7 smartphone camera, at different times of the day, to captures scenes under varying conditions such as lighting changes and altered furniture layouts. They were annotated with manually verified ground-truth 6DoF camera pose in the global coordinate system of the 3D map. The setup provides realistic and challenging indoor localization scenarios, featuring textureless and symmetric areas such as walls, corridors and open spaces, repetitive structures, and significant scene changes due to dynamic elements such as furniture and human movement. The dataset can be freely downloaded from [148], and is made available under the Open Database License.

6.7. RISEdb

The RISEdb (Robust Indoor Localization in Complex Scenarios) dataset [149,150] is a dataset designed for the development and evaluation of visual indoor localization systems. It contains over one million geo-referenced images, organized into 30 sequences that cover five heterogeneous indoor environments, including office buildings, conference rooms, workshops, exhibition areas, and restaurants. The acquisition time was more than 6 h, with a total walking distance of 20.7 km. For each building, the dataset provides: (1) a high-resolution (1 cm) 3D point cloud generated with a portable SLAM-based Mobile Laser Scanning Platform

Table 31
General overview of indoor localization datasets.

Dataset name	Year	Coverage	Technology	Ref.
UJIIndoorLoc	2014	108,703 m^2 (3 multi-story university buildings)	Wi-Fi	[101,102]
SODIndoorLoc	2022	~8000 m^2 (3 buildings, various indoor environments such as corridors, offices, meeting rooms)	Wi-Fi	[107,109]
JUIndoorLoc	2019	2646 m^2 (3 floors in a university building)	Wi-Fi	[110]
Indoor Location Comp. 2.0	2021	Over 200 buildings, area not specified	Wi-Fi, BLE, IMU, MF	[116]
Wi-Fine	2021	9564 m^2 (3 floors in a university building)	Wi-Fi	[114]
IMUWi-Fine	2022	9564 m^2 (3 floors in a university building)	Wi-Fi, IMU	[119]
MagWi	2021	5 multi-floor buildings, area not specified	Wi-Fi, IMU, MF, Barometer	[120,121]
UJIIndoorLoc-Mag	2015	264 m^2 (Laboratory, with corridors and intersections)	MF, IMU	[125,126]
MagPie	2017	~960 m^2 (3 buildings in a university campus)	MF, IMU	[129]
BLE RSSI Dataset	2017	3344 m^2 (single building, library)	BLE	[124]
NAVER LABS	2021	53,036 m^2 (5 floors, 3 from a large mall plus 2 from a metro station)	Vision	[133,134]
Multisource and Multivariate	2016	185.12 m^2 (research laboratories)	IMU, MF, Wi-Fi	[122]
XJTLUIndoorLoc	2018	306 m^2 (2 floors in a university building)	Wi-Fi, MF	[137,139]
Intel Open Wi-Fi RTT	2020	Single building, area not specified	Wi-Fi RTT	[140]
Tampere	2017	22,570 m^2 (5 floors in a university building)	Wi-Fi (Crowdsourcing)	[141,142]
Miskolc IIS Hybrid IPS	2016	2137 m^2 (3 floors of a building, covered area 50%)	Wi-Fi, BLE, MF	[143,145]
OpenCSI	2022	2190 m^2 (research laboratories)	CSI	[146]
InLoc	2018	9290 m^2 (2 multi-story university buildings)	Vision	[147,148]
Continuous Long-Term	2022	Single university building, area not specified	Wi-Fi	[151,152]
RISEdb	2020	5 different environments (office, restaurant, exhibition area, workshop, conference room)	Vision, Wi-Fi, MF, IMU, Ambient Light, Temperature, phone cell towers in range	[149,150]

(MLSP) equipped with two LiDAR sensors and an inertial unit. (2) Continuous image sequences with precise ground-truth poses derived from the MLSP. These sequences contain both stereo image pairs and spherical images, ensuring full visual coverage. (3) Geo-referenced measurements from android smartphone sensors (Wi-Fi signal strengths, phone cell towers in range and noise level every 5 s, MF intensities and direction, ambient temperature, ambient light levels, and inertial sensor data at the fastest rate the device provides). The dataset addresses several limitations of several existing indoor localization datasets, particularly in terms of the scale and diversity of indoor environments, the accuracy of ground-truth poses, and the inclusion of various sensor modalities. It aims at facilitating benchmarking and training of Machine Learning and Deep Learning models for indoor localization and it is available for unrestricted download, with direct and anonymous access for anyone.

6.8. Continuous long-term dataset

The Continuous Long-Term Dataset [151], collected over two years (from 19 february 2019 to 25 march 2021) at the University of Minho in a building hosting offices, laboratories, and classrooms regularly attended by students, professors, and researchers, is a useful resource for evaluating Wi-Fi fingerprinting-based indoor positioning systems. It combines two types of data collection: continuous monitoring by a Wireless Sensor Network (WSN) composed by seven fixed monitoring devices installed in known locations (MDs), and periodic manual site surveys conducted at 49 reference points (RPs). The MDs, implemented using Raspberry Pi 3B+ devices, autonomously collected Wi-Fi signal data every 60 s, operating 24/7, and recorded parameters such as RSSI, frequency channel, and detected access points (APs). This approach resulted in over 7.4 million Wi-Fi samples from 2711 distinct WAPs, with data stored monthly to ensure efficient management and processing. The manual site surveys complemented this continuous data collection by capturing 20 Wi-Fi samples per RP at predefined intervals using a mobile unit consisting of a Raspberry Pi 3B+ mounted on a trolley, ensuring high spatial resolution across the building. All data collection was performed in a controlled environment, with precise marking of RP locations on the floor plan to guarantee accuracy. Anonymization measures were applied to protect privacy by masking MAC addresses' WAPs.

The dataset's structure is well documented, making it valuable for studying long-term radio map degradation, testing interpolation techniques, and benchmarking indoor positioning systems under real-world conditions. The dataset and extensive documentation are freely available at [152] under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, to promote reproducibility and further contributions.

7. Comparative analysis

This section presents an evaluation of the datasets surveyed in Sections 5 and 6. The analysis reveals that multi-sensor ($\geq$2 modalities) approaches are used in 9 out of 20 datasets. Approximately 70% of the selected datasets use Wi-Fi signals, as either the primary (in most cases) or secondary source of information, confirming that Wi-Fi-based indoor localization systems are the most studied within the scientific community. It further highlights the common practice of leveraging Wi-Fi signals in combination with other data sources to improve localization accuracy. In contrast, only one dataset relies exclusively on BLE signals, suggesting that BLE as a standalone method has received limited attention compared to other techniques. Instead, BLE is commonly used in hybrid approaches to enhance localization performance. About 28% of the datasets focus on MF variations caused by building structures, underscoring the relevance of MF-based methods in indoor localization, and 35% of them use IMU as a secondary source of information. Nearly all datasets consider diverse indoor environments, including corridors, open areas, offices, and laboratories. More than 75% are collected within multi-floor structures or across multiple buildings, highlighting the importance of datasets that cover a variety of scenarios to improve the robustness and generalization of ML-based indoor localization algorithms. In particular, the predominance of multi-floor settings reflects their frequency in real deployments (e.g., hospitals, shopping malls, office buildings, and university campuses). Multi-floor datasets allow researchers to examine phenomena that do not arise in single-floor layouts, such as inter-floor signal propagation, floor-identification ambiguity around open stairwells and atriums, and the sensor patterns of vertical movement (stairs, escalators, elevators). These characteristics are fundamental for safety-critical services (emergency response and evacuation guidance) and for commercial applications (floor-aware marketing and crowd analytics). The largest dataset, Indoor Location Competition 2.0, comprises 200 single and multi-floor buildings of various types and provides multimodal data, including Wi-Fi, BLE, IMU, and MF. While environmental changes (furniture movement, human presence, seasonal shifts) pose challenges for indoor localization datasets (often collected within short time frames), some of them explicitly address long-term variations and dynamic scenarios, such as JUIndoorLoc, MagWi, and Continuous Long-Term. Additionally, whereas most datasets provide snapshot-based measurements collected repeatedly (e.g., 30 samples per location), others, such as Wi-Fine, IMUWi-Fine, and MagPIE, capture continuous or sequential data along trajectories, enabling researchers to treat indoor localization as a sequential or time-series problem. Many datasets also gather data from multiple devices (e.g., different smartphone brands and models) to capture the heterogeneity of real-world usage scenarios and reduce device-specific biases. Table 31 presents a general overview of the surveyed datasets, including their name and reference, year of publication, types of sensors used, and coverage. Table 30 summarizes the datasets according to a set of high-level binary criteria, represented by checkmarks or crosses (a dash (–) indicates that the feature is not specified or the information is unavailable).

From this comparative analysis, it emerges that while some datasets are quite large and diverse, certain aspects remain underexplored. With respect to architectural diversity, many datasets focus on offices or university buildings, whereas important building categories — such as hospitals and clinical wards (characterized by dense medical equipment, intermittent electromagnetic interference, strict safety and privacy constraints), warehouses (tall shelving, sparse AP layouts, strong ferromagnetic structures), shopping centers (frequent reconfiguration of aisles, high and variable crowd density), airports (multi-level spaces and crowd "waves"), underground garages/parking facilities and various other industrial settings — are less represented. Capturing these contexts, with multi-session collections, explicit temporal labels, device diversity, and materials/occupancy metadata, would help reduce benchmark bias toward office-like layouts. Furthermore, few datasets include changes in the environments, beyond short-term furniture movement or human traffic, thus providing only partial insight into how indoor localization systems adapt to seasonal shifts or long-term infrastructural updates. Additionally, despite the availability of open repositories, dataset formats vary considerably, often requiring custom preprocessing steps and complicating reproducibility. A standardized format would be desirable.

In addition to these structural considerations, our analysis confirms that the performance of ML models for indoor localization is heavily influenced by specific dataset characteristics. For instance, datasets like JUIndoorLoc (collected over 31 days) and Continuous Long-Term (spanning 2 years) enable models to learn environmental variations, potentially leading to improved robustness against signal drift in real-world deployments at the cost of increased computational complexity during training. Conversely, datasets with limited temporal diversity, such as UJIIndoorLoc-Mag, may yield higher accuracy in controlled settings but often underperform in dynamic environments. Device heterogeneity, as evidenced in datasets such as UJIIndoorLoc and MagWi, introduces sensor noise and calibration biases that complicate model training but enhance generalization. In contrast, homogeneous datasets like IMUWi-Fine (single device) simplify training but risk overfitting. Datasets covering multi-floor/multi-building spaces (e.g., Indoor Location Competition 2.0) challenge models to handle structural variability, often benefiting from transfer learning, whereas models trained on single-building data (e.g., Multisource and Multivariate) may exhibit limited scalability. Finally, multi-sensor datasets allow fusion-based models to outperform single-modality approaches. However, they demand careful synchronization and increase computational costs.

8. Conclusion and future work

To address the lack of detailed descriptions of robust datasets for ML-based indoor localization, this paper has presented a comprehensive survey of 20 publicly available datasets spanning Wi-Fi, BLE, inertial, magnetic, and camera-based sensor modalities. We have highlighted the key features, strengths, and limitations of each dataset and provided a comparative analysis focusing on sensor technologies, data collection methodologies, environment types, and area coverage. The survey revealed that Wi-Fi-based datasets remain the most widely used. However, multi-sensor datasets are gaining popularity due to their richer representation of real-world indoor environments, which lead to improved localization accuracy. For the same reason, many of the selected datasets are collected in multi-floor and multi-building environments. Datasets are still predominantly collected in office or academic buildings, whereas more dynamic environments — such as healthcare facilities, warehouses, supermarkets, and airports — remain underrepresented. These contexts introduce non-stationarities (crowds, moving objects, layout changes, intermittent interference) that are relevant for learning and generalization. The temporal dimension is also limited: few datasets provide multi-session collections with explicit time labels, as collecting temporally rich datasets is more demanding than conducting one-off campaigns. Moreover, the survey identified data-curation challenges (incomplete documentation, heterogeneous formats, and divergent labeling schemes) that further complicate reproducibility and the direct comparison of localization algorithms. Further efforts in data standardization and unified benchmarking practices should be considered a priority by the scientific community, which should converge on a lightweight *Minimum Information for Indoor Localization Datasets schema* covering, for instance, sensors and sampling, ground-truth methods and accuracy, temporal structure, device inventory, environment metadata, privacy measures, and licensing.

In conclusion, selecting an appropriate dataset for benchmarking localization algorithms is a non-trivial task that requires careful consideration of each resource's specific characteristics and limitations (e.g., sensing modalities, temporal structure, ground truth, environmental coverage, and licensing). This survey aims to guide researchers through that process and, to the best of the authors' knowledge, constitutes one of the most comprehensive surveys of indoor localization datasets — both in the number of datasets covered and in the level of detail — currently available in the scientific literature.

CRediT authorship contribution statement

Gaetano Carmelo La Delfa: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Javier Prieto:** Supervision. **Salvatore Monteleone:** Investigation. **Hamaad Rafique:** Investigation. **Maurizio Palesi:** Supervision. **Davide Patti:** Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Gaetano Carmelo La Delfa reports financial support was provided by HORIZON-MSCA- 2022-PF-01 - Postdoctoral Fellowship. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This project has received funding from the European Union's Horizon Europe research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 101110022. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them. The work was partially supported from PNRR MUR project PE0000013-FAIR (Future Artificial Intelligence Research).

Data availability

No data was used for the research described in the article.

References

[1] P. Misra, P.E. Ganga-Jamuna, Global positioning system: signals, measurements, and performance [Book Review], IEEE Aerosp. Electron. Syst. Mag. 17 (2002) 36–37.
[2] F. Zafari, A. Gkelias, K.K. Leung, A survey of indoor localization systems and technologies, IEEE Commun. Surv. Tutor. 21 (3) (2019) 2568–2599.
[3] T.-C. Yang, A. Cabani, H. Chafouk, A survey of recent indoor localization scenarios and methodologies, Sens. (Basel Switz.) 21 (2021).
[4] Á. De-La-Llana-Calvo, J.L. Lázaro, A.G. Vicente, D. Rodríguez-Navarro, I.G. Bravo, F. Espinosa-Zapata, Characterization of multipath effects in indoor positioning systems by AoA and PoA based on optical signals, Sens. (Basel Switz.) 19 (2019).
[5] S.R. Jondhale, R.S. Deshpande, S.M. Walke, A.S. Jondhale, Issues and challenges in RSSI based target localization and tracking in wireless sensor networks, in: 2016 International Conference on Automatic Control and Dynamic Optimization Techniques, ICACDOT, 2016, pp. 594–598.
[6] J. Ni, F. Zhang, J. Xiong, Q. Huang, Z. Chang, J. Ma, B. Xie, P. Wang, G. Bian, X. Li, C. Liu, Experience: pushing indoor localization from laboratory to the wild, in: Proceedings of the 28th Annual International Conference on Mobile Computing and Networking, MobiCom '22, Association for Computing Machinery, New York, NY, USA, 2022, pp. 147–157, http://dx.doi.org/10.1145/3495243.3560546.
[7] M. Dubenova, A. Zderadickova, O. Kafka, T. Pajdla, M. Polic, D-InLoc++: Indoor localization in dynamic environments, in: Pattern Recognition: 44th DAGM German Conference, DAGM GCPR 2022, Konstanz, Germany, September 27–30, 2022, Proceedings, Springer-Verlag, Berlin, Heidelberg, 2022, pp. 246–261, http://dx.doi.org/10.1007/978-3-031-16788-1_16.
[8] G. Ouyang, K. Abed-Meraim, A survey of magnetic-field-based indoor localization, Electronics 11 (6) (2022).
[9] T.-C. Yu, W.-T. Huang, W.-B. Lee, C.-W. Chow, S.-W. Chang, H.-C. Kuo, Visible light communication system technology review: Devices, architectures, and applications, Crystals 11 (9) (2021).
[10] Research and Markets, Indoor location market size, share, growth analysis, by offering (hardware, solutions, services), technology (BLE, UWB, Wi-Fi), application, vertical (retail, healthcare & pharmaceuticals, manufacturing) and region – global forecast to 2029, 2024, URL: https://www.researchandmarkets.com/report/indoor-location, (Accessed 4 June 2025).
[11] H. Obeidat, W. Shuaieb, O. Obeidat, R. Abd-Alhameed, A review of indoor localization techniques and wireless technologies, Wirel. Pers. Commun. 119 (1) (2021) 289–327.
[12] A. Poulose, O.S. Eyobu, D.S. Han, An indoor position-estimation algorithm using smartphone IMU sensor data, IEEE Access 7 (2019) 11165–11177.
[13] L. Antsfeld, B. Chidlovskii, E. Sansano-Sansano, Deep smartphone sensors-WiFi fusion for indoor positioning and tracking, 2020, URL: https://arxiv.org/abs/2011.10799. arXiv:2011.10799.
[14] W. Wang, C. Wang, Z. Wang, X. Zhao, An improved PDR localization algorithm based on particle filter, Comput. Inform. 39 (1–2) (2020) 340–360.
[15] A.K. Panja, C. Chowdhury, S. Neogy, Survey on inertial sensor-based ILS for smartphone users, CCF Trans. Pervasive Comput. Interact. 4 (2022) 319–337.

[16] Q. Wang, M. Fu, J. Wang, H. Luo, L. Sun, Z. Ma, W. Li, C. Zhang, R. Huang, X. Li, Z. Jiang, Q. Liang, Recent advances in pedestrian inertial navigation based on smartphone: A review, IEEE Sens. J. 22 (23) (2022) 22319–22343.

[17] S.-C. Yeh, W.-H. Hsu, W.-Y. Lin, Y.-F. Wu, Study on an indoor positioning system using Earth's magnetic field, IEEE Trans. Instrum. Meas. 69 (3) (2020) 865–872.

[18] H. Ye, T. Gu, X. Tao, J. Lu, B-Loc: Scalable floor localization using barometer on smartphone, in: 2014 IEEE 11th International Conference on Mobile Ad Hoc and Sensor Systems, 2014, pp. 127–135, http://dx.doi.org/10.1109/MASS.2014.49.

[19] M. Zhang, J. Jia, J. Chen, L. Yang, L. Guo, X. Wang, Real-time indoor localization using smartphone magnetic with LSTM networks, Neural Comput. Appl. 33 (16) (2021) 10093–10110.

[20] E. Vinciguerra, E. Russo, M. Palesi, G. Ascia, H. Rafique, Improving lstm-based indoor positioning via simulation-augmented geomagnetic field dataset, in: 2024 IEEE International Conference on Mobility, Operations, Services and Technologies, MOST, IEEE, 2024, pp. 251–259, http://dx.doi.org/10.1109/MOST60774.2024.00034.

[21] H. Rafique, D. Patti, M. Palesi, G.C.L. Delfa, Characteristics-based least common multiple: A novel clustering algorithm to optimize indoor positioning, in: Proceedings of the 21st International Conference on Informatics in Control, Automation and Robotics - Volume 1: ICINCO, INSTICC, SciTePress, 2024, pp. 301–308, http://dx.doi.org/10.5220/0012943900003822.

[22] H. Rafique, D. Patti, M. Palesi, G.C. La Delfa, A characteristics-based least common multiple algorithm to optimize magnetic field-based indoor localization, IEEE Internet Things J. (2024) 1.

[23] H. Rafique, D. Patti, M. Palesi, G.C. La Delfa, V. Catania, Optimization technique for indoor localization: A multi-objective approach to sampling time and error rate trade-off, in: 2023 IEEE Third International Conference on Signal, Control and Communication, SCC, 2023, pp. 01–06, http://dx.doi.org/10.1109/SCC59637.2023.10527609.

[24] H. Rafique, D. Patti, M. Palesi, V. Catania, M-bmc: Exploration of magnetic field measurements for indoor positioning using mini-batch magnetometer calibration, in: 2023 IEEE International Conference on Mobility, Operations, Services and Technologies, MOST, IEEE, 2023, pp. 55–61, http://dx.doi.org/10.1109/MOST57249.2023.00014.

[25] I. Ashraf, Y.B. Zikria, S. Hur, Y. Park, A comprehensive analysis of magnetic field based indoor positioning with smartphones: Opportunities, challenges and practical limitations, IEEE Access 8 (2020) 228548–228571.

[26] M. Sugimoto, M. Suenaga, H. Watanabe, M. Nakamura, H. Hashizume, Smartphone indoor positioning using inertial and ambient light sensors, in: 2023 13th International Conference on Indoor Positioning and Indoor Navigation, IPIN, 2023, pp. 1–6, http://dx.doi.org/10.1109/IPIN57070.2023.10332528.

[27] T. Sato, S. Shimada, H. Murakami, H. Watanabe, H. Hashizume, M. Sugimoto, ALiSA: A visible-light positioning system using the ambient light sensor assembly in a smartphone, IEEE Sens. J. 22 (6) (2022) 4989–5000.

[28] H. Ye, T. Gu, X. Tao, J. Lu, Scalable floor localization using barometer on smartphone, Wirel. Commun. Mob. Comput. 16 (16) (2016) 2557–2571.

[29] M. Yu, F. Xue, C. Ruan, H. Guo, Floor positioning method indoors with smartphone's barometer, Geo-Spatial Inf. Sci. 22 (2) (2019) 138–148.

[30] M. Maheepala, A.Z. Kouzani, M.A. Joordens, Light-based indoor positioning systems: A review, IEEE Sens. J. 20 (8) (2020) 3971–3995.

[31] I. Rishabh, D. Kimber, J. Adcock, Indoor localization using controlled ambient sounds, in: 2012 International Conference on Indoor Positioning and Indoor Navigation, IPIN, 2012, pp. 1–10, http://dx.doi.org/10.1109/IPIN.2012.6418905.

[32] S. King, S. Pinder, D. Fernandez-Lanvin, C. González García, J. De Andrés, M. Labrador, Noise signature identification using mobile phones for indoor localization, Multimedia Tools Appl. (2024).

[33] M. Liu, L. Cheng, K. Qian, J. Wang, J. Wang, Y. Liu, Indoor acoustic localization: a survey, Hum.-Centric Comput. Inf. Sci. 10 (1) (2020).

[34] E. Homayounvala, M. Nabati, R. Shahbazian, S.A. Ghorashi, V. Moghtadaiee, A novel smartphone application for indoor positioning of users based on machine learning, in: Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers, in: UbiComp/ISWC '19 Adjunct, Association for Computing Machinery, New York, NY, USA, 2019, pp. 430–437, http://dx.doi.org/10.1145/3341162.3349300.

[35] S.B. Altaf Khattak, Fawad, M.M. Nasralla, M.A. Esmail, H. Mostafa, M. Jia, WLAN RSS-based fingerprinting for indoor localization: A machine learning inspired bag-of-features approach, Sensors 22 (14) (2022).

[36] H.-Y. Hsieh, S.W. Prakosa, J.-S. Leu, Towards the implementation of recurrent neural network schemes for WiFi fingerprint-based indoor positioning, in: 2018 IEEE 88th Vehicular Technology Conference, VTC-Fall, 2018, pp. 1–5, http://dx.doi.org/10.1109/VTCFall.2018.8690989.

[37] S. Tsuchida, T. Takahashi, S. Ibi, S. Sampei, Machine learning-aided indoor positioning based on unified fingerprints of Wi-Fi and BLE, in: 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA ASC, 2019, pp. 1468–1472, http://dx.doi.org/10.1109/APSIPAASC47483.2019.9023051.

[38] V. Catania, G.C.L. Delfa, S. Monteleone, D. Patti, D. Ventura, G.L. Torre, GOOSE: goal oriented orchestration for smart environments, Int. J. Ad Hoc Ubiquitous Comput. 32 (3) (2019) 159–170.

[39] D. Ventura, S. Monteleone, G. La Torre, G.C. La Delfa, V. Catania, Smart EDIFICE — Smart EveryDay interoperating future devICEs, in: 2015 International Conference on Collaboration Technologies and Systems, CTS, 2015, pp. 19–26, http://dx.doi.org/10.1109/CTS.2015.7210390.

[40] T. Ji, W. Li, X. Zhu, M. Liu, Survey on indoor fingerprint localization for BLE, in: 2022 IEEE 6th Information Technology and Mechatronics Engineering Conference, ITOEC, Vol. 6, 2022, pp. 129–134, http://dx.doi.org/10.1109/ITOEC53115.2022.9734528.

[41] P. Roy, C. Chowdhury, A survey of machine learning techniques for indoor localization and navigation systems, J. Intell. Robot. Syst. 101 (3) (2021).

[42] P. Roy, M.N. Anjum, A survey on ubiquitous WiFi-based indoor localization system for smartphone users from implementation perspectives, CCF Trans. Pervasive Comput. Interact. 4 (3) (2022) 298–318.

[43] Y.-S. Kuo, P. Pannuto, K.-J. Hsiao, P. Dutta, Luxapose: indoor positioning with mobile phones and visible light, in: Proceedings of the 20th Annual International Conference on Mobile Computing and Networking, MobiCom '14, Association for Computing Machinery, New York, NY, USA, 2014, pp. 447–458, http://dx.doi.org/10.1145/2639108.2639109.

[44] C. Danakis, M. Afgani, G. Povey, I. Underwood, H. Haas, Using a CMOS camera sensor for visible light communication, in: 2012 IEEE Globecom Workshops, 2012, pp. 1244–1248, http://dx.doi.org/10.1109/GLOCOMW.2012.6477759.

[45] M.H. Rahman, M.A.S. Sejan, J.-J. Kim, W.-Y. Chung, Reduced tilting effect of smartphone CMOS image sensor in visible light indoor positioning, Electronics 9 (10) (2020).

[46] P. Chen, M. Pang, D. Che, Y. Yin, D. Hu, S. Gao, A survey on visible light positioning from software algorithms to hardware, Wirel. Commun. Mob. Comput. 2021 (1) (2021) 9739577.

[47] A.B.M.M. Rahman, T. Li, Y. Wang, Recent advances in indoor localization via visible lights: A survey, Sensors 20 (5) (2020).

[48] S. Garrido-Jurado, R. Muñoz-Salinas, F. Madrid-Cuevas, M. Marín-Jiménez, Automatic generation and detection of highly reliable fiducial markers under occlusion, Pattern Recognit. 47 (6) (2014) 2280–2292.

[49] E. Olson, AprilTag: A robust and flexible visual fiducial system, in: 2011 IEEE International Conference on Robotics and Automation, 2011, pp. 3400–3407, http://dx.doi.org/10.1109/ICRA.2011.5979561.

[50] J. Wang, E. Olson, AprilTag 2: Efficient and robust fiducial detection, in: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, 2016, pp. 4193–4198, http://dx.doi.org/10.1109/IROS.2016.7759617.

[51] G.C.L. Delfa, S. Monteleone, V. Catania, J.F. de Paz, J. Bajo, Performance analysis of visualmarkers for indoor navigation systems, Front. Inf. Technol. Electron. Eng. 17 (2016) 730–740.

[52] G.C. La Delfa, V. Catania, S. Monteleone, J.F. De Paz, J. Bajo, Computer vision based indoor navigation: A visual markers evaluation, in: A. Mohamed, P. Novais, A. Pereira, G. Villarrubia González, A. Fernández-Caballero (Eds.), Ambient Intelligence - Software and Applications, Springer International Publishing, Cham, 2015, pp. 165–173, http://dx.doi.org/10.1007/978-3-319-19695-4_17.
[53] D. Khan, S. Ullah, S. Nabi, A generic approach toward indoor navigation and pathfinding with robust marker tracking, Remote. Sens. 11 (24) (2019).
[54] M. Nowicki, M. Rostkowska, P. Skrzypczyński, Indoor navigation using QR codes and WiFi signals with an implementation on mobile platform, in: 2016 Signal Processing: Algorithms, Architectures, Arrangements, and Applications, SPA, 2016, pp. 156–161, http://dx.doi.org/10.1109/SPA.2016.7763605.
[55] A. Morar, A. Moldoveanu, I. Mocanu, F. Moldoveanu, I.E. Radoi, V. Asavei, A. Gradinaru, A. Butean, A comprehensive survey of indoor localization methods based on computer vision, Sensors 20 (9) (2020).
[56] P. Mpeis, A. Hadjichristodoulou, I. Ioannides, D. Zeinalipour-Yazti, AnyplaceCV: Infrastructure-less localization in anyplace with computer vision, in: 2022 23rd IEEE International Conference on Mobile Data Management, MDM, 2022, pp. 290–291, http://dx.doi.org/10.1109/MDM55031.2022.00062.
[57] M. Werner, C. Hahn, L. Schauer, DeepMoVIPS: Visual indoor positioning using transfer learning, in: 2016 International Conference on Indoor Positioning and Indoor Navigation, IPIN, 2016, pp. 1–7, http://dx.doi.org/10.1109/IPIN.2016.7743683.
[58] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, Commun. ACM 60 (6) (2017) 84–90.
[59] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.
[60] I. Rish, An empirical study of the naive Bayes classifier, in: IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, Vol. 3, IBM New York, 2001, pp. 41–46.
[61] M. Li, R. Chen, X. Liao, B. Guo, W. Zhang, G. Guo, A precise indoor visual positioning approach using a built image feature database and single user image from smartphone cameras, Remote. Sens. 12 (5) (2020).
[62] D. Khan, Z. Cheng, H. Uchiyama, S. Ali, M. Asshad, K. Kiyokawa, Recent advances in vision-based indoor navigation: A systematic literature review, Comput. Graph. 104 (2022) 24–45.
[63] J. Yan, Z. Huang, X. Wu, Smartphone-based indoor localization using machine learning and multisource information fusion, IEEE Trans. Aerosp. Electron. Syst. 60 (3) (2024) 2722–2734.
[64] X. Kong, C. Wu, Y. You, Y. Yuan, Hybrid indoor positioning method of BLE and PDR based on adaptive feedback EKF with low BLE deployment density, IEEE Trans. Instrum. Meas. 72 (2023) 1–12.
[65] X. Kong, C. Wu, Y. You, Z. Lv, Z. Zhao, Hybrid indoor positioning method of BLE and monocular VINS-based smartphone, IEEE Trans. Instrum. Meas. 72 (2023) 1–13.
[66] M. Karmy, S. ElSayed, A.-E.H. Zekry, Performance enhancement of an indoor localization system based on visible light communication using RSSI / TDOA hybrid technique, J. Commun. 15 (5) (2020) 379–389.
[67] H. Rafique, D. Patti, M. Palesi, G.C.L. Delfa, Fusing visuals with magnetic signals to improve indoor localization using vision transformer, in: 2024 14th International Conference on Indoor Positioning and Indoor Navigation, IPIN, 2024, pp. 1–6, http://dx.doi.org/10.1109/IPIN62893.2024.10786117.
[68] X. Guo, N. Ansari, F. Hu, Y. Shao, N.R. Elikplim, L. Li, A survey on fusion-based indoor positioning, IEEE Commun. Surv. Tutor. 22 (1) (2020) 566–594.
[69] S. Traini, L. Sciullo, A. Trotta, M. Di Felice, Practical indoor localization via smartphone sensor data fusion techniques: A performance study, in: 2019 16th IEEE Annual Consumer Communications & Networking Conference, CCNC, 2019, pp. 1–7, http://dx.doi.org/10.1109/CCNC.2019.8651859.
[70] H.-Y. Huang, C.-Y. Hsieh, K.-C. Liu, H.-C. Cheng, S.J. Hsu, C.-T. Chan, Multi-sensor fusion approach for improving map-based indoor pedestrian localization, Sensors 19 (17) (2019).
[71] M. Kozlowski, R. Santos-Rodriguez, R. Piechocki, Sensor modalities and fusion for robust indoor localisation, EAI Endorsed Trans. Ambient. Syst. 19 (18) (2019).
[72] S. Tufail, H. Riggs, M. Tariq, A.I. Sarwat, Advancements and challenges in machine learning: A comprehensive review of models, libraries, applications, and algorithms, Electronics 12 (8) (2023).
[73] Z. Amiri, A. Heidari, N. Jafari, M. Hosseinzadeh, Deep study on autonomous learning techniques for complex pattern recognition in interconnected information systems, Comput. Sci. Rev. 54 (2024) 100666.
[74] Z. Amiri, A. Heidari, N.J. Navimipour, Comprehensive survey of artificial intelligence techniques and strategies for climate change mitigation, Energy 308 (2024) 132827.
[75] A. Nessa, B. Adhikari, F. Hussain, X.N. Fernando, A survey of machine learning for indoor positioning, IEEE Access 8 (2020) 214945–214965.
[76] S. Tiku, S. Pasricha, Machine Learning for Indoor Localization and Navigation, Springer International Publishing, 2023, http://dx.doi.org/10.1007/978-3-031-26712-3.
[77] S. Pasricha, AI and machine learning driven indoor localization and navigation with mobile embedded systems, 2024, http://dx.doi.org/10.48550/arXiv.2408.04797, arXiv:2408.04797.
[78] H. Rafique, D. Patti, M. Palesi, G.C. La Delfa, Indoor localization by projecting magnetic field signals onto images with vision transformer, Comput. Electr. Eng. 123 (2025) 110074.
[79] J. Hu, C. Hu, A WiFi indoor location tracking algorithm based on improved weighted K nearest neighbors and Kalman filter, IEEE Access 11 (2023) 32907–32918.
[80] X. Zheng, R. Cheng, Y. Wang, RSSI-KNN: A RSSI indoor localization approach with KNN, in: 2023 IEEE 2nd International Conference on Electrical Engineering, Big Data and Algorithms, EEBDA, 2023, pp. 600–604, http://dx.doi.org/10.1109/EEBDA56825.2023.10090664.
[81] D.J. Suroso, R. Rupaksi, A.B. Krisnawan, N.A. Siddiq, Random forest-based fingerprinting technique for device-free indoor localization system, Indones. J. Comput. Eng. Des. (IJoCED) 3 (2) (2021) 79–96.
[82] S. Lee, J. Kim, N. Moon, Random forest and WiFi fingerprint-based indoor location recognition system using smart watch, Hum.-Centric Comput. Inf. Sci. 9 (1) (2019).
[83] H. Lu, L. Zhang, H. Chen, S. Zhang, S. Wang, H. Peng, J. Zou, An indoor fingerprint positioning algorithm based on WKNN and improved XGBoost, Sensors 23 (8) (2023).
[84] C. Zheng, M.S.M. Kasihmuddin, N. Kang, X. Yan, M.A. Mansor, J. Chen, Y. Gao, Fusion and comparative study of fuzzy k-nearest neighbors, random forest, and gradient boosting for indoor positioning systems via WLAN/WiFi fingerprinting, in: 2024 IEEE International Conference on Signal, Information and Data Processing, ICSIDP, 2024, pp. 1–6, http://dx.doi.org/10.1109/ICSIDP62679.2024.10868581.
[85] N. Hernández, I. Parra, H. Corrales, R. Izquierdo, A.L. Ballardini, C. Salinas, I. García, WiFiNet: WiFi-based indoor localisation using CNNs, Expert Syst. Appl. 177 (2021) 114906.
[86] B. Zhang, H. Sifaou, G.Y. Li, CSI-fingerprinting indoor localization via attention-augmented residual convolutional neural network, Trans. Wirel. Comm. 22 (8) (2023) 5583–5597.
[87] K.A. Kordi, M. Roslee, M.Y. Alias, A. Alhammadi, A. Waseem, A.F. Osman, Survey of indoor localization based on deep learning, Comput. Mater. Contin. 79 (2) (2024) 3261–3298.
[88] W. Wu, D. Zhou, L. Shen, Z. Zhao, C. Li, G.Q. Huang, TransAoA: Transformer-based angle of arrival estimation for BLE indoor localization, IEEE Trans. Instrum. Meas. 74 (2025) 1–12.
[89] D. Gufran, S. Tiku, S. Pasricha, VITAL: Vision transformer neural networks for accurate smartphone heterogeneity resilient indoor localization, in: 2023 60th ACM/IEEE Design Automation Conference, DAC, 2023, pp. 1–6, http://dx.doi.org/10.1109/DAC56929.2023.10247684.
[90] Y. Gong, G. Liu, Y. Xue, R. Li, L. Meng, A survey on dataset quality in machine learning, Inf. Softw. Technol. 162 (2023) 107268.

[91] J. Ding, X. Li, An approach for validating quality of datasets for machine learning, in: 2018 IEEE International Conference on Big Data, Big Data, 2018, pp. 2795–2803, http://dx.doi.org/10.1109/BigData.2018.8622640.
[92] L. Li, G. Shen, C. Zhao, T. Moscibroda, J.-H. Lin, F. Zhao, Experiencing and handling the diversity in data density and environmental locality in an indoor positioning service, in: Proceedings of the 20th Annual International Conference on Mobile Computing and Networking, MobiCom '14, Association for Computing Machinery, New York, NY, USA, 2014, pp. 459–470, http://dx.doi.org/10.1145/2639108.2639118.
[93] S. Friedrich, T. Friede, On the role of benchmarking data sets and simulations in method comparison studies, Biom. J. 66 (1) (2024) 2200212.
[94] B. Koch, E. Denton, A. Hanna, J.G. Foster, Reduced, reused and recycled: The life of a dataset in machine learning research, in: Thirty-Fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2), 2021, http://dx.doi.org/10.48550/arXiv.2112.01716.
[95] S. Li, Z. Tang, K.S. Kim, J.S. Smith, 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) (2024).
[96] H. Ge, Z. Sun, Y. Chiba, N. Koshizuka, Accurate indoor location awareness based on machine learning of environmental sensing data, Comput. Electr. Eng. 98 (2022) 107676.
[97] J. Yoon, J. Oh, S. Kim, Transfer learning approach for indoor localization with small datasets, Remote. Sens. 15 (8) (2023).
[98] İ. Onat Korkmaz, T. Özateş, E. Koç, E. Aydın, E. Kor, D. Dilek, M. Alp Güngen, İ. Gökalp Köse, Ç. Akman, Indoor localization with transfer learning, in: 2022 30th Signal Processing and Communications Applications Conference, SIU, 2022, pp. 1–4, http://dx.doi.org/10.1109/SIU55565.2022.9864729.
[99] O. Kerdjidj, Y. Himeur, S.S. Sohail, A. Amira, F. Fadli, S. Atalla, W. Mansoor, A. Copiaco, A. Gawanmeh, S. Miniaoui, D.W. Dawoud, Uncovering the potential of indoor localization: Role of deep and transfer learning, IEEE Access 12 (2024) 73980–74010.
[100] C. Okoli, K. Schabram, A guide to conducting a systematic literature review of information systems research, Soc. Sci. Res. Netw. 10 (26) (2010).
[101] J. Torres-Sospedra, R. Montoliu, A. Martínez-Usó, J.P. Avariento, T.J. Arnau, M. Benedito-Bordonau, J. Huerta, UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems, in: 2014 International Conference on Indoor Positioning and Indoor Navigation, IPIN, 2014, pp. 261–270, http://dx.doi.org/10.1109/IPIN.2014.7275492.
[102] J. Torres-Sospedra, R. Montoliu, A. Martnez-Us, T. Arnau, J. Avariento, UJIIndoorLoc, UCI Machine Learning Repository, 2014, http://dx.doi.org/10.24432/C5MS59.
[103] X. Cao, Z. Yuan, X. Yang, X. Sun, X. Wang, A universal Wi-Fi fingerprint localization method based on machine learning and sample differences, Satell. Navig. 2 (2021).
[104] W. Njima, A. Bazzi, M. Chafii, DNN-based indoor localization under limited dataset using GANs and semi-supervised learning, IEEE Access 10 (2022) 69896–69909.
[105] P. Sadhukhan, K. Dahal, P.K. Das, A novel weighted fusion based efficient clustering for improved Wi-Fi fingerprint indoor positioning, IEEE Trans. Wirel. Commun. 22 (7) (2023) 4461–4474.
[106] N. Agah, B. Evans, X. Meng, H. Xu, A local machine learning approach for fingerprint-based indoor localization, in: SoutheastCon 2023, IEEE, 2023, pp. 240–245, http://dx.doi.org/10.48550/arXiv.2302.10810.
[107] J. Bi, Y. Wang, B. Yu, H. Cao, T. Shi, L. Huang, Supplementary open dataset for WiFi indoor localization based on received signal strength, Satell. Navig. 3 (2022) 1–15.
[108] O. Kerdjidj, Y. Himeur, S. Atalla, A. Copiac, S.S. Sohail, F. Fadli, A. Amira, W. Mansoor, A. Gawanmeh, Exploring 2D representation and transfer learning techniques for people identification in indoor localization, in: 2023 6th International Conference on Signal Processing and Information Security, ICSPIS, 2023, pp. 173–177, http://dx.doi.org/10.1109/ICSPIS60075.2023.10343825.
[109] renwudao24, SODIndoorLoc: Supplementary open dataset for WiFi fingerprint-based indoor localization, 2023, URL: https://github.com/renwudao24/SODIndoorLoc.
[110] P. Roy, C. Chowdhury, D. Ghosh, S. Bandyopadhyay, JUIndoorLoc: A ubiquitous framework for smartphone-based indoor localization subject to context and device heterogeneity, Wirel. Pers. Commun. 106 (2) (2019) 739–762.
[111] P. Roy, C. Chowdhury, Designing an ensemble of classifiers for smartphone-based indoor localization irrespective of device configuration, Multimedia Tools Appl. 80 (13) (2021) 20501–20525.
[112] S. Parsuramka, A.K. Panja, P. Roy, S. Neogy, C. Chowdhury, FABEL: feature association based ensemble learning for positioning in indoor environment, Multimedia Tools Appl. 82 (5) (2023) 7247–7266.
[113] A.T. Abraha, B. Wang, RiLoc: Representative guided subarea to exact localization from crowdsourced samples, in: 2022 IEEE Smartworld, Ubiquitous Intelligence & Computing, Scalable Computing & Communications, Digital Twin, Privacy Computing, Metaverse, Autonomous & Trusted Vehicles, SmartWorld/UIC/ScalCom/DigitalTwin/PriComp/Meta, 2022, pp. 144–151, http://dx.doi.org/10.1109/SmartWorld-UIC-ATC-ScalCom-DigitalTwin-PriComp-Metaverse56740.2022.00046.
[114] Y. Khassanov, M. Nurpeiissov, A. Sarkytbayev, A. Kuzdeuov, H.A. Varol, Finer-level sequential WiFi-based indoor localization, in: 2021 IEEE/SICE International Symposium on System Integration, SII, 2021, pp. 163–169, http://dx.doi.org/10.1109/IEEECONF49454.2021.9382623.
[115] M. Liu, X. Liao, Z. Gao, Q. Li, FT-Loc: A fine-grained temporal features-based fusion network for indoor localization, IEEE Internet Things J. 11 (3) (2024) 4324–4334.
[116] Y. Shu, Q. Xu, J. Liu, R.R. Choudhury, N. Trigoni, V. Bahl, Indoor location competition 2.0 dataset, 2021, URL: https://www.microsoft.com/en-us/research/publication/indoor-location-competition-2-0-dataset/.
[117] M. Zhang, Z. Fan, R. Shibasaki, X. Song, Domain adversarial graph convolutional network based on RSSI and crowdsensing for indoor localization, IEEE Internet Things J. 10 (15) (2023) 13662–13672.
[118] A.T. Abraha, B. Wang, Z. Yu, J. He, OBLoc: Online batch localization for large-scale indoor environments, IEEE Syst. J. 17 (4) (2023) 6552–6563.
[119] M. Nurpeiissov, A. Kuzdeuov, A. Assylkhanov, Y. Khassanov, H.A. Varol, End-to-end sequential indoor localization using smartphone inertial sensors and WiFi, in: 2022 IEEE/SICE International Symposium on System Integration, SII, 2022, pp. 566–571, http://dx.doi.org/10.1109/SII52469.2022.9708854.
[120] I. Ashraf, S. Din, M. Ali, S. Hur, Y. Zikria, Y. Park, MagWi: Benchmark dataset for long term magnetic field and Wi-Fi data involving heterogeneous smartphones, multiple orientations, spatial diversity and multi-floor buildings, IEEE Access 9 (2021) 77976–77996.
[121] I. Ashraf, S. Din, M.U. Ali, S. Hur, Y.B. Zikria, Y. Park, MagWi: Benchmark dataset for long term magnetic field and Wi-Fi data involving heterogeneous smartphones, multiple orientations, spatial diversity and multi-floor buildings, 2021, http://dx.doi.org/10.21227/7g9v-6z48, [Dataset].
[122] P. Barsocchi, A. Crivello, D. La Rosa, F. Palumbo, A multisource and multivariate dataset for indoor localization methods based on WLAN and geo-magnetic field fingerprinting, in: 2016 International Conference on Indoor Positioning and Indoor Navigation, IPIN, 2016, pp. 1–8, http://dx.doi.org/10.1109/IPIN.2016.7743678.
[123] M. Mohammadi, A. Al-Fuqaha, M. Guizani, J.-S. Oh, Semisupervised deep reinforcement learning in support of IoT and smart city services, IEEE Internet Things J. 5 (2) (2018) 624–635.
[124] M. Mohammadi, A. Al-Fuqaha, BLE RSSI Dataset for Indoor Localization and Navigation, UCI Machine Learning Repository, 2017, http://dx.doi.org/10.24432/C54G80.
[125] J. Torres-Sospedra, D. Rambla, R. Montoliu, O. Belmonte, J. Huerta, UJIIndoorLoc-Mag: A new database for magnetic field-based localization problems, in: 2015 International Conference on Indoor Positioning and Indoor Navigation, IPIN, 2015, pp. 1–10, http://dx.doi.org/10.1109/IPIN.2015.7346763.
[126] J. Torres Sospedra, D. Rambla, R. Montoliu, O. Belmonte, J. Huerta, UJIIndoorLoc-Mag, UCI Machine Learning Repository, 2015, http://dx.doi.org/10.24432/C5D311.

[127] Y. Chen, M. Zhou, Z. Zheng, Learning sequence-based fingerprint for magnetic indoor positioning system, IEEE Access 7 (2019) 163231–163244.

[128] M.W.P. Maduranga, H.K.I.S. Lakmal, L.P. Kalansooriya, Three-dimensional wireless indoor localization with machine learning algorithms for location-based IoT applications, in: D. Herath, S. Date, U. Jayasinghe, V. Narayanan, R. Ragel, J. Wang (Eds.), Asia Pacific Advanced Network, Springer Nature Switzerland, Cham, 2024, pp. 1–12, http://dx.doi.org/10.1007/978-3-031-51135-6_1.

[129] D. Hanley, A.B. Faustino, S.D. Zelman, D.A. Degenhardt, T. Bretl, MagPIE: A dataset for indoor positioning with magnetic anomalies, in: 2017 International Conference on Indoor Positioning and Indoor Navigation, IPIN, 2017, pp. 1–8, http://dx.doi.org/10.1109/IPIN.2017.8115961.

[130] L. Wang, H. Luo, Q. Wang, W. Shao, F. Zhao, A hierarchical LSTM-based indoor geomagnetic localization algorithm, IEEE Sens. J. 22 (2) (2022) 1227–1237.

[131] L. Antsfeld, B. Chidlovskii, D. Borisov, Magnetic sensor based indoor positioning by multi-channel deep regression: poster abstract, in: Proceedings of the 18th Conference on Embedded Networked Sensor Systems, SenSys '20, Association for Computing Machinery, New York, NY, USA, 2020, pp. 707–708, http://dx.doi.org/10.1145/3384419.3430419.

[132] Z. Jin, R. Kang, H. Su, Multi-scale fusion localization based on magnetic trajectory sequence, Sensors 23 (1) (2023).

[133] D. Lee, S. Ryu, S. Yeon, Y. Lee, D. Kim, C. Han, Y. Cabon, P. Weinzaepfel, N. Guerin, G. Csurka, M. Humenberger, Large-scale localization datasets in crowded indoor spaces, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, IEEE Computer Society, Los Alamitos, CA, USA, 2021, pp. 3226–3235, http://dx.doi.org/10.1109/CVPR46437.2021.00324.

[134] N. LABS, Indoor localization dataset V1.0 (dept. and metro st.), 2021, URL: https://www.naverlabs.com/en/storyDetail/194.

[135] J. Choi, D. Jung, Y. Lee, D. Kim, D. Manocha, D. Lee, SelfDeco: Self-supervised monocular depth completion in challenging indoor environments, in: 2021 IEEE International Conference on Robotics and Automation, ICRA, 2021, pp. 467–474, http://dx.doi.org/10.1109/ICRA48506.2021.9560831.

[136] O. Kwon, D. Jung, Y. Kim, S. Ryu, S. Yeon, S. Oh, D. Lee, WayIL: Image-based indoor localization with wayfinding maps, in: 2024 IEEE International Conference on Robotics and Automation, ICRA, 2024, pp. 6274–6281, http://dx.doi.org/10.1109/ICRA57147.2024.10610480.

[137] Z. Zhong, Z. Tang, X. Li, T. Yuan, Y. Yang, M. Wei, Y. Zhang, R. Sheng, N. Grant, C. Ling, X. Huan, K.S. Kim, S. Lee, XJTLUIndoorLoc: A new fingerprinting database for indoor localization and trajectory estimation based on Wi-Fi RSS and geomagnetic field, in: 2018 Sixth International Symposium on Computing and Networking Workshops, CANDARW, 2018, pp. 228–234, http://dx.doi.org/10.1109/CANDARW.2018.00050.

[138] P. Alfeld, A trivariate clough-tocher scheme for tetrahedral data, Comput. Aided Geom. Design 1 (2) (1984) 169–181.

[139] C. Lee, Indoor localization dataset, 2018, [Dataset], URL: https://github.com/Carloslee96/Indoor-Localization.

[140] N. Dvorecki, O. Bar-Shalom, L. Banin, Y. Amizur, Intel open Wi-Fi RTT dataset, 2020, http://dx.doi.org/10.21227/h5c2-5439, Dataset.

[141] E.S. Lohan, J. Torres-Sospedra, H. Leppäkoski, P. Richter, Z. Peng, J. Huerta, Wi-Fi crowdsourced fingerprinting dataset for indoor positioning, Data 2 (2017) 32.

[142] E.S. Lohan, J. Torres-Sospedra, P. Richter, H. Leppäkoski, J. Huerta, A. Cramariuc, Crowdsourced WiFi database and benchmark software for indoor positioning, 2017, http://dx.doi.org/10.5281/zenodo.1001662.

[143] Z. Tóth, J. Tamás, Miskolc IIS hybrid IPS: Dataset for hybrid indoor positioning, in: 2016 26th International Conference Radioelektronika, RADIOELEKTRONIKA, 2016, pp. 408–412, http://dx.doi.org/10.1109/RADIOELEK.2016.7477348.

[144] Z. Tóth, ILONA: indoor localization and navigation system, J. Locat. Based Serv. 10 (4) (2016) 285–302.

[145] Z. Tth, J. Tams, Miskolc IIS Hybrid IPS, UCI Machine Learning Repository, 2016, http://dx.doi.org/10.24432/C5CS4R.

[146] A. Gassner, C. Musat, A. Rusu, A. Burg, OpenCSI: An open-source dataset for indoor localization using CSI-based fingerprinting, 2021, CoRR abs/2104.07963.

[147] H. Taira, M. Okutomi, T. Sattler, M. Cimpoi, M. Pollefeys, J. Sivic, T. Pajdla, A. Torii, Inloc: Indoor visual localization with dense matching and view synthesis, in: CVPR 2018 - IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, United States, 2018, pp. 7199–7209, http://dx.doi.org/10.48550/arXiv.1803.10368.

[148] InLoc: Indoor Visual Localization with Dense Matching and View Synthesis, http://www.ok.sc.e.titech.ac.jp/INLOC/, n.d.

[149] C. Sanchez-Belenguer, E. Wolfart, A. Casado-Coscolla, V. Sequeira, RISEdb: a novel indoor localization dataset, in: 2020 25th International Conference on Pattern Recognition, ICPR, 2021, pp. 9514–9521, http://dx.doi.org/10.1109/ICPR48806.2021.9413258.

[150] J.R.C.J. European Commission, RISEDB, 2020, [Dataset]. URL: https://data.europa.eu/89h/b4afbf64-17ab-4872-aaff-d5fc74ac953a.

[151] I. Silva, C. Pendão, A. Moreira, Collection of a continuous long-term dataset for the evaluation of Wi-Fi-fingerprinting-based indoor positioning systems, Sensors 22 (22) (2022).

[152] I. Silva, C. Pendão, A. Moreira, Continuous long-term Wi-Fi fingerprinting dataset for indoor positioning (full version), 2022, http://dx.doi.org/10.5281/zenodo.6928554, [Dataset].