



A survey on ubiquitous WiFi-based indoor localization system for smartphone users from implementation perspectives

Priya Roy¹ · Chandreyee Chowdhury¹

Received: 18 October 2021 / Accepted: 4 January 2022 / Published online: 24 January 2022
© China Computer Federation (CCF) 2022

Abstract

Indoor localization has the capability to change the way of providing location-based services in a closed environment and has more potential than that of GPS if the present shortcomings can be overcome. Thus, developing a ubiquitous Indoor Localization System (ILS) is the need of the day. WiFi-based indoor localization using smartphones is a promising approach to achieve ubiquity since smartphones are widely available today and most of the buildings are WiFi enabled. However, the significant variation of WiFi signal strengths with ambient conditions as well as device configuration badly affects the localization accuracy. Hence, a ubiquitous ILS, which would be widely available and would not require any specialized hardware support is still out of reach. Thus, this paper surveys state-of-the-art WiFi-based indoor localization techniques with a critical analysis of their applicability as a ubiquitous system. Consequently, the main objective of this paper is to highlight the key research challenges of implementing a WiFi-based ILS when system ubiquity is the prime concern with a discussion on future direction. Some of the experiments have been implemented on a real-life dataset in order to indicate the implementation challenges of such a system.

Keywords Indoor localization · WiFi · Fingerprint · Crowdsourcing · Machine learning · Supervised learning

1 Introduction

The primary objective of an Indoor Localization System (ILS) is to estimate the location of a person or object within a multi-storied building. Indoor localization has a widespread application area including rescue operation, warehouse monitoring, asset tracking, indoor positioning, autonomous robot navigation, games, and many more. Hence, this domain has drawn the attention of many researchers over the last two decades due to its potential applications.

The widely accepted and popular outdoor localization system, *GPS* does not work in indoor environment properly due to the attenuation of signals by obstacles like walls, the presence of furniture and human beings and so forth (Yassin et al. 2016). Thus, a number of commercial systems have started to emerge with the escalating demand for indoor

localization. In late 2011, using Google Map 6.0, released by Google, indoor localization and navigation are made available at some shopping malls and airports in countries including US and Japan. However, many building owners do not want to share information about their indoor floor plans in public due to privacy reasons. Some indoor navigation applications like the Tokyo station underground area navigation app,¹ have emerged for helping passengers to navigate through the large indoor spaces of major railway stations. These applications are not fine-grained and have primarily relied upon a manually created building map showing all important places of interest (for example, fare collection counters, ticket vending machine, etc.) which hinders their applicability to other stations.

Over the years, several indoor positioning systems based on different technologies have been explored. Based on the need for dedicated infrastructure these systems are categorized into two groups as depicted in Fig. 1. Each group is further classified into two categories based on the usage of a single technology or multiple technologies. Early systems provided considerable accuracy with pre-configured

✉ Priya Roy
priyaroy.rs@jadavpuruniversity.in
Chandreyee Chowdhury
chandreyee.chowdhury@jadavpuruniversity.in

¹ Computer Science and Engineering Department, Jadavpur University, Kolkata, India

¹ <http://en.rocketnews24.com/2016/02/17/tokyos-busiest-trainstations-have-a-new-free-english-compatible-navigation-app/>.

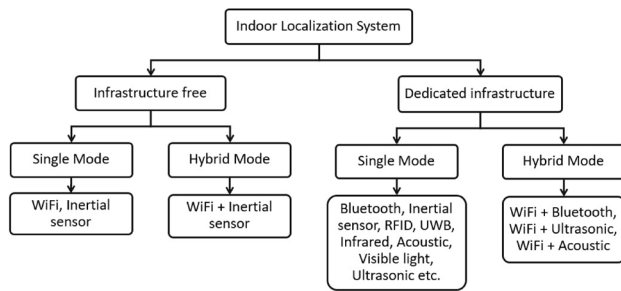


Fig. 1 Overall categorization of existing techniques of Indoor Localization System

dedicated infrastructure including infrared transmitter/receivers, ultrasound receivers, specialized hardware to emit RF/ultrasound beacons, etc. Besides, range-based localization techniques such as Angle of Arrival (AoA), Time of Arrival (ToA), Time Difference of Arrival (TDoA), Time of Flight (ToF), Return Time of Flight (RTToF), Phase of Arrival (PoA) have been used by many early systems. These techniques evaluate the distances from at least three transmitters and incorporate geometrical models for estimating locations. However, sophisticated antennas are required by such systems to provide considerable accuracy. In order to achieve wide-scale success, ILS should be made ubiquitous that would utilize the common devices (such as smartphones) carried by the people and common infrastructure support provided by the public offices, airports, shopping malls, and so on.

Nowadays, most of the buildings (like universities, hospitals, and other public infrastructures) are covered by Wireless LAN-based network infrastructure. Thus, indoor localization based on WiFi fingerprinting can be made ubiquitous as no additional hardware is required. Considerable research efforts are found to incorporate the same (Roy and Chowdhury 2018b, 2021a; Zhang et al. 2017b). This approach uses the Received Signal Strength (RSS) of WiFi Access Points (APs) to predict an unknown location. All smartphones and tablets available in the market have in-built WiFi chips to capture RSS values. An application collects RSS fingerprints of all candidate locations and sends these data to a remote server for storage and analysis. Besides, a test application captures the RSS of a location (unknown to the user) and sends it to the server for predicting the location by analyzing stored fingerprint data. The accuracy of such methods mainly depends on the extent of fingerprint effort as RSS samples are influenced by indoor ambient conditions like opening/closing of door/window, presence/absence of crowd/other interfering devices, and so on (Roy and Chowdhury 2021b).

To improve the accuracy of such systems, Bay et al. (2015), and Ruan et al. (2014) have proposed solutions based on Ultra Wide Band (UWB) radios for computing RSS

variation on narrow channels. Other systems using Radio-Frequency Identification (RFID) (Yang et al. 2016), ZigBee (Gao et al. 2013), Bluetooth Low Energy (BLE) (Cooper et al. 2016), Visible Light Communication (VLC) (Liu et al. 2008), wearable sensing (Ranjan and Whitehouse 2015) also require pre-installed additional hardware for localization which makes this kind of system hard to adapt to the large-scale indoor environment. Video or imaging cameras (Lu et al. 2016) can also be used for localization purposes. However, these techniques require adequate lighting, a direct line of sight, energy resources and so on which increases the installation complexity and cost. More importantly, user privacy may be compromised. The ubiquity of such a system is strangled if different areas of the buildings are covered by different technologies such as RFID, BLE, UWB instead of using the same technology in every location. Indoor navigation can be used conveniently by all citizens including the visually impaired if these types of applications depend on a ubiquitous technology.

In this regard, the researchers have to focus on the reasons behind the vulnerability of RSS coming from an AP and propose remedies without introducing any dedicated infrastructure such as a tracker device or a BLE device to send beacons. In fact, several emerging research challenges are raised during implementation. First of all, to work with RSS fingerprints, signal behavior should be considered for three different domains such as temporal, ambient, and device heterogeneity. Few works can be found on device heterogeneity but the localization problem subject to all three domains, particularly for crowded public infrastructure remains largely unexplored. Two important challenges arising out of it are - detecting stable APs for effective localization and ensuring sustainable localization and navigation accuracy even when training and testing conditions differ.

To the best of our knowledge, ubiquitous ILS with considerable accuracy is still beyond our reach. Even, state-of-the-art survey articles concentrate mostly on the need for integrating different technologies and the cost-effectiveness of a solution. However, the research challenges arising for a ubiquitous ILS remain mostly unexplored.

Therefore, the main contribution of this paper is to thoroughly discuss the research challenges and probable solutions associated with implementing a ubiquitous ILS based on WiFi signals for smartphone users. A survey of existing works on the different phases of a WiFi-based indoor localization framework is also presented. We have shown the effect of the discussed challenges using an RSS fingerprint dataset.

This paper is further organized as follows. Existing surveys that highlight some challenges are reviewed in Sect. 2. Section 3 defines the problem to be addressed followed by a brief description and state-of-the-art works on the different phases of ILS in Sect. 4. Associated research challenges for

developing a ubiquitous ILS and future directions are thoroughly discussed in Sect. 5. Section 6 highlights some open research issues in ubiquitous ILS. Finally, Sect. 7 concludes the paper.

2 Existing surveys on indoor localization systems

A timeline of existing survey articles of this domain is presented in Table 1. These survey articles (Yassin et al. 2016; Stojanović and Stojanović 2014; He and Chan 2016; Xiao et al. 2016; Roy and Chowdhury 2021b) of this domain mainly focused on the indoor localization problem, different technologies, approaches, existing systems, applications, some common challenges, etc. Comparisons among the various methods and technologies according to metrics, such as accuracy, privacy and scalability have also been discussed. It has been highlighted that the integration among several localization systems as well as technologies has improved the quality of indoor location-based services. Moreover, choosing a suitable positioning approach with significant accuracy and granularity and installing costly equipment for combining various non-radio technologies (IMU, visual sensors, and so on.) have also enhanced the localization accuracy (Stojanović and Stojanović 2014; He and Chan 2016; Xiao et al. 2016). Therefore, exploring a cost-effective indoor positioning system that is developed using easily available technology is yet to be explored as mentioned by Yassin et al. (2016).

Liu et al. (2007) have presented an in-depth overview of the existing indoor positioning systems, state-of-the-art localization schemes like triangulation, scene analysis, and proximity. Fischer and Gellersen (2010) have discussed the indoor localization techniques that are useful for assisting emergency responders in challenging ambience such as darkness, smokey, fire-outbreak, power outages, and so on. Besides, Yang et al. (2015) have presented their views on the importance of the mobility information that can be benefited for smartphone-based ILS along with wireless signals. Built-in sensors of the smartphone are used to identify mobility information such as step length, angular velocity, absolute direction, etc. Additionally, a survey on calibration-free indoor positioning systems has been introduced by Hossain and Soh (2015). They have also discussed the associated challenges of traditional fingerprinting like time and manpower, unforeseen environmental changes, device heterogeneity and emphasized calibration-free performance metrics such as map requirement, need for additional sensors, addressing device heterogeneity. Davidson and Piché (2017) have highlighted the lack of any standard procedure for evaluating localization accuracy of various existing algorithms. They have also suggested to design a public

benchmark dataset for evaluating state-of-the-art indoor positioning algorithms as a possible solution.

In this literature, certain research works have been found on mathematical models to track a user but fingerprinting-based techniques can better cope with changing ambient conditions. Though there are fingerprinting-based localization approaches that incorporate different technologies to improve localization accuracy, system ubiquity is not taken into account by most of these works. More specifically, in a few survey articles, the need for integrating different technologies and cost-effectiveness are mentioned but still, the research challenges arising for a ubiquitous ILS remains mostly unexplored. This motivated us to identify and elaborate on the emerging issues of such a ubiquitous ILS from the perspective of practical implementation. Accordingly, the problem statement of indoor localization using WiFi signals is formulated first in the next section.

3 Problem statement

The two vital key parameters of a WiFi fingerprinting-based ubiquitous ILS are Received Signal Strength Indicator (RSSI) and Channel State Information (CSI).

RSSI is basically a measurement of signal power received at the receiver end. This signal power at the receiver end is reduced with the distance due to the propagation of the electromagnetic wave through space, which is known as path loss. The path loss is calculated by a function of the distance between the transmitter and receiver and this relationship between RSSI and distance can be represented using Log Distance Path Loss Model as mentioned by Li et al. (2019), Zafari et al. (2019), and Wu et al. (2018). Generally, this model is expressed by the following equation.

$$PL(d) = PL(d_0) + 10\lambda \log_{10} \frac{d}{d_0} + \xi_\sigma \quad (1)$$

where PL represents signal strength in decibel, λ is known as path loss exponent, d is the distance between the transmitter and receiver, ξ_σ denotes a Gaussian random variable having standard deviation σ , PL_0 is the received power assumed at reference distance d_0 .

The RSSI is defined as a ratio of the received power to a reference power as follows.

$$RSSI = -(10\lambda \log_{10} d + A) \quad (2)$$

where λ is known as propagation exponent, d is the distance between the transmitter and receiver, and A denotes the signal strength received at 1 m of distance.

In addition, CSI is another aspect of wireless signal propagation. CSI denotes the characteristics of a communication channel that shows how the transmitted signal propagates

Table 1 A brief comparison among the existing surveys of indoor localization

Exist. work	Approaches	Technology reviewed	Remarks
Liu et al. (2007)	Triangulation, scene analysis, proximity	GPS, RFID, Cellular, UWB, WLAN, Bluetooth	Qualitative overview and comparison of existing wireless indoor positioning systems and solutions. Highlights on measuring principles, positioning algorithms and performance metrics.
Fischer and Gellersen (2010)	Triangulation, fingerprinting, proximity sensing, dead reckoning, Ad-Hoc relative positioning	Ultrasound, Inertial sensor, Laser range scanner, RFID	Survey on various localization technologies and techniques that are used to assist emergency responders in challenging scenarios they face.
Stojanović and Stojanović (2014)	Triangulation, proximity, scene analysis, dead reckoning	RFID, Bluetooth, UWB, WiFi, Zigbee, Ultrasound, Infrared, Sensor network	Overview of methods, technologies and existing systems required for developing indoor LBS and context-aware services. Present representative examples of indoor LBS.
Yang et al. (2015)	Human mobility information i.e., displacement, direction, integrated information	Accelerometer, Gyroscope, Magnetometer	Discussion about mobility enhancing smartphone-based ILS, different mobility information to assist localization and associated challenges.
Hossain and Soh (2015)	Calibration free localization	WiFi, Inertial sensor, Radio Frequency	Comprehensive understanding on existing calibration-free indoor localization techniques along with their suitability under certain scenarios or requirements.
Yassin et al. (2016)	Triangulation, scene analysis, proximity detection	UWB, WLAN, Zigbee, Ultrasound, Infrared, Sensor-based, Signals of Opportunities	Comprehensive discussion on various methodologies and technologies along with challenges of each technology. Reviewed different location-based applications. Highlighted challenges in terms of accuracy, complexity, cost, security, scalability, etc.
Xiao et al. (2016)	Wireless localization from device perspective	WiFi, Inertial sensor, Acoustic, Bluetooth, GSM, Light, Infrared, UWB, RFID, Ultrasonic	Survey on device-based and device free wireless indoor localization. Highlighted the interests of using smartphone in this domain.
He and Chan (2016)	Temporal pattern, spatial pattern, distance, proximity	WiFi	Discussion about recent advances for reducing offline labor-intensive site-survey, adapting to fingerprint changes, calibrating various devices to collect signals, achieving energy efficiency. Presented how to properly use temporal or spatial signal patterns, user participation, motion sensors.
Zhou et al. (2018)	Ranging, Fingerprinting, 3D map-based, PDR	WiFi, Bluetooth, GSM, FM broadcast, Visible Light, Acoustic, Geomagnetic, Image, Point Cloud, Motion	Introduced a three-layer framework for crowdsourcing-based ILS by combining multiple signals. Presented major insights to combine the available signals together to achieve indoor LBS in different scenarios.
Zafari et al. (2019)	Probabilistic, ANN, KNN, SVM, AoA, ToF, TDoA, RTof, PoA	WiFi, Bluetooth, Light, Zigbee, RFID, UWB, Acoustic, Ultrasound, LoRA, SigFox, IEEE 802.11ah	Discussion about traditional short-range and emerging long-range IoT technologies that can be used for indoor localization.

Table 1 (continued)

Exist. work	Approaches	Technology reviewed	Remarks
Roy and Chowdhury (2021)	Supervised, Semi-supervised, Unsupervised, ELM, Transfer Learning, Deep Learning	WiFi, Bluetooth, RFID, Camera, Inertial sensor	Comprehensive discussion on various machine learning techniques that have been used in this literature. Reviewed different performance metrics, some common challenges and location-based applications.

through the communication channel between transmitters and receivers. It represents the combined effect of wireless communication, such as fading, scattering, shadowing, multipath, and power decay with distance. CSI is calculated at the receiver end as follows.

$$\vec{Y} = \vec{H}\vec{X} + \vec{N} \quad (3)$$

where \vec{X} and \vec{Y} represent the transmitted signal vector and received signal vector, respectively, \vec{N} denotes additive white Gaussian noise and \vec{H} is the channel frequency response, which is referred as CSI. So, CSI is obtained from \vec{X} and \vec{Y} at the receiver end.

Generally, in WiFi fingerprint-based localization systems, the train data are collected from every accessible Location Points (*LPs*). Specifically in the train dataset (*TR*), RSS fingerprints are received from total n number APs. The RSS values of i th fingerprint collected from a location say l_z is notated as $tr_i = \{rss_{i1}, rss_{i2}, \dots, rss_{in} : l_z\}$. The symbol rss_{ij} denotes the RSS value of j th AP presents in the i th fingerprint, where $1 \leq i \leq m$, considering m number of training fingerprints and $1 \leq j \leq n$. All location points and the corresponding train data set are represented as $L = \{l_1, l_2, \dots, l_g\}^T$ and $TR = \{tr_1, tr_2, \dots, tr_m\}^T$ respectively, where $l_z = \{x_z, y_z\}$ is the two-dimensional coordinate of an location in the experimental region. Similarly, the test dataset is represented as $TE = \{te_1, te_2, \dots, te_{m'}\}^T$, where m' denotes the number of test fingerprints and $|TR| > |TE|$. Each test fingerprints, te_i , is represented as $te_i = \{rss'_{i1}, rss'_{i2}, \dots, rss'_{in}\}$.

In this regard, given a labeled train set, TR , and a test set, TE , the indoor localization problem is to predict the unknown location corresponding to each test fingerprints, $te_i \in TE$.

The following section describes the different steps of ILS.

4 Different phases of indoor localization system

This section discusses the general phases of an ILS as depicted in Fig. 2 and highlights the state-of-the-art works in each of these phases.

4.1 Data collection

The two different modes of RSS fingerprint collection are described below.

4.1.1 Dedicated user-based data collection

Using smart devices, the users who are willing to participate in the data collection task acquire RSS of available APs from each location point or reference point. The location

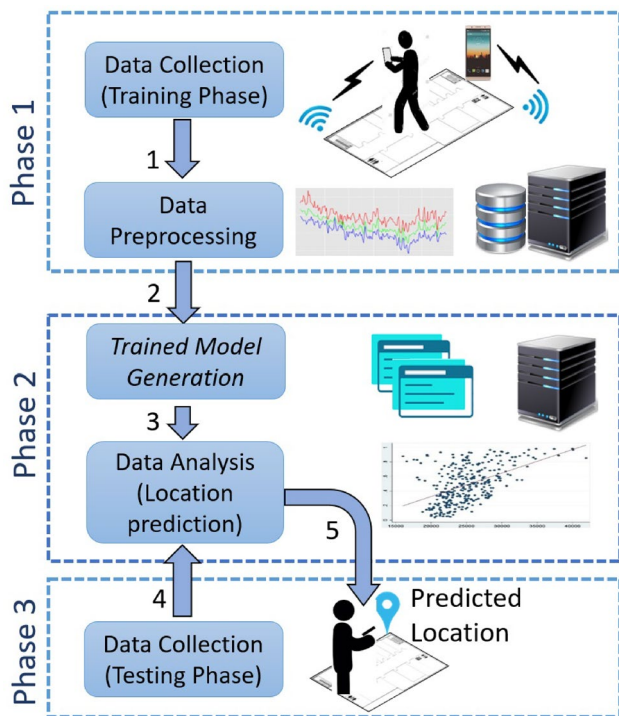


Fig. 2 Different phases of indoor localization system

points are chosen according to ground truth decided by the work. For instance, data can only be collected from meaningful location points of the experimental region. Torres-Sospedra et al. (2014) have collected RSS data from two locations per room i.e., one position inside the room and one position just outside the room for room-level localization. In order to provide fine-grained localization, Ghosh et al. (2016) have divided their experimental region into 2×2 sq. meter grids and RSS fingerprints have been collected by the user from each grid. Generally, in this data collection mode, proper labels i.e., locations are tagged with the collected fingerprints.

4.1.2 Crowdsourcing-based data collection

In this data collection process, the users, carrying smart devices, walk around the building as usual for their daily activities. Their smart devices record RSS fingerprints and other relevant information from various positions of their movement path along with the traveled walking distances (Wu et al. 2015; Lohan et al. 2017). More importantly, it is difficult to properly label the crowdsourced data because in many cases, the users are even unaware of their involvement in the data collection task. Hence, these fingerprints are grossly labeled based on the feedback of the crowd. Wu et al. (2015) have collected a large volume of

crowdsourced data and predicted the user's current location based on the number of footsteps from a previously known location. These footsteps are obtained by the accelerometer sensor of smartphones. However, these number of footsteps, as well as sensor readings, may vary from user to user. Lohan et al. (2017) have proposed another crowdsourcing approach where the current position of a user has been taken as manual input from the user. Thus, the labeling of the crowdsourced dataset may be incorrect as it depends on the crowd behavior.

4.2 Data preprocessing

The raw data needs preprocessing before analysis. The WiFi data are preprocessed in the following manner.

- *Interpolating missing entries of unheard APs:* In every location point, the signals of all APs are not heard due to the limited coverage range of WiFi signals and other indoor environmental factors. These missing entries (rss_i, rss_i') of both *TR* and *TE* need to be interpolated before analysis. There are different ways of handling the missing values such as deleting the observation, discarding the feature, imputation with mean/median/mode, etc. Generally, RSS values lie between 0 to -120 dBm. Wu et al. (2015) have assigned the signal strengths of unheard APs with 0 whereas Cooper et al. (2016) have assigned the same with the minimum observed value of RSS.
- *Removal of inconsistent AP:* Generally, APs of nearby buildings or even hotspots are heard during data collection. Those APs may not be available all the time and also have weak signal strength. A value close to 0 dBm indicates the strong signal and less than -80 dBm indicates a weak signal (poor distance sensitivity), which may not be useful. Hence, those rss_i need to be discarded at the time of location prediction. Moreover, WiFi hotspots are movable and alive for a short duration. Keeping these kinds of signal strengths for analysis incur noise in location prediction.

4.3 Data analysis

The collected datasets are analyzed to find a meaningful distribution of RSS over different locations. Statistical approaches, as well as various machine learning algorithms, are used to discover the pattern of RSS to estimate an unknown location. In this context, a brief review of the existing literature based on statistical and machine learning approaches is presented below.

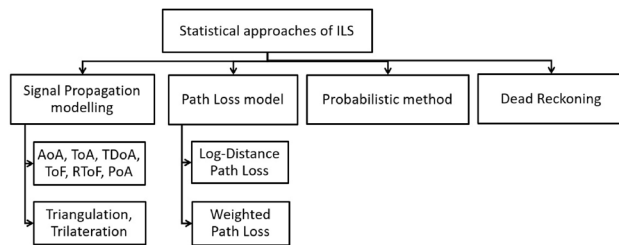


Fig. 3 Different statistical approaches for indoor localization system

4.3.1 Statistical analysis techniques

The representative statistical approaches of indoor localization are depicted in Fig. 3. In this domain, a large number of research efforts have been adopted the fingerprint approach as a basic scheme of location estimation. So, over the years fingerprinting has been applied in different technologies including WiFi, RFID, acoustic, visible light, magnetic field, and so on. However, WiFi-based fingerprinting is the most preferable one due to its ubiquity in indoor regions. In early 2000, *RADAR* (Bahl and Padmanabhan 2000) has been proposed as an indoor position tracking system and used the existing WLAN technology. It is one of the first significant works in this field. Along with the triangulation-based localization technique, this system incorporates the Rayleigh fading model and Rician distribution model. Another well known system is *Horus* (Youssef and Agrawala 2005), which requires less computational resources. This system has been implemented by many researchers in the last decades. To achieve better accuracy, different modules including Clustering, Discrete Space Estimator, Correlation Modelling and Handling, Continuous Space Estimator, and Small-Scale Compensator have been proposed in *Horus* to address different causes of wireless channel variations.

Apart from these, some recent research efforts have been found to utilize inertial sensors for improving localization accuracy (Kang and Han 2015; Koroglu and Yilmaz 2017). Generally, inertial sensor-based localization systems use Pedestrian Dead Reckoning (PDR) approach. In PDR, the distance traveled from a known or initial starting position has been computed. Embedded inertial sensors of smartphones have been used to track pedestrians. The displacement of a user has been determined by the complex human mobility information like step counting, stride length estimation, heading direction estimation, trajectory, walking, running, stair, elevator, and so on. *SmartPDR* (Kang and Han 2015) is one of the well-known PDR approaches for smartphone users. *SmartPDR* uses various modules including step event detection, heading direction estimation, step length estimation for location estimation. Generally, biases, bias stability, and thermo-mechanical noise are the most common errors that affect those inertial measurements.

In addition, CSI has been considered as a stable signature for achieving higher localization accuracy (Xiao et al. 2012; Sen et al. 2012; Zhang et al. 2020). Using CSI reliable and fine-grained information about the wireless channel has been acquired. Zhang et al. (2020) have introduced Cramer–Rao Lower Bound (CRLB) concept to analyze the localization errors of their proposed CSI-based indoor localization model. Their proposed technique has considered the relationship between the localization accuracy and the path loss, shadowing effect, multipath effect, and asynchronous effect in order to obtain the localization error due to the pedestrian motion. However, these techniques require huge calibration effort for building a fingerprint database via wardriving.

Interestingly, the major drawbacks of the fingerprint approach like the cost of time and manpower have been reduced by the crowdsourcing technique in which a fingerprint database has been constructed at the time of normal traversing with a smartphone. Rai et al. (2012) has developed a system called *Zee* to enable a zero-effort crowdsourcing approach while collecting WiFi signal strength and inertial sensor readings. In addition, *Zee* has incorporated the augmented particle filtering method to represent the uncertainty in location prediction. Similarly, in *LiFS* (Yang et al. 2012), authors have leveraged user trajectories to construct a fingerprint database that maps between fingerprints and the floor plan. They have introduced the concept of a stress-free floor plan so that the geometrical distances between any two points in the high dimension space are reflected by the real walking distances of the users. So, they transform a floor plan into a stress-free floor plan, generate a fingerprint space, and map the fingerprints to the real locations. In the localization phase, *LiFS* uses the nearest neighbor algorithm to find the target location.

Besides, of those above-mentioned approaches, some geometrical models have been built up to find out an unknown location. Unlike searching from the previously stored fingerprint data, an unknown location has been calculated using a model such as Log-Distance Path Loss (LDPL) model (Li et al. 2018; Prasad and Bhargava 2021), Weighted Path Loss model (Poulose and Han 2019) etc. Ji et al. (2006) and Lim et al. (2010) have deployed WiFi sniffers at known locations for collecting the RSS of various APs. Then an RSS map has been constructed using the LDPL model. Moreover, a more sophisticated model called the ray-tracing model has been used by Ji et al. (2006). However, these types of models have required knowledge about the locations of APs. To address this issue, Chintalapudi et al. (2010) have used a genetic algorithm along with the LDPL model in their proposed system, *EZ*, for solving the RSS-distance equations. However, *EZ* has been dependent on the available GPS information at some specific locations like the entrance of a room or near a window. In addition, a complex computation process has been involved in *EZ* and the physical localization method

has generated a lot of miss-detections in rooms. Recently, Prasad and Bhargava (2021) have designed a localization model from the RSS considering unknown transmit power and LDPL exponent.

Apart from the RSS-based model, other geometric models based on the relationship between the transmitted and received signals have been also utilized in this domain. These type of systems include CUPID (Sen et al. 2013) based on Angle of Arrival (AoA), Guoguo (Liu et al. 2013) based on Time of Arrival (ToA) and Cricket (Priyantha et al. 2000) based on Time Difference of Arrival (TDoA). Wang et al. (2015) have proposed a novel WiFi-based scheme using Curve Fitting (CF) and location search techniques. The CF technique has been used to construct a fitted RSS-distance function for each AP in each subarea. The two-step online positioning phase has been designed to determine the subarea of a device and identify the appropriate location within the selected subarea using two location search algorithms. Yang et al. (2020) have proposed a novel RSS-based Trilateration algorithm for indoor localization. First, they have preprocessed the raw data using a Gaussian filter to reduce the influence of measurement noise. Using a novel Least-Squares CF (LSCF) method, they have estimated the transmit power and the path loss exponent. However, these model-based approaches usually require the deployment of additional infrastructure, remodeling of some existing products, and knowledge about the hardware configurations.

In the recent past, the Fresnel zones model has been used in this literature to determine the elliptic region of the target (Fei et al. 2020; Wu et al. 2021). There are multiple propagation paths in the indoor environment from transmitter to receiver due to NLoS propagation. Each Fresnel phase is the phase difference between the NLoS path and LoS path. Conclusively, the resided Fresnel Zone of a target is determined according to various mathematical calculations that use the difference of the Fresnel phases. Fei et al. (2020) have applied the Fresnel zones model to obtain the elliptic region of the target according to the phase of CSI. They have implemented their proposed model in two different multipath indoor areas to evaluate its feasibility. Wu et al. (2021) have considered two traditional sensing models such as the Fresnel zone model and CSI-ratio model to extract some insightful properties for localization and a variety of device-free sensing applications.

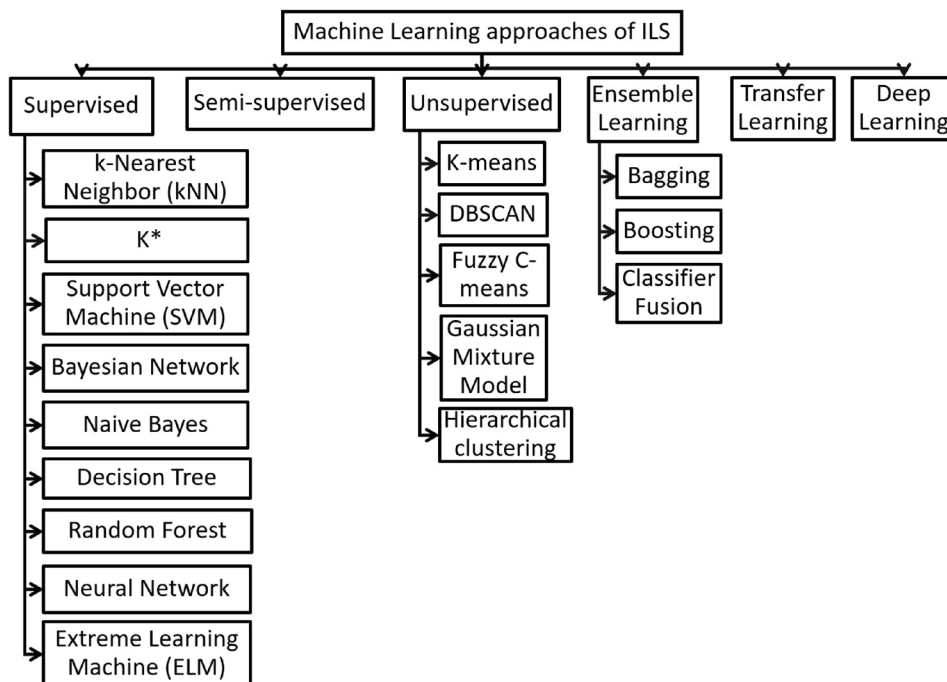
4.3.2 Machine learning techniques

In machine learning, the collected dataset is divided into train set, validation set, and test set, while the train set is used to build the model and the validation set is used to validate it. The trained models are then used to detect an unknown location using the recorded data in the test phase as input. Labeling of data is an important issue

here. Labeled train sets are suited for supervised learning algorithms. Specifically, proper data labels should be maintained in the labeled train set or else we will not be able to get significant accuracy for the test data. The Semi-supervised learning algorithms are chosen when a dataset contains few labeled data and a large amount of unlabeled data. Besides, the Unsupervised learning algorithms are applicable for the unlabeled datasets.

Consequently, the representative machine learning approaches of WLAN-based indoor localization techniques, that have been proposed and implemented in the last two decades are summarized in Fig. 4. Researchers have utilized the well-known supervised learning approaches like k-Nearest Neighbor (kNN) (Kriz et al. 2016; Roy and Chowdhury 2018a), K* (Mascharka and Manley 2015), Support Vector Machine (SVM) (Yu et al. 2014; Rossi et al. 2013), Bayesian Network (BN) (Xu et al. 2017), Naive Bayes (Zhang et al. 2014), Decision Tree (Zia et al. 2018), Random Forest (Ramadan et al. 2018), Neural Network (Meng et al. 2019; Roy and Chowdhury 2021a) in this domain. Kriz et al. (2016) have applied weighted kNN in their proposed technique where BLE beacons along with WiFi RSS have been used for estimating unknown locations. Along with several classifiers, Mascharka and Manley (2015) have used K* to analyze their collected dataset. Yu et al. (2014) have proposed an SVM-based algorithm to effectively minimize the fingerprint calibration effort while improving localization accuracy and stability. In order to develop a robust and accurate floor localization method, Xu et al. (2017) have designed a technique based on the Bayesian Network to identify the floor level of a pedestrian in a multi-storied building. Zhang et al. (2014) have investigated the performance of Bayes learning algorithms and have identified the common problem of zero probability caused by data incompleteness that affects the localization accuracy. So, they have proposed an improved Naive Base algorithm to overcome this problem. Zia et al. (2018) have investigated the performance of several machine learning techniques including Decision Tree in order to localize an object in indoor spaces. In this literature, the Decision Tree algorithm has been used to form another technique like Gradient Boosted Decision Tree, Random Forest, etc. To resolve the problem of Non-line-of-sight (NLoS) identification, Ramadan et al. (2018) have employed a Random Forest-based technique. Furthermore, willing to design a stable fingerprint database by reducing the fluctuation of WiFi RSS, Meng et al. (2019) have considered Radial Basis Function (RBF) Neural Network. Besides, Zhang et al. (2018) have addressed certain drawbacks of the existing PDR approaches such as the accumulation of errors due to noisy sensors by introducing a novel PDR-based ILS using Online Sequential Extreme Learning Machine (OS-ELM).

Fig. 4 Different machine learning approaches for indoor localization system



Apart from these, the semi-supervised (Zhou et al. 2017; Wang et al. 2018) and unsupervised learning (Wu et al. 2013; Wang et al. 2019) algorithms have been also found in this literature that aims to reduce the effort of fingerprint collection. Zhou et al. (2017) have employed a semi-supervised manifold alignment approach where unlabeled samples along with timestamps have been used to construct the radio map. Besides, Wu et al. (2013) have also eliminated the effort of site surveys by designing an unsupervised approach using k-means clustering and a logical floor plan mapping method. Wang et al. (2019) have used DBSCAN (Density Based Spatial Clustering of Applications with Noise) to cluster the RSS fingerprint database and divide the entire region into several regions based on the clustering results. They have claimed that besides improving the localization accuracy, the computational complexity and location prediction time have been also reduced by clustering the fingerprint database.

In addition, some existing works have used ensemble learning techniques like Bagging (Trawiński et al. 2013), Boosting (Cooper et al. 2016), classifier fusion (Belmonte-Fernández et al. 2018). Those techniques have been found to outperform individual machine learning algorithms. Among them, Cooper et al. (2016) have utilized the AdaBoost technique to reduce computation complexity and improve runtime performance. They have used Bluetooth Low Energy (BLE) beacons along with WiFi signals for precise indoor positioning and their system has been achieved 96.6% room-level accuracy. Belmonte-Fernández et al. (2018) have designed an ensemble classifier based on the sum of

probability estimates of six classifiers namely, Bayesian Network, kNN, Multi-Layer Perceptron (MLP), Random Forest, SVM, and Sequential Minimal Optimization (SMO) to estimate user's position.

Recently, Transfer learning and Deep learning-based indoor localization schemes have become very popular among researchers. Based on the knowledge transferred from a related source environment to a related target environment, Liu et al. (2017) have proposed a Transfer learning-based framework to enhance the scalability of fingerprint-based localization by reducing the effort of fingerprint collection for the target indoor regions. Zhang et al. (2017a) have presented an indoor localization scheme using Deep Neural Network and Deep Belief Network in which WiFi RSS and magnetic field data have been fused to enhance the localization accuracy. However, the performance of their technique is closely related to the numbers of APs, location points, and labeled fingerprints in training sets. Koike-Akino et al. (2020) have used ResNet-inspired Deep Neural Network (DNN) to identify the location and orientation of a client.

The adoption of machine learning techniques in indoor positioning and their effectiveness in extracting knowledge, discovering, learning, and improving localization accuracy can be observed in the literature. These approaches are very effective than the traditional mathematical models for solving the problem of indoor localization. Due to the dynamic nature of indoors (like fluctuation of RSS, changing ambience, device heterogeneity), the localization problem becomes too complicated for handwritten rules and/or equations. Machine learning techniques can provide a scalable

solution to the problem for large indoor spaces as the classifiers, can be easily tuned with the updation of datasets (Kim et al. 2018; Abbas et al. 2019; Zou et al. 2016). Moreover, unlike the traditional statistical approaches, machine learning techniques can be easily extended to provide a stable performance under various ambient conditions (Abbas et al. 2019; Jiang et al. 2014; Zou et al. 2015b). The online learning ability of classifiers allows to incrementally adapt to changing ambient scenarios which are very difficult with traditional approaches (Zou et al. 2015b; Jiang et al. 2014). Even, the knowledge gained through training for one ambient condition or one experimental region can be transferred to learn a new but related ambient condition or region, respectively through transfer learning mechanisms (Liu et al. 2017; Zou et al. 2017). More importantly, in the inertial sensing-based ILS, the adoption of unsupervised learning techniques more specifically, the clustering algorithms becomes a very effective approach to group the similar type of user-contributed trajectories together (Shang et al. 2015; Lashkari et al. 2018; Luo et al. 2018). The machine learning techniques have been also integrated for step detection and inertial navigation (Pasricha et al. 2015). Furthermore, in geomagnetic sensing-based ILS, Deep Neural Network has been incorporated to effectively classify the sequences of magnetic patterns that are very much sensitive with the indoor ambience (Lee et al. 2018).

Thus, from the above discussion and to the best of our knowledge, it is hard to find out any research work focusing on the essential features of a ubiquitous solution to the problem of indoor localization. However, ubiquity is an essential feature for the wide-scale commercial success of indoor localization-based applications. Consequently, the emerging research challenges for implementing a ubiquitous ILS and probable solutions are discussed in the following section.

5 Research challenges of a ubiquitous ILS and future directions

Common research challenges of indoor localization are: reducing the effort of fingerprint collection, fusion of proper technologies, selection of proper learning algorithm(s), elimination of unpredictable noise, improving reliability, and so on. In this regard, a comparison among the existing literature with respect to these research challenges is presented in Table 2.

However, in order to achieve a low-cost ubiquitous solution of indoor localization, apart from these general challenges we need to overcome the following emerging challenges.

- Designing datasets and learning techniques to handle different contexts.

- Identifying *stable* infrastructure in order to provide a ubiquitous solution.

5.1 Designing datasets and learning techniques to handle different contexts

This challenge calls the need for (i) designing a fine-grained dataset to analyze the RSS fingerprints collected from public infrastructure at different times in different ambient conditions using a number of devices having varying hardware configurations, and (ii) designing and testing localization techniques that work effectively even when the training and testing conditions vary.

5.1.1 Data acquisition subject to temporal, ambience and device heterogeneity for public infrastructure

A. Challenges faced:

The RSS of WiFi APs significantly varies from one device to another at any specific location point. The signal strength also varies at different times in a day, and due to weather conditions as well. Moreover, signal strengths also fluctuate for different indoor ambience such as the presence or absence of human beings and other interfering devices, furniture change, as well as the presence of obstacles. Thus, indoor positioning techniques should be validated using datasets that contain RSS data of different contexts such as temporal, ambience, and device heterogeneity. Additionally, proper location point labeling of the RSS data in the train set should be maintained for appropriate estimation of an unknown position. In this field, many researchers have proposed their techniques by explaining their experiments and providing results using their own datasets which are not disclosed. In other research fields such as image processing, natural language processing, bioinformatics, it is a common practice to validate any newly proposed technique using a huge number of available datasets in a public repository like the University of California Irvine (UCI) Machine Learning Repository.² This domain has fewer publicly available datasets as compared to other domains. Most of the available datasets have not considered the above-mentioned heterogeneous contexts. In addition, the available datasets have been taken either from university buildings or from office buildings. The existing techniques retain significant performance in such environments. In a public infrastructure (like shopping malls, railway stations, hospitals, etc.), the WiFi signal strengths become noisy due to the movement of the crowd, nearby interference of mobile devices, and many more. Thus, in order to develop a ubiquitous ILS, the existing or newly proposed techniques should be validated

² <https://archive.ics.uci.edu/ml/index.php>.

Table 2 Comparison among the existing literature in terms of the various research challenges

Exist. work, year	Challenges considered	Technique	Technology	Remarks
Surround-Sense Azizyan et al. (2009), 2009	<i>Fusion of technologies</i> : Acquire fingerprints of ambient sound, light, color, WiFi	k-means, SVM	WiFi, Acc, camera	Integrates optical, acoustic, motion data to generate identifiable fingerprints for logical localization. Achieved an average localization accuracy of 87% considering all sensing modalities.
UnLoc Wang et al. (2012), 2012	<i>Elimination of noise</i> : Extracting accurate location of landmarks from erroneous locations	Unsupervised clustering	Acc, Gyr, Mag, WiFi	Combines dead-reckoning, urban sensing, WiFi-based fingerprinting into a framework in order to deal with noisy sensory data, optimize energy level. Median localization errors are 0.89 m (offline), 1.69 m (online).
Yang et al. (2016), 2016	<i>Selection of proper learning algorithm</i> : Real-time data analyzing technique for processing huge signals	Online Sequential ELM (OSELM)	RFID	Tracks manufacturing objects in dynamic shop floor. Reduces training time by learning data chunk-wise. Enhances accuracy by 6% over ELM.
LoCo Cooper et al. (2016), 2016	<i>Fusion of technologies</i> : BLE & WiFi-based ILS to improve reliability	Adaptive Boosting	WiFi, BLE	Applied Boosting to the room-specific classifiers for room-level localization. Obtains the decision stumps of RSS by greedy iterative procedure to distinguish every room. Achieved accuracies are 0.966 and 0.991 considering WiFi and BLE, respectively.
Kriz et al. (2016), 2016	<i>Fusion of technologies</i> : Integration of BLE with WiFi-based ILS for more accurate localization	Weighted kNN	WiFi, BLE	Creates radio map and updates it continuously. Combines WiFi and BLE signals within one signal space. Median accuracy is 0.77 m.
DeepFi Wang et al. (2017), 2017	<i>Improving reliability</i> : Use CSI to mitigate inconsistent nature of WiFi signal	Deep learning	WiFi	Incorporates greedy learning algorithm to minimize computational complexity. Validates the system using various parameters (number of antennas, test packets, packet per batch). Achieved a localization error of 1.829 m.
Liu et al. (2017), 2018	<i>Reducing effort of fingerprinting</i> : Enhance scalability by reducing offline training	Transfer learning	WiFi	Minimizes site survey overhead without compromising accuracy. Validates the framework using various indoor environments as source and target domains. Achieved an accuracy of 85%.
Meng et al. (2019), 2019	<i>Improving reliability</i> : Inaccurate positioning of WiFi-based ILS	RBF-NN, fast clustering algorithm	WiFi	Optimizes training process by applying RBF NN. Improves training speed for more robust results. Mean square error is 1.459 m using 25 neurons.
Yoo and Park (2020), 2020	<i>Reducing effort of fingerprinting</i> : Collection of unlabeled data instead of costly labeled data	Principal Component Analysis, Gaussian Process	WiFi	Validates trustworthiness of unlabeled data. Distinguishes minor and major variations of RSS using variation detection algorithm.

Acc accelerometer, Gyr gyroscope, Mag magnetometer

using the datasets of public infrastructure. However, the benchmark datasets for public infrastructure are still hard to find in public repositories. Moreover, comparing the performances of different techniques become very hard due to the following factors.

- The size of a grid or cell, considered as location point, is not uniform. It varies from a small squared or rectangular area to the size of a room.
- During train dataset preparation, the number of data samples collected per location point i.e. sampling rate varies a lot.
- The units of RSS data are found to vary in state-of-the-art works. Either these are in linear scale or nonlinear such as dBm.

The main characteristics of some publicly available datasets are described in Table 3. The CRAWDAD (King et al. 2008), KIOS (Laoudias et al. 2013) and IPIN 2016 Tutorial (Montoliu et al. 2017) datasets contain RSS fingerprints of very small regions. In CRAWDAD (King et al. 2008) dataset, the distance between any two location points is 1.5 m. However, the positioning granularity (i.e. size of each location point or cell) is not mentioned and even the distance between any two location points is not uniform in KIOS (Laoudias et al. 2013) and IPIN 2016 Tutorial (Montoliu et al. 2017) datasets. The UJIIndoorLoc (Torres-Sospedra et al. 2014) dataset contains RSS fingerprints of a vast indoor area of a university campus. Another dataset, UJIIndoorLoc-Mag (Torres-Sospedra et al. 2015) highlights the variations of the magnetic field. This dataset consists of 40,159 discrete captures containing inertial sensor data. However, these large-scale datasets (Torres-Sospedra et al. 2014, 2015) are neither fine-grained nor contain data for various ambient conditions. In JUIIndoorLoc (Roy et al. 2019) dataset, the RSS fingerprints have been taken at a granularity level of 1 sq. meter from a university building. Moreover, these dataset contains RSS fingerprints of different times, ambience (open/closed room, presence/absence of human), and devices. In the recent past, few crowdsourced indoor localization datasets (Mendoza-Silva et al. 2018; Lohan et al. 2017) are published. Among them, the dataset described by Lohan et al. (2017) contains user traces of a significantly large area. However, one of the major drawbacks of the crowdsourcing approach is that many users are involved in constructing the radio map in the offline training phase. So, the labeling (i.e. tagging locations with samples) can also be varied from one user to the other. Thus, localization solely based on the crowdsourced RSS values can cause significant localization errors.

The localization accuracies of state-of-the-art classifiers for various publicly available datasets are shown in Fig. 5.

According to this figure, for UJIIndoorLoc (Torres-Sospedra et al. 2014) dataset, the accuracy of every classifier is better than the other datasets since the room level accuracy is always better than the fine-grain accuracy. Moreover, the localization accuracy varies with ambient conditions and positioning granularity. The area covered by each cell is significant in measuring the localization error in distance metrics. Furthermore, significant localization accuracy can be achieved with a granularity level of 1 sq. meter or 2 sq. meter using sophisticated machine learning techniques.

B. Probable solution:

The RSS data need to be collected for various contexts to analyze the robustness of localization algorithms with heterogeneous fingerprints. The different contexts can be as follows.

1. **Temporal:** Data can be recorded at different time slots in a day (say morning, afternoon, evening, night, etc.), to deal with the varying nature of RSS.
2. **Ambience:** Various public places have different types of ambience. So, the data can be collected in various ambient conditions such as the presence of a dense crowd in a railway station, presence of heavy electrical appliances in a factory, semi-open spaces in the museum, railway stations, hospitals, and so on.
3. **Device:** Smartphones with different configurations can be used for data collection to understand the variation of signal strength with respect to hardware change.

Few existing works (Torres-Sospedra et al. 2014, 2015; Lohan et al. 2017) consider device heterogeneity but do not consider temporal or ambience heterogeneity. In order to show the effectiveness of the emerging challenges discussed here, we have formed a small fingerprint dataset. The RSS data of the available WiFi APs have been collected from a faculty room that has approximately 36 sq. meter area. The distance between any two neighboring location points is 1 m. Moreover, for understanding the dynamic nature of RSS, the data have been recorded for 20 days at different times. However, this dataset does not contain RSS for varying contexts throughout the entire floor or multi-floor due to operational problems. The variation of the signal strengths of 5 APs with respect to time, using the same device and at the same location point, is shown in Fig. 6. The RSS scan is repeated every 15 min from 11:00 AM to 07:00 PM. At a particular time instant, for every APs an average of statistical RSS values received in the scan duration is considered. It can be observed that the signal strengths in the morning (11:00 AM) and evening (07:00 PM) are almost the same with a drop at around 02:00 PM to 03:00 PM for all the APs. In the morning and evening, the location is less crowded. Thus, the number

Table 3 Main characteristics of the publicly available datasets for indoor localization

Year	Dataset	Coverage area (sq.m.)	#Floor	#LP	#Sample	#AP	#Device
2008	CRAWDAD ^a King et al. (2008)	221	1	130	14,300	25	NA
2013	KIOS ^b Laoudias et al. (2013)	560	1	105	2100	60	5
2014	UJIIndoorLoc ^c Torres-Sospedra et al. (2014)	108,703	4	933	21,049	520	25
2015	UJIIndoorLoc-Mag ^d Torres-Sospedra et al. (2015)	260	1	NA	40,159	93	4
2016	IPIN 2016 Tutorial ^e Montoliu et al. (2017)	120	1	105	1629	168	1
2017	WiFi Crowdsourced ^e Lohan et al. (2017)	22,570	5	4648	4648	991	21
2018	Long-Term WiFi Fingerprint ^f Mendoza-Silva et al. (2018)	308.4	2	212	63,504	448	1
2019	JUIndoorLoc ^g Roy et al. (2019)	2646	3	1000	25,364	172	4

^a <https://crawdad.org/mannheim/compass/20080411/fingerprint>

^b https://www.researchgate.net/publication/256482916_KIOS_WiFi_RSS_dataset

^c <http://indoorlocplatform.uji.es/>

^d <http://archive.ics.uci.edu/ml/datasets/UJIIndoorLoc-Mag>

^e 10.5281/zenodo.889798

^f <https://zenodo.org/record/3748719#.YCP9IGzbiU>

^g https://drive.google.com/open?id=1_z1qhoRIcpineP9AHkfVGCfB2Fd_e-fD

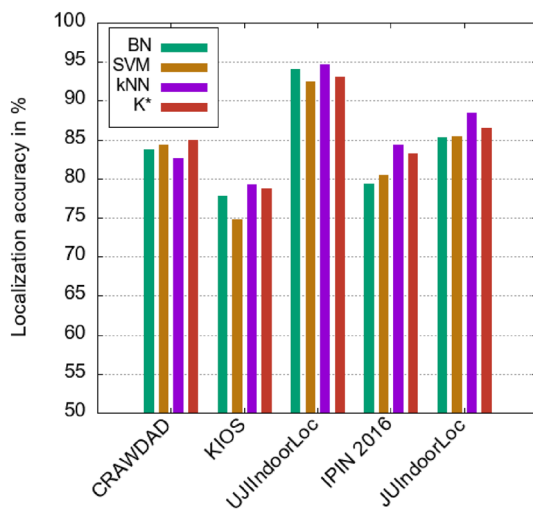


Fig. 5 Localization accuracies of state-of-the-art classifiers for various datasets

of nearby interfering devices is also less while at around 02:00 PM to 03:00 PM the place is more crowded.

C. Future direction: In JUIndoorLoc (Roy et al. 2019) dataset, the WiFi fingerprints have been collected from different floors of a university building for varying devices and minor variations of ambient conditions. However, the university campus is not that much crowded as public infrastructures, such as shopping malls or railway stations. Secondly, the floors do have a long corridor but do not contain vast open space as can be found in most airports, shopping malls, hospitals, and railway stations. A dataset

containing RSS from such indoor/semi-indoor crowded open spaces is important as the RSS behavior of such places would not match with that collected from a room or a moderate-sized closed seminar hall. The semi-indoor spaces contain unique characteristics, unlike closed indoor spaces. The available datasets mentioned in Table 3 are useful for testing commercial applications such as navigation in office/university buildings. Thus, new benchmark RSS datasets of such overcrowded places subject to the temporal, ambience, and device heterogeneity are needed for conducting experimentation on critical services, such as emergency evacuation, especially for crowded public places. So, preparing datasets from such indoor/semi-indoor environments that are very common for public places is very crucial for effective localization and navigation services.

5.1.2 Designing techniques for precise indoor localization and navigation, when training and testing contexts are different

A. Challenges faced: Considering the dynamic nature of the indoor environment, it may not be possible to take the train set and test set in the same context (collection time, indoor ambience, and scanning device). Interestingly, the training context is known but not the testing context. The test set can be collected for an ambience or device for which no training instances are available. Moreover, the context may vary while a user moves around the experimental region for collecting the test instances. Providing navigation service becomes even more challenging for fine-grained localization because of labeling ambiguity. As a user moves in an indoor

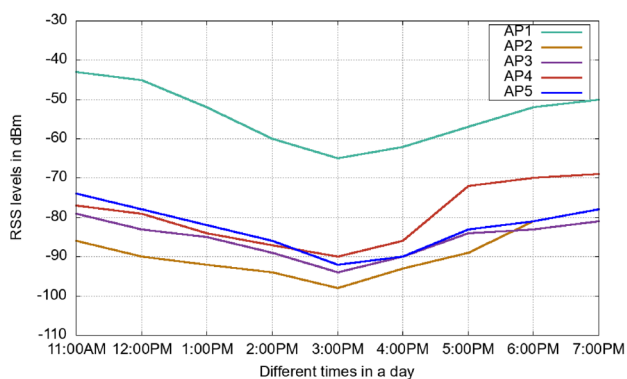


Fig. 6 Variation of RSS with different times at a specific location point

region, if the cell sizes are less (such as 1×1 sq. meter), the user moves from one cell to the other during the collection of test instances. Thus, sufficient instances from one cell may not be present in the test set and precise labeling of data also becomes very difficult. Hence, a supervised machine learning classifier may not provide satisfactory localization and navigation accuracy under different conditions. In this domain, few existing works utilized BLE beacons (Cooper et al. 2016) and RFID signals (Calderoni et al. 2015) for precise indoor positioning. The main advantage of using those signals is to get good distance sensitivity in close range and consume low power. However, the ubiquity of the ILS is not maintained as such additional devices need to be deployed as part of the infrastructure.

B. Modified problem formulation: Given a labeled train set, $TR^{(t,am,d)} = \{tr_1^{(t,am,d)}, tr_2^{(t,am,d)}, \dots, tr_m^{(t,am,d)}\}^T$ (where m is the number of training fingerprints), of known contextual heterogeneity subject to temporal (t), ambience (am), and device (d) and a test set, $TE^{uc} = \{te_1^{uc}, te_2^{uc}, \dots, te_{m'}^{uc}\}^T$ (where m' is the number of test fingerprints), of an unknown context (uc), the objective of a machine learning classifier is to predict an unknown location, $l_z \in LP$ of each $te_i^{uc} \in TE^{uc}$ with considerable localization accuracy.

C. Probable solution: Some experiments have been conducted to show the effect of different training and testing conditions with respect to the dataset mentioned in Sect. 5.1.1. Table 4 highlights the performance of the individual classifiers for predicting an unknown location when the configuration of the device in which the classifiers are trained and the device used for collecting test data is different. Each subset of train dataset, $TR(d_g, d_h)$, contains RSS of all the available APs from the corresponding locations taken by the smartphones say d_g and d_h (where $g, h = 1$ to 4 and $g \neq h$) at different times in a day. As can be observed from

Table 4 Localization accuracies in % obtained by training and testing of different condition-specific datasets using four classifiers

Train set	Test set	BN	SVM	kNN	K*
$TR(d_1, d_2)$	$TE(d_3)$	81.14	78.99	85.52	81.82
$TR(d_1, d_3)$	$TE(d_4)$	78.94	75.71	79.15	77.56
$TR(d_2, d_3)$	$TE(d_1)$	89.89	88.35	90.46	89.38
$TR(d_2, d_4)$	$TE(d_1)$	88.96	88.50	84.38	89.22
$TR(d_3, d_4)$	$TE(d_2)$	87.62	88.29	84.48	89.22

Table 4, the decision of individual classifiers is not significant enough to estimate a location. Thus, it is difficult for an individual classifier to retain generality while maintaining precision. In such a case, an ensemble of different condition-specific classifiers would be a better choice where a classifier is tuned separately for the conditions as shown in Table 5. The test dataset considered in Table 5 contains instances of all four devices as of the train dataset. However, the same instances are not present in the train and test datasets. Integrating the prediction results of these condition-specific individual base classifiers using the majority voting method the localization accuracy can be improved to 92% as shown in Table 5. Hence, the unified decision of all the individual classifiers is able to cover all the conditions. The condition-specific classifiers can be based on temporal, ambience, and device-specific data or the combination of those data.

This technique has been proposed by Ghosh et al. (2016) and implemented on a relatively small experimental zone. The localization accuracy of each individual classifier ranges between 58% to 85%. However, applying the majority voting method the accuracy increased to almost 96% in their dataset. Roy et al. (2021) have presented a weighted voting algorithm based on Dempster-Shafer belief theory to address this issue subject to device heterogeneity on the JUIndoor-Loc dataset (Roy et al. 2019).

More importantly, a condition-based ensemble classifier performs effectively when the condition in which the test dataset is collected is similar to one or more of the training conditions. However, a huge number of smart devices with different hardware configurations are available in the market. In reality, many contexts may appear as well. Hence, it is infeasible to record data for all conditions using every device. If many contexts are taken into consideration, the number of base classifiers increases exponentially for all probable combinations of contexts. Moreover, the signal strength variation in neighboring location points of 1 sq. meter granularity is very negligible, and this variation is significantly high with different contexts which may result in lower classification accuracy. Thus, a condition-based ensemble classifier may not be able to provide accurate

Table 5 Localization accuracies in % obtained by each condition-specific base classifier and the Ensemble method using kNN with the test dataset featuring instances of all devices, d_1 to d_4

Classifier	Train set	kNN	Accuracy
Base classifier 1	$TR(d_1, d_2)$	7NN	84.46
Base classifier 2	$TR(d_1, d_3)$	3NN	77.11
Base classifier 3	$TR(d_2, d_3)$	13NN	79.83
Base classifier 4	$TR(d_2, d_4)$	13NN	82.83
Base classifier 5	$TR(d_3, d_4)$	13NN	81.74
Ensemble (majority voting)	–	–	92.00

results always. Hence, it is a vital challenge for researchers to explore a suitable technique to mitigate this problem.

D. Future direction: Interestingly, *Transfer Learning* is a new paradigm for improving the learning performance when the train and test data are collected for different conditions. In such a case, transferring the knowledge gained from the source environment may effectively improve the learning performance of the new environment or target domain (Liu et al. 2017). Hence, the overhead of site-survey for the target domain gets reduced and the scalability of the system is enhanced. Generally, metric learning and metric transfer are the main phases of any Transfer Learning-based framework. In metric learning, the distance metric from the source domain is learned by maximizing the statistical dependency between the WiFi signal features and corresponding location labels. The metric transfer phase then determines the most appropriate metric for the target domain by minimizing the inconsistency between the two domains. In this context, the source environment should have sufficient labeled fingerprints to achieve satisfactory localization performance. However, getting a sufficient amount of labeled fingerprints for all representative WiFi signal features is difficult. The dependency relation between the features may also vary with time, ambience, and device.

Moreover, it is difficult to locate a user precisely while on the move due to inherent labeling ambiguity induced by the movement. Here, data can only be grossly labeled. However, these individual instances with a gross label cause ambiguity. *Multiple Instance Learning (MIL)* is a semi-supervised learning technique that can be used to solve this issue. In MIL, not an individual instance, but a bag of instances are assigned a label with the requirement that at least one instance of the bag belongs to that label. Thus, a location-specific bag contains at least one RSS instance that is actually collected from that location and some other grossly labeled instances collected on the move. In this way, MIL techniques can be explored for indoor localization with few accurately labeled and other grossly labeled RSS instances collected using different devices.

5.2 Selecting stable infrastructure in order to provide a stable solution

A. Challenges faced: The localization accuracy mainly depends on the RSS fingerprints of all available APs. Signal strengths of certain APs at a location (say l_1) collected in some ambience (say am_1) may match with the signal strengths of that AP at another nearby location (say, l_2) for a different ambience (say am_2) due to signal fluctuations. As a result, incorrect location can be predicted at the time of classification if the ambience of fingerprint data is not considered. For smaller grid sizes, that is, fine-grained location points, this effect is even more apt. Interestingly, those APs are found to be stable across ambience which exhibits strong signal strength at a location. However, a common set of APs cannot exhibit strong signal strength across the entire coverage area. Thus, building properties need to be considered along with AP signal variations for steady coverage.

Jiang et al. (2015) have selected the important APs using the signal feature-based MaxMean approach. Similarly, in order to identify important APs, Lin et al. (2014) and Xue et al. (2019) have also used signal feature-based approaches such as Group Discriminant, Access Point Discrimination Criterion, respectively. They found improvement in accuracy than other existing methods. However, the physical distribution of the APs across the entire region is not considered in those methods. Thus, the localization error may increase in some locations as the selected important APs are not evenly distributed throughout the experimental region. Another drawback of the signal feature-based approach is to determine the threshold value which is used to select only the relevant APs and exclude the irrelevant ones. Moreover, information theory-based approaches like Information gain (Zhou et al. 2013; Zou et al. 2014) and Mutual information (Zou et al. 2015a) have been used in this domain for AP selection. However, the information theory-based approaches follow a univariate way to select the important APs, so, they cannot handle the redundant APs. Besides, Kim et al. (2017) have divided a target area into several rectangular clusters and each cluster has been divided into eight subzones to uniformly distribute important APs. However, this type of region division technique may not provide sustainable results for all experimental regions.

B. Modified problem formulation: Given a reduced train set, $RTR^{(t,am,d)} = \{rtr_1^{(t,am,d)}, rtr_2^{(t,am,d)}, \dots, rtr_m^{(t,am,d)}\}^T$, of n' APs (where $n' = |AP_{min}|$ and $|AP_{min}|$ is a minimal set of stable APs) and a reduced test set, $RTE^{(uc)} = \{rte_1^{(uc)}, rte_2^{(uc)}, \dots, rte_{m'}^{(uc)}\}^T$, the objective is to select $|AP_{min}|$ in such a way that the unknown locations of the test set, $RTE^{(uc)}$, should be predicted by a machine learning classifier with considerable localization accuracy.

C. Probable solution: Machine learning-based feature selection techniques such as Correlation Attribute

Evaluation, Information Gain evaluate relevant features in a relation. Moreover, in order to take care of the stability of the selected APs, the mean and standard deviation of the RSS of each AP at every location point should be considered. Before applying feature selection, the stable APs can be short-listed based on mean and standard deviation. Following this above-mentioned mechanism, experiments are conducted using state-of-the-art classifiers through cross-validation on our collected dataset. The obtained localization accuracies are depicted in Fig. 7. Instead of considering all the APs, if a minimum number of stable APs obtained from the feature selection technique are taken into consideration, the performance of the classifiers is improved significantly as shown in Fig. 7. However, the feature selection method is not the only solution for stable AP selection. The ranking of APs may change with different times, ambience, device configuration, and more importantly with the zone where localization is considered. An AP may show stable signal strength for one region while in another region it may not show that much stability. Thus, another related question regarding the stable AP selection is how to divide the region into different zones and identify the stable APs per zone.

D. Future direction: In public indoor regions like airports, railway stations, shopping malls, museums the APs are not deployed with the aim of providing localization. Moreover, in those places, many WiFi hotspots are alive than the pre-deployed WiFi APs. The signal strength of the hotspots can degrade the performance of localization as they are movable in nature and alive for a short duration. Thus, besides the selection of stable APs, the zones must be identified in order to deploy the APs in such a way that the APs cover the entire indoor region. In addition, their signal strength must be strong enough in the target region in order to show stability for heterogeneous conditions. Identifying stable APs not only reduces the cost of maintenance of APs and the dimension of location classification problem but also ensures sustainable localization performance.

However, according to the building properties, the indoor ambient properties vary, and consequently, the localization capability of the APs differs in various regions of an experimental area. Therefore, the whole region can be divided into clusters having similar signal properties of different APs. At the time of cluster formation, it can be observed that some APs may provide distinguishing characteristics for the two adjoining clusters while some APs give no predictive information for localization. Consequently, a proper technique should be explored for the identification of the optimal number and size of clusters based on the similar RSS behavior and the selection of important and stable AP set that distinguishes among the clusters.

Moreover, AP selection is a combinatorial optimization problem where an optimal set of relevant APs are selected

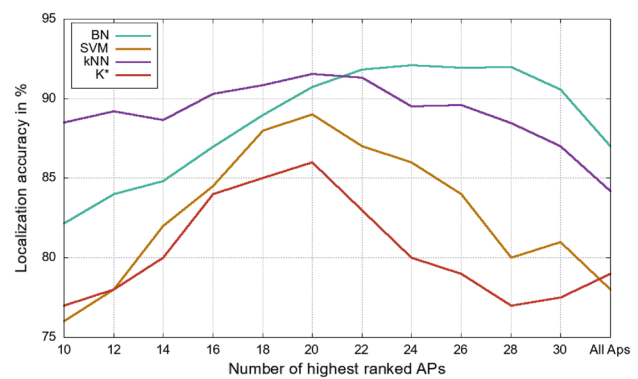


Fig. 7 Localization accuracies in % of different classifiers for the different number of highest ranked APs

from a large set of APs. Evaluating the performance of all possible subsets of APs from a large search space is generally infeasible in practice as huge computational effort is required. **Meta-heuristics techniques** find a near-optimal solution to an optimization problem like AP selection through exploring and exploiting a larger search space. Hence, Meta-heuristics techniques should be explored to select stable and important APs that generate a robust model for indoor localization.

The important APs that are less susceptible to the change of different contexts can be identified by **Deep Neural Networks** with multiple hidden layers. It can also be applied for extracting useful features from a dataset having instances from different devices subject to varying ambient conditions at different times in a day. Generally, it hierarchically learns multiple levels of representation and corresponding different levels of abstraction.

6 Discussion on open research issues in ubiquitous ILS

Despite the above-mentioned research challenges and existing research efforts, some open issues still exist. These are discussed below.

- **Adaptation with the change in feature space:** The performance of a localization system often gets affected due to some significant changes in the existing WiFi network infrastructure like addition, replacement, drop-off, and shifting of WiFi APs. At the time of emergency conditions, like fire outbreaks, some WiFi APs of one region may be inactive. In such scenarios, the training model needs to be updated to provide localization services. Thus, a feature space mapping technique needs to be explored in

order to adapt the old training model to the new feature space.

- *Adaptation with the unlabeled data:* The collection of huge training fingerprints is a very laborious and time-consuming task. Therefore, an investigation is needed to determine the minimum volume of fingerprint collection that covers the entire region. Accordingly, the system needs to be made effective and accurate by collecting unlabeled fingerprints from anonymous users. Moreover, an extensive investigation is also required to enhance the system's performance.
- *Avoidance of crowd in public places:* Crowd identification and mitigation are important concerns for public health and safety especially during the pandemic period in order to contain the spread of infectious diseases (such as Covid-19). Indoor localization techniques can be used to identify the mobility pattern of the crowd. Accordingly, crowd mitigation strategies can be investigated and possible crowd formation spots could be identified. For example, a dense crowd can be formed in front of a LED display screen at public places like some regions of a railway station, airports, shopping malls, etc. Analyzing the movement of the crowd or the formation of the crowd, if another LED display screen were placed in the nearby area, then there is a possibility to avoid the formation of the crowd.
- *Maintaining a trade-off among various performance metrics:* The localization or tracking accuracy is the most vital requirement of any localization system. A system can be better if it has a high localization accuracy. In order to increase the accuracy, often the other characteristics including, scalability, robustness, energy efficiency, cost get overlooked. So, a proper trade-off between accuracy and other characteristics needs to be maintained to develop an efficient system. If a large localization area becomes very crowded, the wireless signal channels get more congested. Hence, more calculations or analyses may be required for localization. Thus, a localization system should be scalable so that it can ensure its usual localization performance when the localization scope gets increased. Moreover, a localization algorithm should be less complex and executed in the server end due to the lack of strong processing power and long battery life of the client end mobile device. Therefore, a system needs to be energy efficient so that it can consume less power. ILS is mostly used for live-location tracking of the users and real-time navigation. Thus, an efficient system is required with low network latency. To achieve this small volume of data (pre-processed data) should be transferred among the server and client.

7 Conclusion

The aim of this research is to provide the motivation for a ubiquitous WiFi-based ILS and the emerging research challenges associated with it. A brief discussion about the different phases of an ILS, problem definition, and a review of previous works are presented. Our key contribution is to give a detailed categorization of research challenges in ILS when system ubiquity is the prior concern. This is needed for almost all applications of indoor localization including indoor navigation, asset tracking, emergency evacuation, etc.

The associated research challenges and possible future scope are studied thoroughly. Designing a fine-grained comprehensive dataset for public infrastructure, designing techniques for precise indoor positioning with the train and test dataset of various conditions, and identifying minimal but more importantly, stable infrastructure are the prime challenges as detailed in Sect. 5. These issues need to be fixed for developing a wide-scale ILS that provides significant localization accuracy in crowded indoor spaces.

Acknowledgements This research work is partially supported by the project entitled—“Developing Framework for Indoor Location Based Services with Seamless Indoor Outdoor Navigation by expanding Spatial Data Infrastructure”, funded by the Ministry of Science and Technology, Department of Science and Technology, NGP Division, Government of India, ref no. NRDMS/UG/NetworkingProject/e-13/2019(G).

Declarations

Conflict of interest The authors declare that they have no conflict of interests.

References

- Abbas, M., Elhamshary, M., Rizk, H., Toriki, M., Youssef, M.: WiDeep: WiFi-based accurate and robust indoor localization system using deep learning. In: 2019 IEEE International Conference on Pervasive Computing and Communications (PerCom), pp. 1–10 (2019). <https://doi.org/10.1109/PERCOM.2019.8767421>
- Azizyan, M., Constandache, I., Roy Choudhury, R.: SurroundSense: mobile phone localization via ambience fingerprinting. In: Proceedings of the 15th Annual International Conference on Mobile Computing and Networking, Association for Computing Machinery, New York, NY, USA, MobiCom'09, pp. 261–272 (2009). <https://doi.org/10.1145/1614320.1614350>
- Bahl, P., Padmanabhan, V.N.: RADAR: An in-building RF-based user location and tracking system. In: INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, vol. 2, pp. 775–784 (2000)
- Bay, A., Carrera, D., Fosson, S.M., Fragneto, P., Grella, M., Ravazzi, C., Magli, E.: Block-sparsity-based localization in wireless sensor networks. EURASIP J. Wirel. Commun. Netw. **1**, 182 (2015)
- Belmonte-Fernández, Ó., Montoliu, R., Torres-Sospedra, J., Sansano-Sansano, E., Chia-Aguilar, D.: A radiosity-based method to avoid

- calibration for indoor positioning systems. *Expert Syst. Appl.* **105**, 89–101 (2018). <https://doi.org/10.1016/j.eswa.2018.03.054>
- Calderoni, L., Ferrara, M., Franco, A., Maio, D.: Indoor localization in a hospital environment using random forest classifiers. *Expert Syst. Appl.* **42**(1), 125–134 (2015)
- Chintalapudi, K., Padmanabha Iyer, A., Padmanabhan, V.N.: Indoor localization without the pain. In: *Proceedings of the Sixteenth Annual International Conference on Mobile Computing and Networking*. ACM, pp. 173–184 (2010)
- Cooper, M., Biehl, J., Filby, G., Kratz, S.: LoCo: boosting for indoor location classification combining Wi-Fi and BLE. *Pers. Ubiquit. Comput.* **20**(1), 83–96 (2016). <https://doi.org/10.1007/s00779-015-0899-z>
- Davidson, P., Piché, R.: A survey of selected indoor positioning methods for smartphones. *IEEE Commun. Surv. Tutor.* **19**(2), 1347–1370 (2017)
- Fei, H., Xiao, F., Huang, H., Sun, L.: Indoor static localization based on Fresnel zones model using COTS Wi-Fi. *J. Netw. Comput. Appl.* **167**(102), 709 (2020)
- Fischer, C., Gellersen, H.: Location and navigation support for emergency responders: a survey. *IEEE Pervasive Comput.* **9**(1), 38–47 (2010). <https://doi.org/10.1109/MPRV.2009.91>
- Gao, Y., Niu, J., Zhou, R., Xing, G.: ZiFind: Exploiting cross-technology interference signatures for energy-efficient indoor localization. In: *2013 Proceedings IEEE INFOCOM*, pp. 2940–2948 (2013). <https://doi.org/10.1109/INFCOM.2013.6567105>
- Ghosh, D., Roy, P., Chowdhury, C., Bandyopadhyay, S.: An ensemble of condition based classifiers for indoor localization. In: *2016 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*, pp. 1–6 (2016). <https://doi.org/10.1109/ANTS.2016.7947872>
- He, S., Chan, S.H.G.: Wi-Fi fingerprint-based indoor positioning: recent advances and comparisons. *IEEE Commun. Surv. Tutor.* **18**(1), 466–490 (2016)
- Hossain, A.M., Soh, W.S.: A survey of calibration-free indoor positioning systems. *Comput. Commun.* **66**, 1–13 (2015). <https://doi.org/10.1016/j.comcom.2015.03.001>
- Ji, Y., Biaz, S., Pandey, S., Agrawal, P.: Ariadne: a dynamic indoor signal map construction and localization system. In: *Proceedings of the 4th International Conference on Mobile Systems, Applications and Services*. ACM, pp. 151–164 (2006)
- Jiang, X., Liu, J., Chen, Y., Liu, D., Gu, Y., Chen, Z.: Feature adaptive online sequential extreme learning machine for lifelong indoor localization. *Neural Comput. Appl.* **27**(1), 215–225 (2014). <https://doi.org/10.1007/s00521-014-1714-x>
- Jiang, P., Zhang, Y., Fu, W., Liu, H., Su, X.: Indoor mobile localization based on Wi-Fi fingerprint's important access point. *Int. J. Distrib. Sensor Netw.* **11**(4), 429 (2015)
- Kang, W., Han, Y.: SmartPDR: smartphone-based pedestrian dead reckoning for indoor localization. *IEEE Sens. J.* **15**(5), 2906–2916 (2015)
- Kim, H., Hwang, D., Kim, K.H., Jung, J.J.: Reducing positioning errors in the important access point selection method for fingerprint localization by spatial partitioning. In: *Information Networking (ICOIN), 2017 International Conference on IEEE*, pp. 445–447 (2017)
- Kim, K.S., Lee, S., Huang, K.: A scalable deep neural network architecture for multi-building and multi-floor indoor localization based on Wi-Fi fingerprinting. *Big Data Anal.* **3**(1), 4 (2018). <https://doi.org/10.1186/s41044-018-0031-2>
- King, T., Kopf, S., Haenselmann, T., Lubberger, C., Effelsberg, W.: CRAWDAD dataset mannheim/compass (v. 2008-04-11). Downloaded from <https://crawdad.org/mannheim/compass/20080411/fingerprint>, traceset: fingerprint (2008)
- Koike-Akino, T., Wang, P., Pajovic, M., Sun, H., Orlik, P.V.: Fingerprinting-based indoor localization with commercial MMWave WiFi: a deep learning approach. *IEEE Access* **8**, 84879–84892 (2020). <https://doi.org/10.1109/ACCESS.2020.2991129>
- Koroglu MT, Yilmaz A (2017) Pedestrian inertial navigation with building floor plans for indoor environments via non-recursive bayesian filtering. In: *2017 IEEE Sensors*, pp. 1–3. 10.1109/ICSENS.2017.8234216
- Kriz, P., Maly, F., Kozel, T.: Improving indoor localization using bluetooth low energy beacons. *Mobile Inf. Syst.* (2016) 11. 10.1155/2016/2083094
- Laoudias, C.C., Piché, R., Panayiotou, C.: KIOS WiFi RSS Dataset. (2013) <http://goo.gl/u7IoG>
- Lashkari, B., Rezazadeh, J., Farahbakhsh, R., Sandrasegaran, K.: Crowdsourcing and sensing for indoor localization in IoT: a review. *IEEE Sens. J.* **19**(7), 2408–2434 (2018)
- Lee, N., Ahn, S., Han, D.: AMID: accurate magnetic indoor localization using deep learning. *Sensors* **18**(5), 1598 (2018)
- Li, W., Yuan, Z., Yang, S., Zhao, W.: Error analysis on RSS range-based localization based on general log-distance path loss model. In: *2018 IEEE 15th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*, pp. 469–474 (2018). 10.1109/MASS.2018.00071
- Li, W., Chen, Z., Gao, X., Liu, W., Wang, J.: Multimodel framework for indoor localization under mobile edge computing environment. *IEEE Internet Things J.* **6**(3), 4844–4853 (2019). <https://doi.org/10.1109/JIOT.2018.2872133>
- Lim, H., Kung, L.C., Hou, J.C., Luo, H.: Zero-configuration indoor localization over IEEE 802.11 wireless infrastructure. *Wireless Netw.* **16**(2), 405–420 (2010)
- Liu, K., Liu, X., Li, X.: Guoguo: enabling fine-grained indoor localization via smartphone. In: *Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, pp. 235–248 (2013)
- Lin, T., Fang, S., Tseng, W., Lee, C., Hsieh, J.: A group-discrimination-based access point selection for WLAN fingerprinting localization. *IEEE Trans. Veh. Technol.* **63**(8), 3967–3976 (2014). <https://doi.org/10.1109/TVT.2014.2303141>
- Liu, H., Darabi, H., Banerjee, P., Liu, J.: Survey of wireless indoor positioning techniques and systems. *IEEE Trans. Syst. Man Cybern. C Appl. Mater.* **37**(6), 1067–1080 (2007). <https://doi.org/10.1109/TSMCC.2007.905750>
- Liu, X., Makino, H., Maeda, Y.: Basic study on indoor location estimation using Visible Light Communication platform. In: *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2377–2380 (2008). <https://doi.org/10.1109/IEMBS.2008.4649677>
- Liu, K., Zhang, H., Ng, J.K.Y., Xia, Y., Feng, L., Lee, V.C.S., Son, S.H.: Towards low overhead fingerprint-based indoor localization via transfer learning: design, implementation and evaluation. *IEEE Trans. Ind. Inform.* **14**(3), 898–908 (2017). <https://doi.org/10.1109/TII.2017.2750240>
- Lohan, E.S., Torres-Sospedra, J., Leppäkoski, H., Richter, P., Peng, Z., Huerta, J.: Wi-Fi crowdsourced fingerprinting dataset for indoor positioning. *Data* **2**(4) (2017). 10.3390/data2040032
- Lu, G., Yan, Y., Ren, L., Saponaro, P., Sebe, N., Kambhmettu, C.: Where am i in the dark: exploring active transfer learning on the use of indoor localization based on thermal imaging. *Neurocomputing* **173**, 83–92 (2016)
- Luo, C., Hong, H., Chan, M.C., Li, J., Zhang, X., Ming, Z.: MPiLoc: self-calibrating multi-floor indoor localization exploiting participatory sensing. *IEEE Trans. Mob. Comput.* **17**(1), 141–154 (2018). <https://doi.org/10.1109/TMC.2017.2698453>
- Mascharka, D., Manley, E.: Machine learning for indoor localization using mobile phone-based sensors. arXiv preprint [arXiv:1505.06125](https://arxiv.org/abs/1505.06125) (2015)
- Mendoza-Silva, G.M., Richter, P., Torres-Sospedra, J., Lohan, E.S., Huerta, J.: Long-term Wi-Fi fingerprinting dataset and

- supporting material. MDPI Data **3**, (2018). <https://doi.org/10.3390/data3010003>
- Meng, H., Yuan, F., Yan, T., Zeng, M.: Indoor positioning of RBF neural network based on improved fast clustering algorithm combined with LM algorithm. *IEEE Access* **7**, 5932–5945 (2019). <https://doi.org/10.1109/ACCESS.2018.2888616>
- Montoliu, R., Sansano, E., Torres-Sospedra, J., Belmonte, O.: Indoor-Loc platform: a public repository for comparing and evaluating indoor positioning systems. In: 2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1–8 (2017). <https://doi.org/10.1109/IPIN.2017.8115940>
- Pasricha, S., Ugave, V., Anderson, C.W., Han, Q.: LearnLoc: a framework for smart indoor localization with embedded mobile devices. In: 2015 International Conference on Hardware/Software Codesign and System Synthesis (CODES+ ISSS), IEEE, pp. 37–44 (2015)
- Poulose, A., Han, D.S.: Indoor localization using PDR with Wi-Fi weighted path loss algorithm. In: 2019 International Conference on Information and Communication Technology Convergence (ICTC), pp. 689–693 (2019). [10.1109/ICTC46691.2019.8939753](https://doi.org/10.1109/ICTC46691.2019.8939753)
- Prasad, K.N.R.S.V., Bhargava, V.K.: RSS localization under gaussian distributed path loss exponent model. *IEEE Wireless Commun. Lett.* **10**(1), 111–115 (2021). <https://doi.org/10.1109/LWC.2020.3021991>
- Priyantha, N.B., Chakraborty, A., Balakrishnan, H.: The cricket location-support system. In: Proceedings of the 6th Annual International Conference on Mobile Computing and Networking. ACM, pp. 32–43 (2000)
- Rai, A., Chintalapudi, K.K., Padmanabhan, V.N., Sen, R.: Zee: zero-effort crowdsourcing for indoor localization. In: Proceedings of the 18th Annual International Conference on Mobile Computing and Networking. ACM, pp. 293–304 (2012)
- Ramadan, M., Sark, V., Gutierrez, J., Grass, E.: NLOS identification for indoor localization using random forest algorithm. In: WSA 2018; 22nd International ITG Workshop on Smart Antennas, pp. 1–5 (2018)
- Ranjan J, Whitehouse K (2015) Object Hallmarks: Identifying Object Users Using Wearable Wrist Sensors. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, New York, NY, USA, UbiComp '15, pp. 51–61. <https://doi.org/10.1145/2750858.2804263>
- Rossi, M., Seiter, J., Amft, O., Buchmeier, S., Tröster, G.: RoomSense: an indoor positioning system for smartphones using active sound probing. In: Proceedings of the 4th Augmented Human International Conference. ACM, pp. 89–95 (2013)
- Roy, P., Chowdhury, C.: Indoor localization for smart-handhelds with stable set of wireless access points. In: 2018 Fifth International Conference on Emerging Applications of Information Technology (EAIT). **2018**.8470401 <https://doi.org/10.1109/EAIT.2018.8470401>
- Roy, P., Chowdhury, C.: Smartphone based indoor localization using stable access points. In: Proceedings of the Workshop Program of the 19th International Conference on Distributed Computing and Networking, ACM, Workshops ICDCN '18, pp. 17.1–17.6 (2018b). <https://doi.org/10.1145/3170521.3170538>
- Roy, P., Chowdhury, C.: Designing an Ensemble of Classifiers for Smartphone-based Indoor Localization Irrespective of Device Configuration. *Multimedia Tools and Applications*, pp. 1–25 (2021a). <https://doi.org/10.1007/s11042-020-10456-w>
- Roy, P., Chowdhury, C.: A survey of machine learning techniques for indoor localization and navigation systems. *J. Intell. Robot. Syst.* **101**(3), 1–34 (2021b). <https://doi.org/10.1007/s10846-021-01327-z>
- Roy, P., Chowdhury, C., Ghosh, D., Bandyopadhyay, S.: JUIndoorLoc: a ubiquitous framework for smartphone-based indoor localization subject to context and device heterogeneity. *Wireless Pers. Commun.* (2019). <https://doi.org/10.1007/s11277-019-06188-2>
- Roy, P., Chowdhury, C., Kundu, M., Ghosh, D., Bandyopadhyay, S.: Novel weighted ensemble classifier for smartphone based indoor localization. *Expert Syst. Appl.* **164**(113), 758 (2021)
- Ruan, W., Yao, L., Sheng, Q.Z., Falkner, N.J., Li, X.: Tagtrack: Device-free localization and tracking using passive RFID tags. In: Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), pp. 80–89 (2014)
- Sen, S., Radunovic, B., Choudhury, R.R., Minka, T.: You are facing the mona lisa: spot localization using phy layer information. In: Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services. ACM, pp. 183–196 (2012)
- Sen, S., Lee, J., Kim, K.H., Congdon, P.: Avoiding multipath to revive inbuilding wifi localization. In: Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services. ACM, pp. 249–262 (2013)
- Shang, J., Gu, F., Hu, X., Kealy, A.: Apfiloc: an infrastructure-free indoor localization method fusing smartphone inertial sensors, landmarks and map information. *Sensors* **15**(10), 27251–27272 (2015)
- Stojanović, D., Stojanović, N.: Indoor localization and tracking: methods, technologies and research challenges. *Facta Univers. Ser. Autom. Control Robot.* **13**(1), 57–72 (2014)
- Torres-Sospedra, J., Montoliu, R., Martínez-Usó, A., Avariento, J.P., Arnau, T.J., Benedito-Bordonau, M., Huerta, J.: UJIIndoorLoc: a new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems. In: Indoor Positioning and Indoor Navigation (IPIN), 2014 International Conference on IEEE, pp. 261–270 (2014)
- Torres-Sospedra, J., Rambla, D., Montoliu, R., Belmonte, O., Huerta, J.: UJIIndoorLoc-Mag: a new database for magnetic field-based localization problems. In: Indoor Positioning and Indoor Navigation (IPIN), 2015 International Conference on IEEE, pp. 1–10 (2015)
- Trawiński, K., Alonso, J.M., Hernández, N.: A multiclassifier approach for topology-based WiFi indoor localization. *Soft. Comput.* **17**(10), 1817–1831 (2013). <https://doi.org/10.1007/s00500-013-1019-5>
- Wang, H., Sen, S., Elgohary, A., Farid, M., Youssef, M., Choudhury, R.R.: No need to war-drive: unsupervised indoor localization. In: Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services. ACM, pp. 197–210 (2012)
- Wang, B., Zhou, S., Liu, W., Mo, Y.: Indoor localization based on curve fitting and location search using received signal strength. *IEEE Trans. Ind. Electron.* **62**(1), 572–582 (2015)
- Wang, X., Gao, L., Mao, S., Pandey, S.: CSI-based fingerprinting for indoor localization: a deep learning approach. *IEEE Trans. Veh. Technol.* **66**(1), 763–776 (2017)
- Wang, D., Wang, T., Zhao, F., Zhang, X.: Improved graph-based semi-supervised learning for fingerprint-based indoor localization. In: 2018 IEEE Global Communications Conference (GLOBECOM), pp. 1–6 (2018). <https://doi.org/10.1109/GLOCOM.2018.8647621>
- Wang, K., Yu, X., Xiong, Q., Zhu, Q., Lu, W., Huang, Y., Zhao, L.: Learning to improve WLAN indoor positioning accuracy based on DBSCAN-KRF algorithm from RSS fingerprint data. *IEEE Access* **7**, 72308–72315 (2019). <https://doi.org/10.1109/ACCESS.2019.2919329>
- Wu, C., Yang, Z., Liu, Y., Xi, W.: WILL: wireless indoor localization without site survey. *IEEE Trans. Parallel Distrib. Syst.* **24**(4), 839–848 (2013). <https://doi.org/10.1109/TPDS.2012.179>

- Wu, C., Yang, Z., Liu, Y.: Smartphones based crowdsourcing for indoor localization. *IEEE Trans. Mob. Comput.* **14**(2), 444–457 (2015). <https://doi.org/10.1109/TMC.2014.2320254>
- Wu, C., Yang, Z., Xiao, C.: Automatic radio map adaptation for indoor localization using smartphones. *IEEE Trans. Mob. Comput.* **17**(3), 517–528 (2018). <https://doi.org/10.1109/TMC.2017.2737004>
- Wu, D., Zeng, Y., Zhang, F., Zhang, D.: WiFi CSI-based device-free sensing: from Fresnel zone model to CSI-ratio model. *CCF Trans. Pervasive Comput. Interact.* pp 1–15 (2021)
- Xiao, J., Wu, K., Yi, Y., Ni, L.M.: Fifs: Fine-grained indoor fingerprinting system. In: *ICCCN, Citeseer*, pp. 1–7 (2012)
- Xiao, J., Zhou, Z., Yi, Y., Ni, L.M.: A survey on wireless indoor localization from the device perspective. *ACM Comput. Surv. (CSUR)* **49**(2), 25 (2016)
- Xu, Z., Wei, J., Zhu, J., Yang, W.: A robust floor localization method using inertial and barometer measurements. In: *2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1–8 (2017). <https://doi.org/10.1109/IPIN.2017.8115952>
- Xue, M., Sun, W., Yu, H., Tang, H., Lin, A., Zhang, X., Zimmermann, R.: Locate the mobile device by enhancing the WiFi-based indoor localization model. *IEEE Internet Things J.* **6**(5), 8792–8803 (2019). <https://doi.org/10.1109/JIOT.2019.2923433>
- Yang, Z., Wu, C., Liu, Y.: Locating in fingerprint space: wireless indoor localization with little human intervention. In: *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking*. ACM, pp. 269–280 (2012)
- Yang, Z., Wu, C., Zhou, Z., Zhang, X., Wang, X., Liu, Y.: Mobility increases localizability: a survey on wireless indoor localization using inertial sensors. *ACM Comput. Surv. (Csur)* **47**(3), 54 (2015)
- Yang, Z., Zhang, P., Chen, L.: RFID-enabled indoor positioning method for a real-time manufacturing execution system using OS-ELM. *Neurocomputing* **174**, 121–133 (2016). <https://doi.org/10.1016/j.neucom.2015.05.120>
- Yang, B., Guo, L., Guo, R., Zhao, M., Zhao, T.: A novel trilateration algorithm for RSSI-based indoor localization. *IEEE Sens. J.* **20**(14), 8164–8172 (2020). <https://doi.org/10.1109/JSEN.2020.2980966>
- Yassin, A., Nasser, Y., Awad, M., Al-Dubai, A., Liu, R., Yuen, C., Raulefs, R., Aboutanios, E.: Recent advances in indoor localization: a survey on theoretical approaches and applications. *IEEE Commun. Surv. Tutor.* **19**(2), 1327–1346 (2016)
- Yoo, J., Park, S.: Fingerprint variation detection by unlabeled data for indoor localization. *Pervasive Mob. Comput.* **67**(101), 219 (2020). <https://doi.org/10.1016/j.pmcj.2020.101219>
- Youssef, M., Agrawala, A.: The Horus WLAN location determination system. In: *Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services*, ACM, pp. 205–218 (2005)
- Yu, F., Jiang, M.H., Liang, J., Qin, X., Hu, M., Tao, P., Hu, X.R.: An improved indoor localization of wifi based on support vector machines. *Int. J. Future Gener. Commun. Netw.* **7**(5), 191–206 (2014)
- Zafari, F., Gkelias, A., Leung, K.: A survey of indoor localization systems and technologies. *IEEE Commun. Surv. Tutor.* **21**(3), 2568–2599 (2019). <https://doi.org/10.1109/COMST.2019.2911558>
- Zhang, W., Wang, L., Qin, Z., Zheng, X., Sun, L., Jin, N., Shu, L.: INBS: an improved Naive Bayes simple learning approach for accurate indoor localization. In: *2014 IEEE International Conference on Communications (ICC)*, pp. 148–153 (2014). <https://doi.org/10.1109/ICC.2014.6883310>
- Zhang, W., Sengupta, R., Fodero, J., Li, X.: Deep positioning: intelligent fusion of pervasive magnetic field and WiFi fingerprinting for smartphone indoor localization via deep learning. In: *16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 7–13 (2017a). <https://doi.org/10.1109/ICMLA.2017.0-185>
- Zhang, J., Sun, J., Wang, H., Xiao, W., Tan, L.: Large-scale WiFi indoor localization via extreme learning machine. In: *2017 36th Chinese Control Conference (CCC)*, pp. 4115–4120 (2017b). <https://doi.org/10.23919/ChiCC.2017.8028004>
- Zhang, M., Wen, Y., Chen, J., Yang, X., Gao, R., Zhao, H.: Pedestrian dead-reckoning indoor localization based on OS-ELM. *IEEE Access* **6**, 6116–6129 (2018). <https://doi.org/10.1109/ACCESS.2018.2791579>
- Zhang, Z., Nie, W., Wang, Y., Xie, L.: Channel state information based indoor localization error bound leveraging pedestrian random motion. *IEEE Access* **8**, 153–321 (2020). <https://doi.org/10.1109/ACCESS.2020.3017730>
- Zhou, Y., Chen, X., Zeng, S., Liu, J., Liang, D.: AP selection algorithm in WLAN indoor localization. *Inf. Technol. J.* **12**(16), 3773 (2013). <https://doi.org/10.3923/itj.2013.3773.3776>
- Zhou, M., Tang, Y., Tian, Z., Geng, X.: Semi-supervised learning for indoor hybrid fingerprint database calibration with low effort. *IEEE Access* **5**, 4388–4400 (2017). <https://doi.org/10.1109/ACCESS.2017.2678603>
- Zhou, X., Chen, T., Guo, D., Teng, X., Yuan, B.: From one to crowd: a survey on crowdsourcing-based wireless indoor localization. *Front. Comp. Sci.* **12**(3), 423–450 (2018). <https://doi.org/10.1007/s11704-017-6520-z>
- Zia, K., Iram, H., Aziz-ul-Haq, M., Zia, A.: Comparative study of classification techniques for indoor localization of mobile devices. In: *2018 28th International Telecommunication Networks and Applications Conference (ITNAC)*, pp. 1–5 (2018). <https://doi.org/10.1109/ATNAC.2018.8615220>
- Zou, G., Ma, L., Zhang, Z., Mo, Y.: An indoor positioning algorithm using joint information entropy based on WLAN fingerprint. In: *Fifth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, pp. 1–6 (2014). <https://doi.org/10.1109/ICCCNT.2014.6963033>
- Zou, H., Luo, Y., Lu, X., Jiang, H., Xie, L.: A mutual information based online access point selection strategy for WiFi indoor localization. In: *Automation Science and Engineering (CASE), 2015 IEEE International Conference on IEEE*, pp. 180–185 (2015a)
- Zou, H., Lu, X., Jiang, H., Xie, L.: A fast and precise indoor localization algorithm based on an online sequential extreme learning machine. *Sensors* **15**(1), 1804–1824 (2015b). <https://doi.org/10.3390/s150101804>
- Zou, H., Huang, B., Lu, X., Jiang, H., Xie, L.: A robust indoor positioning system based on the procrustes analysis and weighted extreme learning machine. *IEEE Trans. Wireless Commun.* **15**(2), 1252–1266 (2016). <https://doi.org/10.1109/TWC.2015.2487963>
- Zou, H., Zhou, Y., Jiang, H., Huang, B., Xie, L., Spanos, C.: Adaptive localization in dynamic indoor environments by transfer kernel learning. In: *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, (2017). <https://doi.org/10.1109/WCNC.2017.7925444>



Priya Roy is presently working as an Assistant Professor in the Department of Computer Science and Engineering, Sister Nivedita University, Kolkata, India. She is currently pursuing her Ph.D. degree in Computer Science and Engineering Department at Jadavpur University, Kolkata, India under the supervision of Dr. Chandreyee Chowdhury. She received her B.Tech. degree in Computer Science and Engineering from Dream Institute of Technology, Kolkata, India in 2012 and M.Tech.

degree in Computer Technology from Jadavpur University in 2016. Her research interest is primarily in the domain of Indoor Localization and Machine Learning.



Chandreyee Chowdhury is an Associate Professor in the Department of Computer Science and Engineering at Jadavpur University, Kolkata, India. She received Ph.D. in Engineering from Jadavpur University in 2013 and M.E. in Computer Science and Engineering from Jadavpur University in 2005. Her research interests include routing issues of Wireless Body Area Networks, mobile crowd-sensing, and applications of smartphone sensing. She is a member of IEEE

and IEEE Computer Society.