MagWi: Benchmark Dataset for Long Term Magnetic Field and Wi-Fi Data Involving Heterogeneous Smartphones, Multiple Orientations, Spatial Diversity and Multi-Floor Buildings

IMRAN ASHRAF[1], **SADIA DIN**[1], **MUHAMMAD USMAN ALI**[2], **SOOJUNG HUR**[1], **YOUSAF BIN ZIKRIA**[1], **(Senior Member, IEEE), AND YONGWAN PARK**[1]

[1]Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38541, South Korea
[2]Department of Computer Science, University of Gujrat, Gujrat 50700, Pakistan

Corresponding authors: Yongwan Park (ywpark@yu.ac.kr) and Yousaf Bin Zikria (yousafbinzikria@ynu.ac.kr)

ABSTRACT The wide use of mobile devices introduced several new services for the consumer market which are collectively called location-based services, the name being indicative of the significance of the consumer position. Consequently, a rich variety of positioning technologies have been adopted to provide and enhance user location information. The mass deployment of Wi-Fi access points (APs) and the ubiquity of the magnetic field data make them attractive candidates for indoor positioning. Additionally, the availability of embedded magnetic and Wi-Fi sensors in smartphones helps to achieve positioning without additional infrastructure. Even though Wi-Fi and magnetic field data offer complementary characteristics for enhancing positioning accuracy, several challenges for these technologies remain unresolved. However, the lack of publicly available datasets for the magnetic field and Wi-Fi makes it very difficult to extensively investigate these characteristics. Also, the proposed approaches cannot be tested on common benchmark datasets to analyze the results of the state-of-the-art approaches. To resolve these issues, this study presents a dataset that comprises the magnetic field, Wi-Fi, and the data from the inertial measurement unit (IMU) sensors of the smartphone including accelerometer, gyroscope, and barometer. First, the important characteristics of both the Wi-Fi and the magnetic field that require further investigation are highlighted, and later the data are collected. The data are collected over a longer period spanning approximately five years involving five different smartphones used by four different users, both female, and males. Different path geometries are followed in different multi-floor buildings which are physically separated, comprising both small and large areas. Besides, three different orientations of the smartphone are considered for data collection covering corridors, halls, and laboratories. The data from the stairs help to test 'stairs up' and 'stairs down' events and approaches aiming at multi-floor positioning can be tested with the provided dataset.

INDEX TERMS Indoor positioning, smartphone sensors, magnetic field data, Wi-Fi data, inertial measurement unit, benchmark dataset.

I. INTRODUCTION

Indoor positioning and localization have been an area of wide research interest over the past few years. Precise location information serves location-based services (LBS) and offers

The associate editor coordinating the review of this manuscript and approving it for publication was Vlad Diaconita.

invaluable help for rescue operations during an emergency. Although both outdoor and indoor positioning is necessary, the indoor position becomes significantly important because humans spend 80% to 90% of their time indoors [1]. As a result, 70% of cellular calls and 80% of data connections originate from indoor environments which include airports, universities, train and bus stations, and offices, etc. [2].

Unlike outdoor positioning where the global positioning system (GPS) and its variants like assisted-GPS (A-GPS), GPS-inertial navigation system (GPS-INS), and GPS with camera provide reliable position information, the indoor environment poses several additional challenges. GPS faces many physical barriers to performing indoor positioning including frequency blocking, attenuation caused by walls and roofs, less number of visible satellites, and similar other problems. So, the calculated position shows high error which may be higher than the indoor area itself, especially for small buildings.

To overcome the challenges of indoor positioning, several positioning technologies have been proposed and adopted for indoor environments. For example, radio frequency identification (RFID), infrared (IR) and ultrawideband (UWB) have been adopted for indoor positioning [3]–[5]. Similarly, Bluetooth low energy (BLE) has been a topic of interest due to its low cost and easy implementation. However, these technologies require installing additional sensors such as tags or beacons to perform positioning and are called infrastructure-based technologies collectively. On the other hand, infrastructure-less technologies do not need additional infrastructures, such as pedestrian dead reckoning (PDR), Wi-Fi, and the earth's magnetic field-based positioning. PDR provides the relative position of a user and requires a starting or previous position contrary to the Wi-Fi and magnetic field-based positioning that have no such presumption. Wi-Fi is categorized under the infrastructure-less category owing to the wide deployment of Wi-Fi access points (APs) and its ability to calculate user's position in sparsely deployed APs [6].

Wi-Fi positioning approaches have been investigated over the last two decades and offer an average accuracy of 4 to 6 m [7], [8]. However, several challenges of Wi-Fi positioning remain unresolved and require further investigation. For example, the propagation losses of wireless propagation cause a substantial change in the received signal strength (RSS) and the performance of the fingerprinting approaches is severely affected [9], [10]. Similarly, signal absorption and shading, multipath shadowing, and dynamic environments involving human mobility such as airports, and shopping malls cause signal fluctuation and introduce high positioning errors. Heterogeneity of hardware and antenna design from various vendors and signal absorption by the human body are also reported to influence the RSS value and degrade positioning performance [11], [12].

Magnetic field-based positioning has emerged as a potential solution for indoor positioning due to its ubiquity, simplicity, and offered positioning accuracy. Despite being smooth outdoor, the magnetic field data is disturbed by the ferromagnetic materials present in the construction material and indoor infrastructure that leads to magnetic disturbances also called anomalies. Such anomalies are reported to have unique distribution in the indoor and can be leveraged to identify various locations [13]. Smartphone embedded micro-electromechanical system (MEMS) magnetometer measures such anomalies which can be used as fingerprints. Although magnetic field-based positioning is simple to adapt and an effective technique for indoor positioning, it has several limitations [14]. For example, smartphone heterogeneity leads to different positioning results even with the same positioning algorithms. Smartphone companies use the embedded magnetometers from various manufacturers with different noise tolerance levels and precision which shows the difference in the measured data. Similarly, the impact of different orientations of the smartphone is not studied very well for such approaches. However, it offers several supportive features to enhance the Wi-Fi positioning performance. For example, the magnetic field data tend to show long-term stability than Wi-Fi data. Similarly, dynamic environments involving human mobility have less influence on the magnetic field data.

Wi-Fi and magnetic field data offer complementary features that can help to enhance the positioning performance of hybrid systems. However, lack of the publicly available datasets containing both Wi-Fi and magnetic field data, makes it very difficult to study such systems. Data collection is a laborious and time-consuming process that requires experienced users. Already available datasets lack several of the previously discussed characteristics. Hence, this study provides a hybrid dataset of Wi-Fi and magnetic field data and makes the following contributions in essence

- A detailed analysis of the existing datasets has been conducted to analyze their pros and cons. Essential elements of a hybrid dataset are outlined.
- A large dataset MagWi is presented which contains the data for Wi-Fi and the magnetic field for indoor positioning. The data are collected over a long period of approximately 5 years.
- Beside Wi-Fi and magnetic field, inertial measurement unit (IMU) data are provided from the accelerometer, motion sensors, and barometer involving four users both males and females. Five different buildings are used for data collection and five smartphones including Galaxy S8, LG G6, Galaxy A8, LG 7, and Galaxy S9+ are used.
- To study the influence of the dynamic environment, various scenarios are considered involving low to high human mobility at a public exhibition hall where the data are collected from different smartphones.
- The data collection also involves various orientations of the smartphone such as navigation, call listening, and swinging. Apart from the data from multi-floor building, the data collection environments include stairs in a multi-floor building, and 'stairs up' and 'stairs down' events can be tested.

The rest of the paper is organized as follows. Already available public datasets are discussed and analyzed in Section II. Important elements of both Wi-Fi and magnetic field datasets are highlighted in Section III. The description of the data collection process, path geometry used for data collection, details of the smartphones and their orientations, and data collection scenarios are provided in Section IV. Section V

contains the structure of the dataset including both Wi-Fi and magnetic field data. The guide on how to use the data for positioning is described in Section VI along with several positioning results. In the end, discussions and conclusions are given in Section VII.

II. RELATED WORK

Despite a large number of proposed approaches and research papers on indoor positioning and localization involving the use of the magnetic field and Wi-Fi, publicly available datasets are not in tandem. The available datasets are analyzed in the following sections.

A. MAGNETIC FIELD DATASETS

A large number of research papers can be found in literature during the last few years that focus on using magnetic field data for indoor positioning. However, the number of publicly available datasets is only a few. For example, a large dataset containing the magnetic field data is presented in [15]. Indoor magnetic field anomalies are recorded using two different smartphones including Motorola Moto Z Play and Lenovo Phab 2 Pro. For data collection two platforms are used: a hand-held smartphone and a wheeled robot-mounted smartphone. The data are collected simultaneously at the predefined location points in two different buildings. The data for various temporary interfering items are collected by placing an industrial fan, iron cupboard, and a rack of personal computers alongside the path used for data collection.

A dataset called, UJIIndoorLoc-Mag is presented in [16] that contains the magnetic field data collected in a laboratory environment. Consisting of 8 corridors, the area has a dimension of 260 m^2 which is separated by bookcases and desktop computers. Space is further divided into smaller areas where the data are collected between starting and ending points at each 0.1 s. Data collection involves two different smartphones including Google Nexus 4 and LG G3, both operating on Android 5.0. Dataset also contains the IMU sensor data collected while the user is walking at a consistent speed. Also, multiple users collected the data over a short period of few days.

B. WI-FI FINGERPRINTING DATASETS

Two public datasets containing the data from Wi-Fi APs are presented for open areas with no GPS coverage in [17]. First, the dataset is collected from four different users in a busy open area called Bush Court in Murdoch University, while the second dataset contains auto-generated records from Wi-Fi APs received from users' devices. Four smartphones are used for the data collection so that the impact of smartphone heterogeneity can be investigated. PRIMO GH7I with Android 8.1, Oppo F1 Plus with Android 6.0, and LG G6 and Samsung S8 both operating on Android 7.0 are used for data collection.

A Wi-Fi RSS dataset is presented in [18] to analyze the fluctuations in RSS over a longer period. A trained professional collected the data spanning over 15 months on pre-defined location points. Various datasets are provided for each month containing both training and testing data separately. Data collection is done using Samsung Galaxy S3 in an indoor building with several floors.

Similarly, the authors present a Wi-Fi dataset called UTSIndoorLoc for indoor positioning in [19]. The dataset contains Wi-Fi fingerprints from a 16 story building of the University of Technology, Sydney. Covering an area of approximately 44,000 m^2, a total of 1,840 sample points are provided in the dataset. It provides approximately 9,107 training samples and 387 test samples from 589 different Wi-Fi APs.

Data collection for Wi-Fi or magnetic field fingerprints is a labor-intensive task that requires a substantial amount of time by expert users. An alternative strategy is to involve multiple non-professional users through an approach called crowdsourcing. In crowdsourcing, the data from multiple users are collected which are later combined into one database after necessary preprocessing and data cleansing. Following a similar procedure, a crowdsourced Wi-Fi fingerprinting dataset is provided by [20]. The data are collected in a five-floor university building with a footprint of 208 × 108 m^2. Eight volunteers participated in the data collection process with 21 android devices. The data are collected along with different directions, and the details of how the data are joined from 8 users are not provided.

C. HYBRID DATASETS FOR WI-FI AND MAGNETIC FIELD

Due to the complementary features of Wi-F- and magnetic field data, several hybrid datasets have been presented over the last few years that comprise both the Wi-Fi and magnetic field data for indoor positioning.

A public dataset is introduced in [21] that contains the magnetic field and Wi-Fi data. Sony Xperia M2 and LG W110G Watch R is used for the data collection in an indoor space comprising offices, corridors, and connected corridors. Additionally, the data from IMU sensors are included as well such as accelerometer and gyroscope. The data from the smartwatch helps to analyze user's various orientations and their impact on the data. The covered area is large and the path trajectory is complex involving several turns in different directions. Besides, the Wi-Fi and IMU data can be used to enhance the positioning accuracy when used with the magnetic field data.

A hybrid dataset is provided in [22] where the data are collected using several technologies to enhance the accuracy of the indoor positioning. The dataset contains the data for the Wi-Fi, Bluetooth, and magnetic field, and Samsung Galaxy Young GT-S5360 operating on Android 4.4.4 is used for data collection. Thirty Wi-Fi APs and 9 Bluetooth devices are installed in an indoor area of 465.75 m^2 for data collection. The data are collected using a client-server architecture where the smartphone has a lightweight client application for data collection while the storage is on the server-side. The ground truth points for the data collection are marked by manually

calculating the distance. Besides, the data collection environment is a multi-level building.

A Wi-Fi and magnetic field dataset is presented in [23] to evaluate hybrid positioning approaches using Wi-Fi and the magnetic field data. IMU data are also incorporated for providing the heading direction and orientation information of the user. The data are collected following two different directions in corridors and intersections. A single smartphone Google Nexus 4 operating on Android 5.0.1 is used for data collection. Data collection time spans over 6 months to study the time-related mutation of the magnetic field and Wi-Fi data.

IPIN2016 tutorial dataset is an alternative to the UJIIndoor-Loc dataset and uses comparatively smaller indoor scenarios for positioning [24]. The data are collected in an indoor corridor, located in the School of Engineering of the University of Alcala, Spain. The dataset contains both training and testing samples separately comprising 927 and 702 records, respectively. The data are collected for a total of 168 APs on manually marked location points in the corridor. AP names, RSS, basic service set identifier (BSSID), location coordinates, and other important information are provided with the dataset.

D. LIMITATIONS OF EXISTING DATASETS

Although a large number of separate datasets are publicly available for Wi-Fi and magnetic field data, as well as, hybrid datasets for the same, they lack in several aspects. First, magnetic field datasets often involve using single or at most two devices which shows their inability to investigate device heterogeneity. For example, [15] provides the magnetic field data of two smartphones while [22], [23] provide the data using only one smartphone. The availability of a large range of smartphone companies including Samsung, iPhone, LG, Huawei, and Nokia necessitate a dataset containing the data from more number of smartphones. Second, the use of fixed orientation of the smartphone is the predominant way of data collection for the magnetic field data as large variations are observed when changing smartphone orientations. None of the publicly available magnetic field datasets use more than one orientation for the data collection. For the most part, the smartphone is held in front of the user's body to collect the data where the user is allowed to change the directions but not the smartphone orientation. Thirdly, for the Wi-Fi data, the impact of long-term fluctuations in RSS is not studied very well. Although the data of 15 months is provided in [18], the data are collected using a single smartphone, i.e., Galaxy S3. The influence of various smartphones on RSS fluctuation over time is not covered. Fourthly, Wi-Fi has been investigated for almost two decades and it provides a solution for multi-floor indoor positioning, the datasets do not cover different scenarios for multi-floor environments. For example, uneven floors pose a great challenge for multi-floor Wi-Fi positioning, however, none of the available datasets provide the Wi-Fi data for such environments to test the Wi-Fi positioning approaches. Last but most important,

dynamic environments with human mobility cause problems and reduce the positioning performance of Wi-Fi positioning. Similarly, for magnetic field data, the height of the user, walking speed of the user, and temporal indoor changes are important aspects of the positioning which are not covered in the existing datasets.

III. IMPORTANT ELEMENTS OF INDOOR POSITIONING DATASETS

Magnetic field and Wi-Fi data have different characteristics that are regarded as important for indoor positioning. The characteristics that can influence indoor positioning accuracy are highlighted here.

A. CHARACTERISTICS OF MAGNETIC FIELD DATA FOR INDOOR POSITIONING

The magnetic field data can be denoted by two representations: using magnetic x, y, and z or F, D, and I, where x, y, and z represent the North, East and vertical values of the magnetic field while F, D, and I show the total magnetic intensity, declination and inclination of the magnetic field, respectively [25]. The inclination and declination in the 2nd representations are angles that are very sensitive to the attitude of the device and conventionally not used for positioning. The elements in the first representation change with the orientation of the smartphone and are not very well studied. Figure 1 shows the magnetic field data for different orientations of the smartphone including phone in the pocket, call listening, and phone swinging. Two orientations are considered for the phone in the pocket mode where the phone is put in an upside-down position with a camera at the top and an opposite position with the camera at the bottom. The objective of these orientations is to show that even a single orientation can have multiple sub-orientations that affect the magnetic field data.

Ideally, the navigation mode is used where the user holds the smartphone in front of his body with the smartphone LCD facing upward. In this orientation, the y-axis is directing towards the user's moving direction. The user moves in various directions with the smartphone's fixed orientation. To increase accuracy and reduce preprocessing the frame of reference should be always aligned with the global coordinates, however, this is not very practical for real-world positioning scenarios. Existing datasets are collected with a fixed orientation and lack the diversity of orientation.

Despite its long-term stability, the magnetic field data is mutated over time and the world magnetic model is revised every five years to tackle the mutation [26]. Studies involving long-term magnetic field data confirm that the mutation of the magnetic field data is slower than the Wi-Fi signals [27]–[30]. Similarly, the materials used in man-made constructions affect the magnetic field data differently. This effect depends on the type and quantity of the ferromagnetic materials. Consequently, the nature of the indoor environment and its settings like the place of electric doors, elevators, vending machines, and electric cupboards, etc. interfere with the
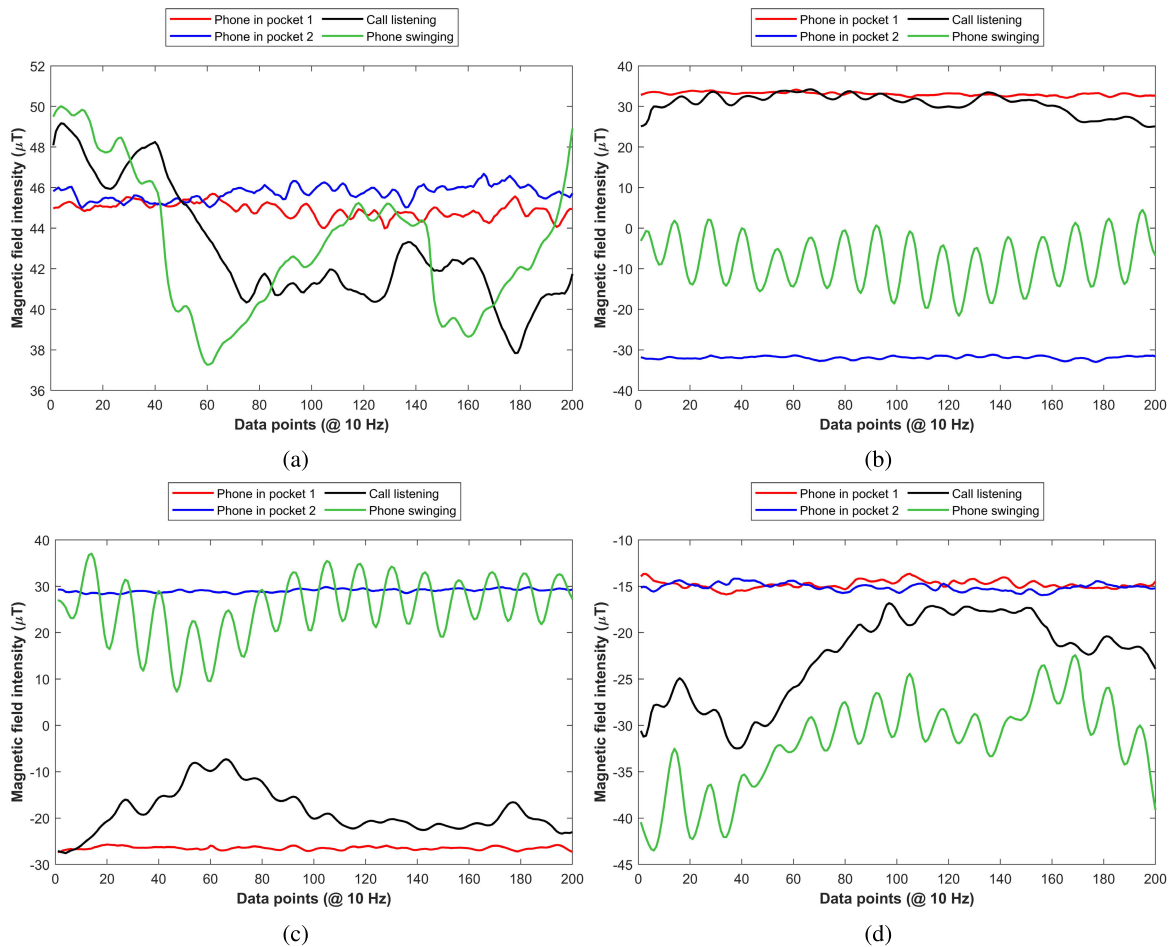
FIGURE 1. The magnetic field data for different orientations, (a) Total magnetic field intensity *F*, (b) Magnetic *x* component, (c) Magnet *y* component, and (d) Magnetic *z* component.

magnetic data and offer different distribution of the magnetic field data. To analyze the performance of the magnetic field-based indoor positioning approaches, the dataset should contain data from different types of buildings.

One of the bigger challenges of magnetic field-based positioning is to cope with smartphone heterogeneity. In the majority of the cases, positioning approaches utilize a single smartphone for performance evaluation which does not show the capability of the approach to work seamlessly with various smartphones. Smartphone heterogeneity poses two challenges: the attitude of the smartphone magnetometer and the diversity of the magnetometer. Various smartphone companies use IMU sensors from different manufacturers even for the different models of the same product line. The embedded magnetometers have different levels of sensitivity and noise tolerance and result in different magnetic field values even for the same location. Even though the indoor infrastructure, as well as, human mobility is the same, variations can easily be observed. A sample for this phenomenon of the magnetic field data is shown in Figure 2 which shows the data collected at the same location using four different smartphones.

The presence of magnetometers from different manufacturers and their different models tends to show different magnetic field data even when the data are collected for the same location from various smartphones. This leads to different positioning accuracy even while the same positioning approach is used. Device calibration helps to minimize the change in the magnetic field readings from the same device where calibration is achieved by moving the smartphone making a shape like a digit 8 along three axis [31]. Besides, offset value [32] and sequential measurements [33] have been used for the same purpose as well. However, the majority of the existing datasets do not provide the sequential measurements to test the positioning approaches. Similarly, the datasets provide the magnetic field data collected from a single user which is not appropriate to analyze the impact of the user's height on the positioning accuracy.

B. IMPORTANT CHARACTERISTICS OF WI-FI DATA FOR INDOOR POSITIONING

Wi-Fi indoor positioning has been investigated for almost two decades now, yet it has several challenging and unresolved problems. For example, the direction of the smartphone at the
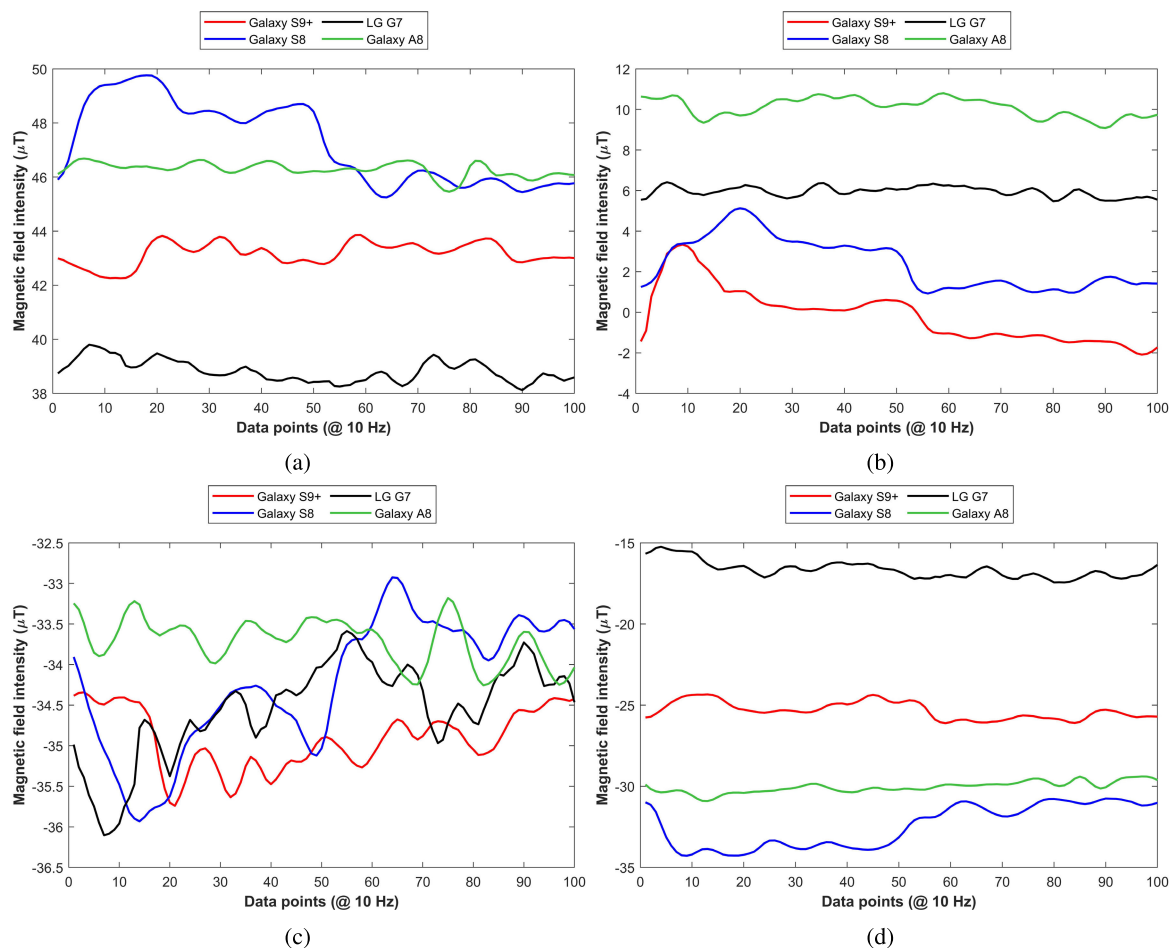
FIGURE 2. Diverse attitude of Galaxy S9+, Galaxy S8, LG G7, and Galaxy A8 smartphones at the same location, (a) Total magnetic field intensity *F*, (b) Magnetic *x* component, (c) Magnet *y* component, and (d) Magnetic *z* component.

time of data collection can potentially influence the collected RSS value. The change in the RSS value may occur because changing the direction can make a line of sight (LOS) AP to be a non-line of sight (NLOS) AP. Figure 3 shows the RSS values collected while the smartphone is placed on the table and following different directions. It indicates that even when there is no human mobility and the smartphone is placed on the table, different scans show different RSS values.

Similar to the data collected from different directions, the orientation of the smartphone has a substantial impact on the collected RSS. Since the direction of the smartphone is changed with the orientation, the RSS value is changed, even when the user and the position of the user are not changed. To illustrate this characteristic, Wi-Fi data are collected using three different orientations of the smartphone. For this purpose, the 'navigation', 'phone in the pocket', and 'phone swinging' modes are used. Figure 4 shows the Wi-Fi APs that are visible at the location of data collection and their RSS value. It appears that the collected RSS values for visible Wi-Fi APs are not the same for all orientations. Even though the same smartphone is used to collect the data for the same indoor location, the RSS value varies from one

orientation to another. The data are collected over a short period of 10 minutes, so the RSS values are not influenced by time. Moreover, AP-7 is not visible in 'phone swinging' orientation and assigned an RSS value of -1 for illustration. Wi-Fi data are collected at a sampling rate of 1 Hz. The frequency of independent scans of the Wi-Fi environment depends on hardware response so it varies from mobile to mobile. Latest mobile devices like Samsung Note 9 scans WiFi signal at less than 1Hz, whereas earlier mobile take 2 to 3 seconds and sometimes even more due to the hardware response to the mobile OS. However, it is also seen that multiple same scans are returned by the mobile within a particular scan frequency so in that case the time stamp associated with each scan helps to avoid repetitive RSS values. A threshold is defined as that if the time difference between two consecutive scans is less than 1Hz and RSS difference is zero that scan will be ignored.

Apart from the impact of smartphone orientations, a large range of smartphone models influences the collected RSS value significantly. Wi-Fi hardware and software configurations of different smartphones lead to RSS fluctuations. Similarly, the antenna design found in different smartphones also
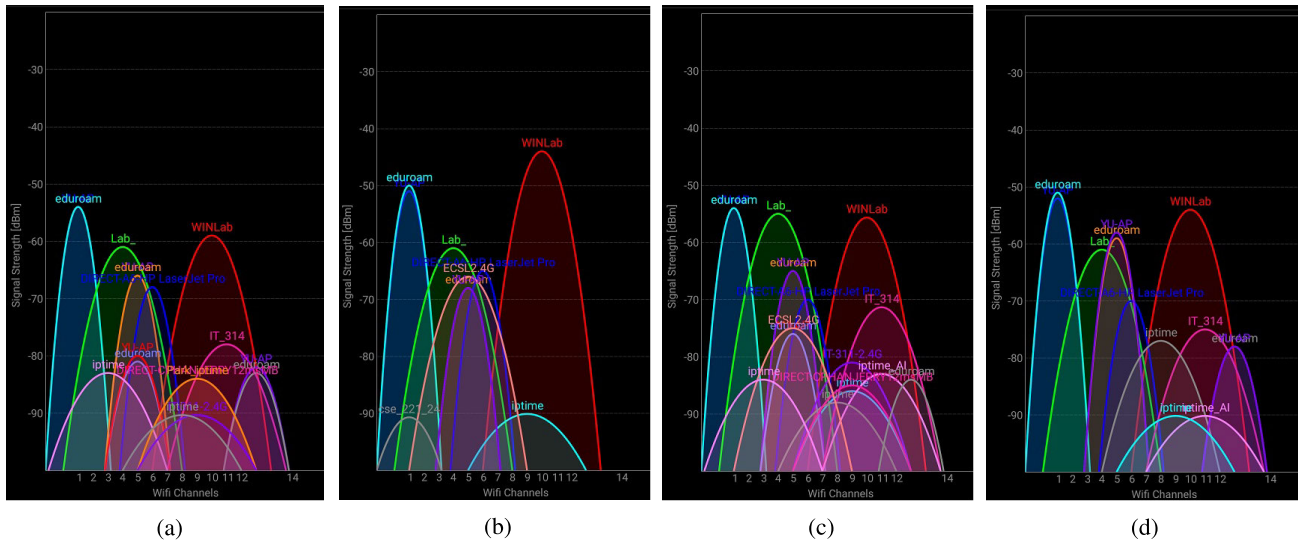
FIGURE 3. Visible APs with received signal strength from Galaxy S8 smartphone, (a) Smartphone direction to North, (b) Smartphone directing to West, (c) Directing towards East, and (d) Towards South direction. 'WINLab' AP is placed in the West direction from the place of data collection using WiFi Analyzer application [34].
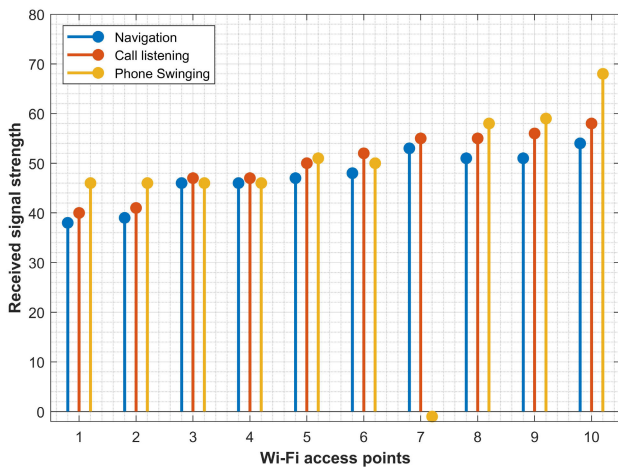

FIGURE 4. Visible Wi-Fi access points and their RSS values. RSS values are shown positive for better illustration. A value of −1 indicates that the AP is not found during the scanning.

affects the RSS value [11]. The impact of various chipsets installed in a smartphone is analyzed in [12] which states that different chipsets in heterogeneous models affect the performance of Wi-Fi positioning. In addition to the RSS fluctuation, the number of detected APs may be different when different smartphones are used for scanning. To corroborate these findings, Wi-Fi data are collected using four different smartphones including Galaxy S8, Galaxy S9+, LG G7, and Galaxy A8.

Figure 5 shows the RSS values of 10 Wi-Fi APs with good RSS values using Galaxy S8 smartphones. The APs found in the S8 scan are searched in the scans from other smartphones and their associated RSS values are recorded if the AP is found, otherwise, an RSS of −1 is assigned to the missing AP. For better illustration in the bar graph, RSS values are displayed as +ve quantities. Figure 5 indicates that the RSS

values are not the same for all the smartphones used to collect the data. Moreover, one AP from the Galaxy A8 scan is missing in the LG G7 scan while three APs are missing in the Galaxy S9+ scans. Studies indicate that the RSS values are influenced by the change in the smartphones, our experiments confirm these findings [35], [36]. Owing to the impact of device heterogeneity, the Wi-Fi dataset should contain the Wi-Fi data from several different smartphones to analyze the performance of the state-of-the-art indoor positioning approaches. However, the majority of the existing datasets contain the data from a single smartphone and do not fulfill the requirement of device heterogeneity.

One of the most important issues of Wi-Fi positioning is the influence of dynamic indoor environments such as a change in the indoor infrastructure due to placing or removing furniture or cupboards. Even the doors opening and closing lead to RSS fluctuations which cause positioning degradation [37], [38]. Human mobility has a strong impact on the collected RSS value. For example, Wi-Fi data collected with user standing is less noisy than the data collected while walking [39]. The human body is reported to influence the propagation of wireless signals and fluctuation in RSS value [40]. The fluctuation in the RSS values has a strong relationship with the distance of the human body from APs, as well as, the number of humans present in that environment [41]. Because of these findings, an important characteristic of the Wi-Fi dataset is to consider several indoor environments with different levels of human mobility so that the performance of the Wi-Fi positioning approaches can be evaluated extensively. Existing publicly available Wi-Fi datasets do not provide the data for indoor dynamic conditions and traditionally the data are collected inside university campuses that do not involve substantial human activity. So, it is highly desirable to collect the Wi-Fi data at a public indoor place like an exhibition hall where human mobility of various levels can
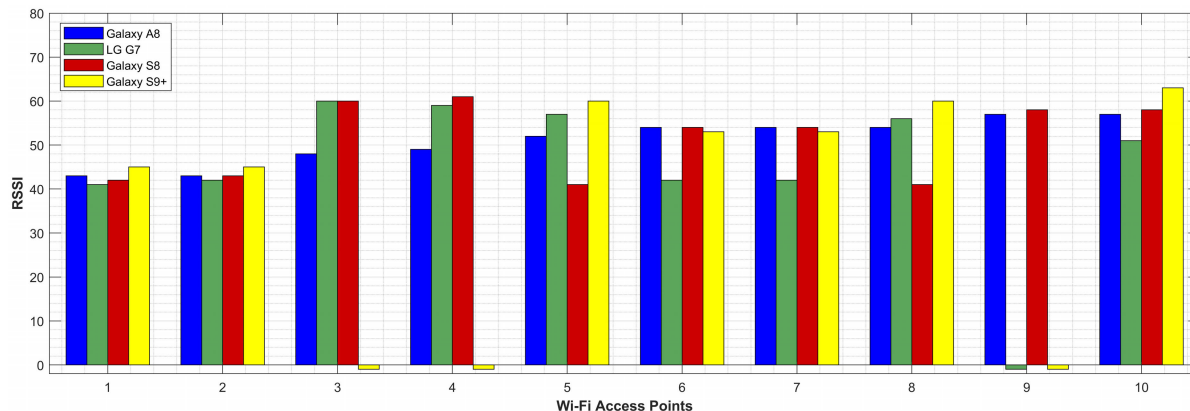
FIGURE 5. Wi-Fi APs and the RSS values using different smartphones. A value of −1 indicates that the AP is not found in the scan. RSS value is shown +ve for better illustration.

TABLE 1. A brief overview of existing publicly available datasets.

Ref.	Smartphone	User	Time	Orientation	Dynamicity
Magnetic Field datasets					
[15]	Multiple	Single	Short	Single	N/A
[16]	Multiple	Multiple	Short	Single	N/A
Wi-Fi datasets					
[17]	Multiple	Single	Short	Single	No
[18]	Single	Single	Long	Single	No
[20]	Multiple	Multiple	Short	Single	No
[19]	Single	Multiple	Short	Single	No
Hybrid datasets					
[21]	Multiple	Single	Short	Single	No
[22]	Single	Single	Short	Single	No
[23]	Single	Single	Long	Single	No
[24]	Single	Single	Short	Single	No

be recorded. The dimensions (area) of the place used for data collection are also significant, as the positioning in small indoor areas tends to show higher positioning performance.

Environmental dynamics is another point to be considered for Wi-Fi data as RSS-based Wi-Fi positioning is vulnerable to various weather conditions. Different weather conditions are reported to introduce attenuation for wireless signal propagation which affects the positioning performance [42]. Change in the temperature is also associated with the change in the measured RSS value which creates problems for fingerprint-based Wi-Fi solutions [43]. The data collected in the winter and summer will help to investigate the influence of both temperature and time. Wi-Fi signals are depleted over time and the positioning performance is high if the training and test data are collected over a short duration. The higher the gap between the training and test data is, the higher the chances are that the positioning performance will be poor [44], [45]. A brief comparison of existing datasets is provided in Table 1 considering the above-discussed elements of the magnetic field and Wi-Fi data concerning indoor positioning.

IV. DATA COLLECTION PROCESS FOR HYBRID DATASET

Multiple smartphones are used to collect the data from several different buildings involving multiple users to overcome the limitations of the existing datasets. Similarly, the data are

collected with various orientations so that the analysis of the impact of orientation can be done using the dataset. Data collection is carried out during different times of each year, however, most of the data are collected from October to December and June to August at least once every year.

A. DATA COLLECTION USING HETERGOENEOUS SMARTPHONES

Since smartphone heterogeneity is one of the major problems for both Wi-Fi and magnetic field data, several smartphones from different companies are used to collect the data. Five smartphones are used for this purpose, which has different software and hardware specifications, as shown in Table 2. Specifications of the smartphones are given because the vendors, as well as, the models of various sensors from the same vendor are different.

B. USE OF DIFFERENT ORIENTATIONS FOR DATA COLLECTION

Owing to the influence of smartphone orientation on the magnetic field and Wi-Fi data, the data are collected using multiple orientations so that its impact on the positioning accuracy can be studied. For this purpose, this study considers the three most commonly used orientations of the smartphone including 'navigation', 'call listening', and 'phone swinging'. Navigation is widely used where the phone is held in the hand in front of the user's body approximately at navel height. The user moves in different directions but does not change the orientation of the smartphone. An illustration of this orientation is given in Figure 6.

Call listening is yet another common orientation used by phone users. In this orientation, the smartphone is placed beside the year facing upside down direction with the camera at the top of the smartphone. The screen of the smartphone is facing the user's face and the phone is tilted approximately 45 degrees clockwise. Although this orientation is not frequently used with magnetic field-based positioning, it is quite helpful to formulate positioning approaches for real-world situations. The details for magnetic field x, y, and

TABLE 2. Names and descriptions of smartphone sensors used to collect the magnetic field data.

| Sensor	Vendor	Description
		Samsung Galaxy A8
SM-A530N Galaxy A8	Samsung	Octa-core, Adreno 540 GPU, Android 9.0 (Pie), 4 GB RAM
Magnetometer (AK09918C)	Asahi Kasei	3-axis, 16-bit, sensitivity 0.15 mT/LSB, temperature –30 to +85 °C, 6.0 mA [46]
Accelerometer (LSM6DSL)	Asahi Kasei	3-axis, 16-bit, sensitivity 0.061 mg/LSB, Temperature –40 to +85 °C, 0.25 mA [47]
Gyroscope (LSM6DSL)	STMicroelectronics	3-axis, 16-bit, sensitivity 125 mdps/LSB, Temperature –40 to +85 °C, 6.1 mA
		Samsung Galaxy S8
SM-G950N Galaxy S8	Samsung	Octa-core, Adreno 550 GPU, Android 8.0.0 (Oreo), 4 GB RAM
Magnetometer (AK09916C)	Asahi Kasei	3-axis, 16-bit, sensitivity 0.15 mT/LSB, temperature –30 to +85 °C, 6.0 mA [48]
Accelerometer (LSM6DSL)	STMicroelectronics	3-axis, 16-bit, sensitivity 0.061 mg/LSB, Temperature –40 to +85 °C, 0.25 mA [47]
Gyroscope (LSM6DSL)	STMicroelectronics	3-axis, 16-bit, sensitivity 125 mdps/LSB, Temperature –40 to +85 °C, 6.1 mA
		Samsung Galaxy S9+
SM-G965U Galaxy S9+	Samsung	Octa-core, Adreno 630 GPU, Android 10, 6 GB RAM
Magnetometer (AK09916C)	Asahi Kasei	3-axis, 16-bit, sensitivity 0.15 mT/LSB, temperature –30 to +85 °C, 6.0 mA [48]
Accelerometer (LSM6DSL)	STMicroelectronics	3-axis, 16-bit, sensitivity 0.061 mg/LSB, Temperature –40 to +85 °C, 0.25 mA [47]
Gyroscope (LSM6DSL)	STMicroelectronics	3-axis, 16-bit, sensitivity 125 mdps/LSB, Temperature –40 to +85 °C, 6.1 mA
		LG G6
LGM-G600L LG G6	LG	Quad-core, Adreno 530 GPU, Android 7.0 (Nougat), 4 GB RAM
Magnetometer (AK09915C)	Asahi Kasei	3-axis, 16-bit, sensitivity 0.15 mT/LSB, temperature –30 to +85 °C, 6.0 mA [49]
Accelerometer (BMI-160)	Bosch	3-axis, 16-bit, Temperature –40 to +85 °C, 0.18mA [50]
Gyroscoope (BMI-160)	Bosch	3-axais, 16-bit, Temperature –40 to +85 °C, 0.9 m [50]
		LG G7
LM-G710N LG G7 ThinQ	LG	Octa-core, Adreno 630 GPU, Android 10.0, 4 GB RAM
Magnetometer (AK09918C)	Asahi Kasei	3-axis, 16-bit, sensitivity 0.15 mT/LSB, temperature –30 to +85 °C, 1.1 mA [46]
Accelerometer (IAM-20680)	TDK-Invesense	3-axis, 16-bit, Temperature –40 to +85 °C, 0.24mA [51]
Gyroscoope (IAM-20680)	TDK-Invesense	3-axais, 16-bit, Temperature –40 to +85 °C, 1.25 mA [51]
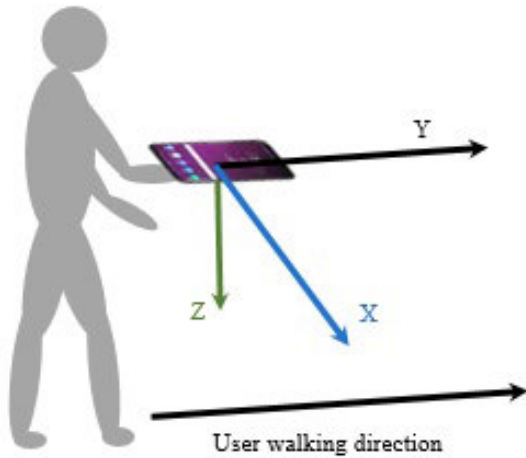
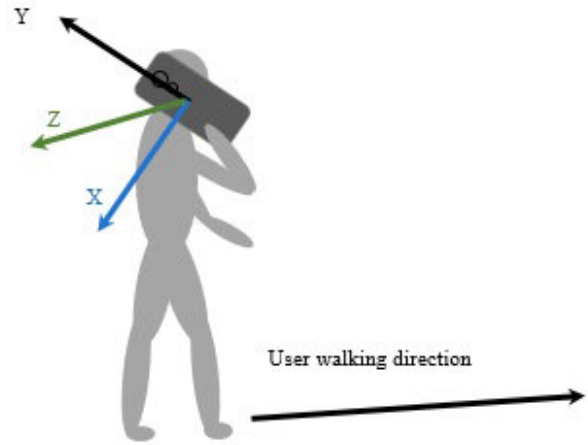**FIGURE 6.** Illustration of the navigation orientation of the smartphone.

FIGURE 7. Call listening orientation of the smartphone with magnetic field axes.

z axis are provided in Figure 7. The angle of the smartphone is an approximation and may change with different users depending on their style of attending the phone call.

The third orientation considered for data collection is the phone swinging in the hand of the user while walking. It may also involve data collection while the user is standing however the orientation of the smartphone is the same. This orientation is complex than those of navigation and call listening and not used frequently for the magnetic field data or Wi-Fi-based positioning because the generated data contains noise and positioning performance is affected. Figure 8 shows the magnetic field axes for this orientation. As shown in the figure, there is a movement window for this orientation, so when the

phone is swung the generated magnetic field data changes with it. It makes the data preprocessing difficult and affects the performance of the positioning accuracy. The purpose of this orientation is to provide the data so that more complex real-life scenarios can be tested for positioning. The selected orientations are used when collecting the data from different smartphones.

C. PATH TRAJECTORY AND SPATIAL DIVERSITY USED FOR DATA COLLECTION

Path trajectory is an important aspect to collect magnetic field and Wi-Fi data. So, different path trajectories are defined involving both simple straight paths, as well as, complex
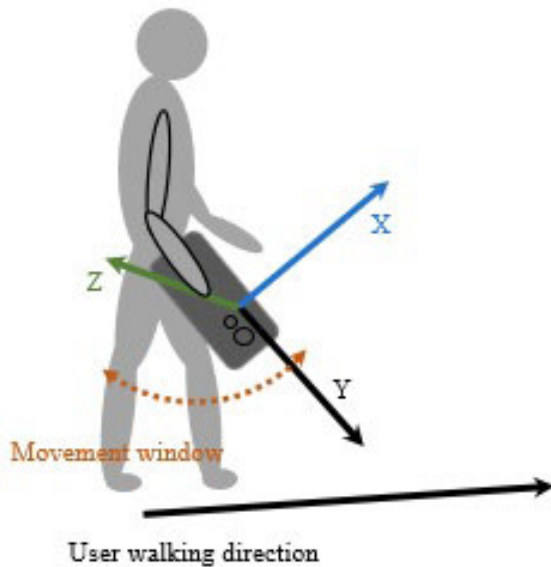
FIGURE 8. Axes of the smartphone for phone swinging orientation.

paths involving multiple turns in all directions. In addition, different buildings are used to collect the data wherein each building has different data collection scenarios. Also, the dimension of each building concerning the available area for positioning is different. Depending on the path length used, data collection time varies from one building to another. However, for static data collection where the user needs to collect the data while standing at each ground truth location point, approximately 25 to 30 min are required for the longest path while the shortest path needs 15 to 18 min. For continuous data collection where the user has to walk along a dedicated path, the longest path is covered in 2 to 2.5 min while the shortest path needs a 1.5 to 2 min walk. The data for longer periods are collected from a single user (User 2) using Galaxy S8 and LG G6 especially. For the most part, the data collection time is 25 to 45 minutes and the time of the day varies for different data collection campaigns. Following the norms of the datasets, the collected data are transformed into a standard format where the original SSIDs and BSSIDs are replaced. For SSIDs, wireless access points (e.g., WAP123) are used while original BSSIDs are changed with randomly generated unique BSSIDs of the same format and length.

This study selected five buildings with different indoor infrastructure and different deployments of Wi-Fi APs. Similarly, the number of available Wi-Fi APs is also different. Five buildings include information technology (IT), computer science (CS), electrical engineering (EE), regional innovation center (RIC), and business & economics (BE) department building. For each building, the path followed for data collection is different concerning the space available for positioning. Figures 9a, 9b, and 9c show the paths followed for IT building scenarios 1 and 2 and CS building, respectively. The green and yellow circles on the map indicate the starting and

ending points of the data collection, respectively, while the arrows indicate the walking direction of the data collectors. The data are collected at manually marked location points located in a grid form.

The Wi-Fi data can be collected at 2 to 5 m to reduce the data collection time, however, this study considers the data collection at high resolution and selects a distance of 1 m for data collection. The same procedure is followed in other buildings as shown in Figure 10. The purpose of following multiple scenarios is to provide the data for both simple and complex paths involving multiple turns so that the performance of the state-of-the-art positioning approaches can be evaluated. Similarly, the data collection buildings have different sizes, indoor settings, and Wi-Fi APs which makes it easier to analyze the influence of these factors on the positioning performance.

One major challenge for Wi-Fi-based floor identification schemes is the uneven structure of floors in some buildings which includes ramps or a few steps on a floor. In such structures, the floor is not changed, however, due to the up and down level of the floor, the performance of the floor identification approaches is affected. Unfortunately, existing Wi-Fi datasets do no provide the Wi-Fi data for such floor structures and the performance of the floor identification algorithm cannot be determined appropriately. This study collects the data for uneven floor structure from BE building, as shown in Figure 11.

Figure 11a shows the followed path for data collection in BE building. The path is subdivided into two sub-paths for better presentation and indicated as path 1 and path 2 and shown as magenta and blue circles, respectively, in Figure 11a. Path 1 contains two descending stairs containing 8 steps each where the height of each step is 13.30 cm. On the other hand, path 2 has three ascending stairs with 8, 5, and 8 steps, respectively. The surveyor starts from path 1 and follows through path 2 till the end. Path 1 and path 2 are used to explain the path geometry; for the data collection, both paths are connected. Lines of the same color, shown in Figure 11b for the two paths indicate that they have the same height. For example, the end of path 1 and the start of path 2 which are shown in red color, has the same height. The data are collected at manually labeled points which are separated by 1 m. The data are collected on the 2nd floor of the BE building with very little human mobility.

Predominantly, the Wi-Fi benchmark datasets do not contain multi-floor data and the floor identification schemes cannot be tested. To overcome this limitation, this study considers the data collection on multiple floors of a building. It explicitly involves the data from 3-floor buildings where the data are collected for both Wi-Fi and magnetic field. Also, the data are collected on the stairs which helps to test the performance of stairs-related events such as 'stairs up', 'stairs down', and floor change, etc. Figure 12 shows the structure of stairs used for the data collection. For stairs events, the data are collected in both directions, i.e., starting from the 3rd floor to the 1st floor and vice versa.
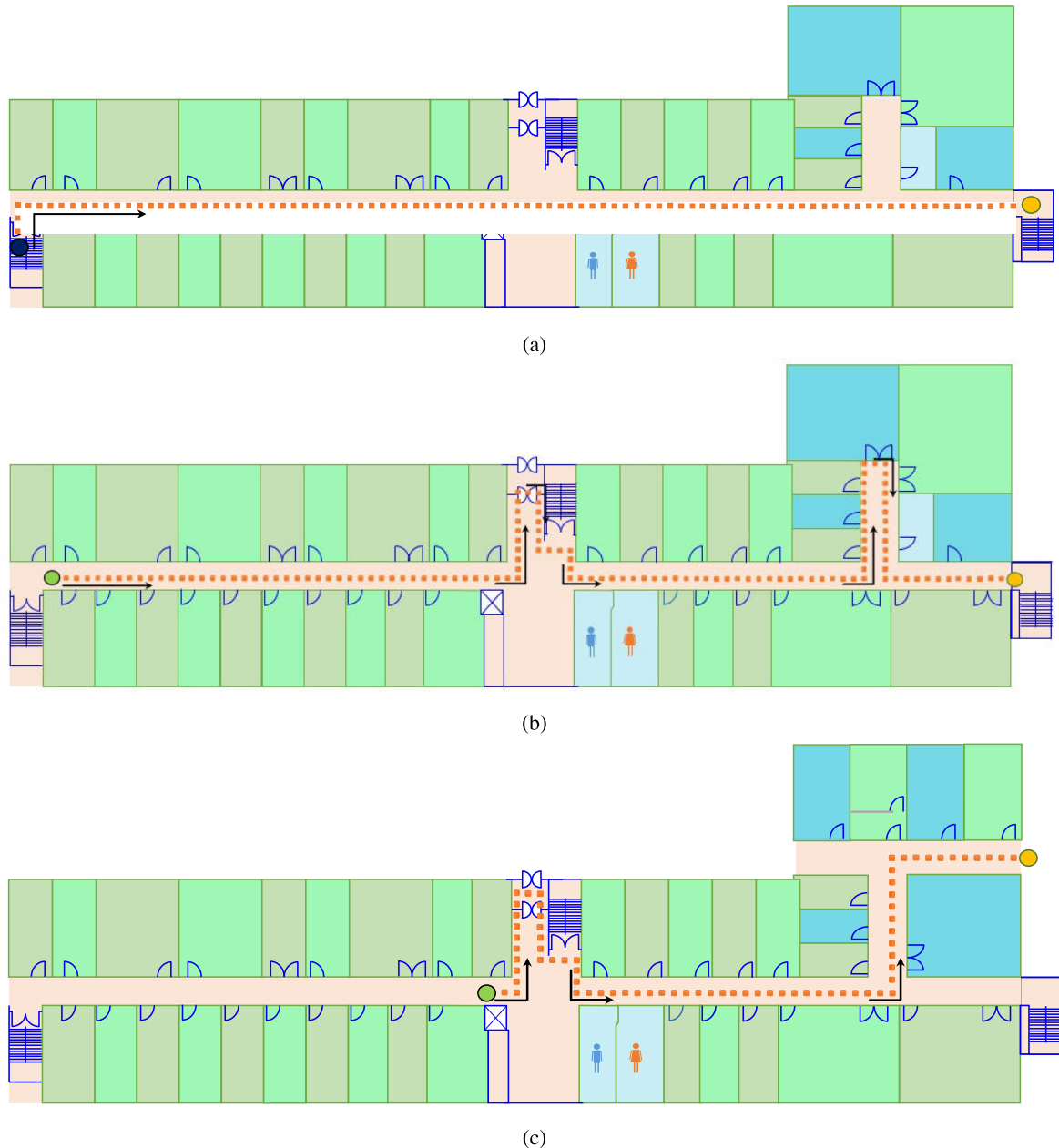
FIGURE 9. Paths used for data collection, (a) IT building scenario 1, (b) IT building scenario 2, and (c) CS engineering. Arrows on the maps shows the walking direction of the surveyor.

D. DATA COLLECTION FOR DIFFERENT LEVELS OF HUMAN MOBILITY

Human mobility has a significant influence on the change in the RSS values and affects the performance of Wi-Fi fingerprinting systems. Existing Wi-Fi benchmarks do not provide human mobility information for the provided data. This study considers a public exhibition hall called Starfield COnvention and EXhibition (COEX) center situated in the Gangnam area of Seoul, Korea. The data are collected in a large exhibition hall with the dimension 108×106 m^2. Data collection for various human mobility conditions are considered in the light of the reported results in the literature

which states that the human body changes the RSS value [52], obstructs the signal and causes RSS variation [53] and the higher number of people's presence leads to large positioning errors than the absence of people [11], [54].

Two human mobility conditions are considered for the data collection

- Medium human mobility involving approximately 50 to 350 people present at the time of the data collection.
- High human mobility when more than 350 people are present in the hall.

The data are collected during an exhibition in progress where different stalls from various companies are installed
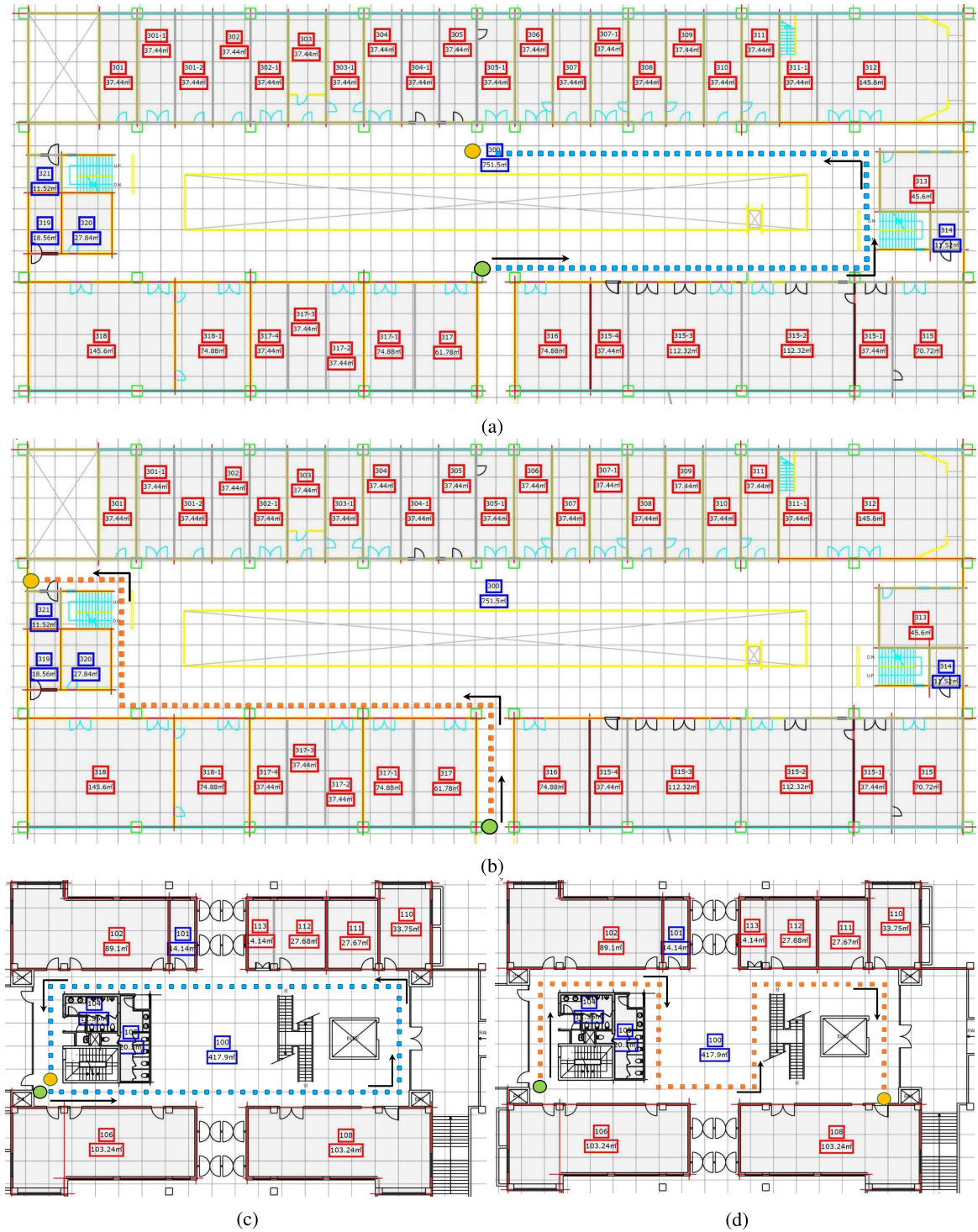
FIGURE 10. Paths trajectories followed for data collection, (a) EE building scenario 1, (b) EE building scenario 2, (c) RIC building scenario 1, and (d) RIC buildings scenario 2.

in the hall. Consequently, the hall is divided into multiple corridors and open spaces. However, there are no concrete walls in the hall, although, few concrete pillars are there. Due to the large size of the hall, the data collection points are separated by 2 m. Data collection involves Wi-Fi and magnetic field data at 228 points in the hall. An illustration of the data collection points and the view of the exhibition stalls are given in Figure 13.

V. STRUCTURE OF THE MagWi DATASET

MagWi is a hybrid dataset providing both the Wi-Fi and the magnetic field data. The dataset aims to accelerate the indoor
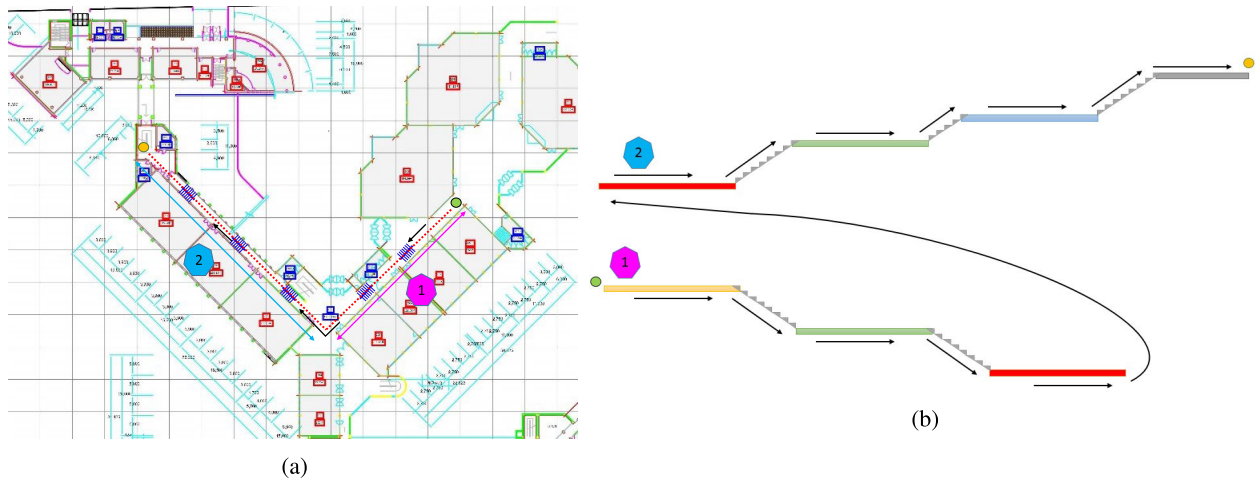
FIGURE 11. Data collection scenario for BE building, (a) Path followed for data collection, and (b) Details for the path and stairs.
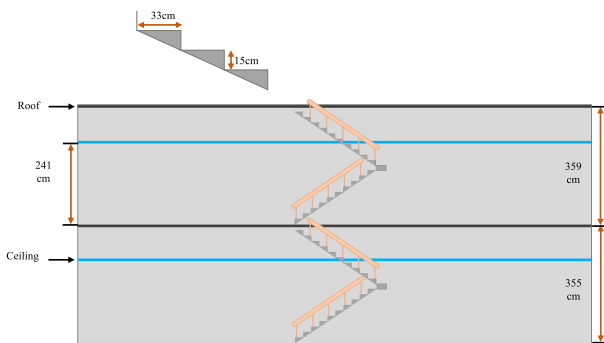

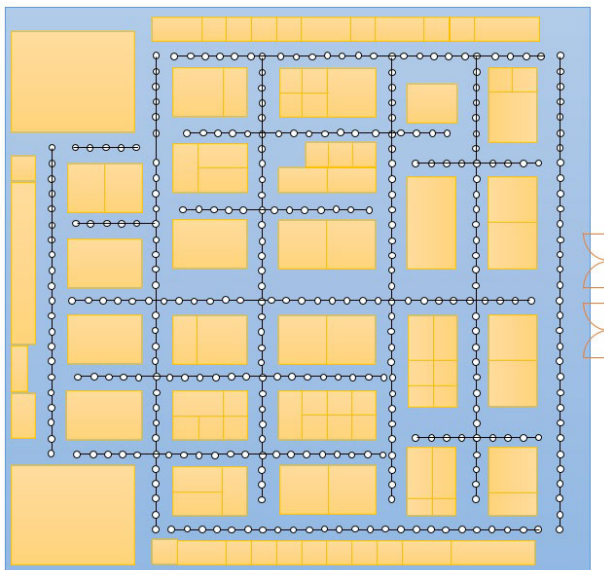**FIGURE 12.** Structure of 3 floor stairs used for data collection.

FIGURE 13. Data collection points in COEX exhibition hall.

positioning research using sensor fusion and is publicly available from the 'IEEE Data Port' at link[1] with the digital object identifier (DOI) https://dx.doi.org/10.21227/7g9v-6z48.

[1] https://ieee-dataport.org/open-access/magwi-benchmark-dataset-long-term-magnetic-field-and-wi-fi-data-involving-heterogeneous

The dataset contains both the magnetic field and Wi-Fi data which are stored in separate folders. The data are collected for Wi-Fi and magnetic field on the same location points which indicates that the data in different folders correspond to the same locations for different scenarios and users. The hierarchical structure of the dataset is shown in Figure 14. The MagWi contains two subfolders, one each for the magnetic field and Wi-Fi data, respectively. For the magnetic field data, two types of data are collected: continuous data and static data. The latter is collected at specified location points while the surveyors stand still at manually marked location points which are at a distance of 1 m from each other. The static data are collected 125 to 150 data samples at the sampling rate of 10 Hz. On the other hand, the former involves the data collection while the surveyor walks along the chosen path with a consistent speed between the starting and ending points as indicated on maps given in Figures 9 to 11. The sampling rate for continuous data is also 10 Hz. The former can be used to make the training data/fingerprint database while the latter aims at providing the data for various walking speeds of users. The data for the former category can be resampled depending upon the walking speed of the user.

Apart from two types of data for the magnetic field, the rest of the structure is the same for both Wi-Fi and magnetic field datasets. Figure 15 shows the structure of the magnetic field data under MagWi. The folder and subfolder structure are the same for static and continuous magnetic field data, as well as, the Wi-Fi data. The data for five selected buildings are stored in separate folders for magnetic field and Wi-Fi data, followed by the folders for three different orientation styles. Besides 'call listening', 'swinging', and 'navigation', the data for stairs are stored separately. The data for stairs is gathered only for the IT building and contains the data for three-floor stairs including both 'stairs down' and 'stairs up' events where 'stairs down' even refers to the user walking from 3rd to 1st floor and vice versa for 'stairs up' event.

Three scenarios are followed for each orientation where each scenario follows a different path geometry, as well as, starting and ending points. Five smartphones are used for data
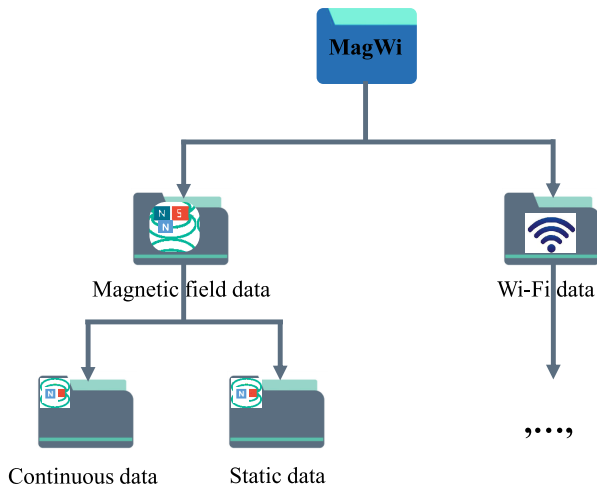
FIGURE 14. Folder and subfolder structure of MagWi dataset.

collection including A8, S8, and S9+ from Samsung Galaxy while G6 and G7 from LG. As mentioned previously, four users participate in the data collection process and the data for each user is stored separately. In the end, for each user, the data are stored as Excel files with '.xlsx' extension for small data size and easy processing. The following format is used to name the files, "IMU_IT Engineering_Scenario 1_User 2 (M-174cm)_Navigation_2018.02.20 120647.xlsx".

where

- IMU - indicates the inertial measurement unit data containing the magnetic field, accelerometer, gyroscope, and barometer sensor data.
- IT Engineering - the name of the building for data collection.
- Scenario 1 - shows the scenario followed for data collection.
- User 2 (M-174cm) - the user who collected the data, 'M' refers to 'male' while 174 cm is the height of the user.
- 2018.02.20 120647 - shows the time stamp in 'yyyy.mm.dd' format followed by the time in 'hh.mm.ss' format.

The same format is used for magnetic field and Wi-Fi file names except for the initial name where 'IMU' is replaced with 'WiFi' for the Wi-Fi data, as in "WiFi_CS Engineering_Scenario 2_User 2 (M-174cm)_Navigation_2021.08.12 170502.xlsx". Figure 16 shows the snapshot of the Excel sheet for the magnetic field data.

A. DETAILS OF WI-FI DATA RECORDS

Excel sheet for the Wi-Fi data contains 6 columns and the detail of each column is as follows.

① 'Time' indicates the date and time when the data are collected using the format 'yyyy.mm.dd hh.mm.ss'.

② 'X_pos' shows the x coordinate of the location point where the data are collected.

③ 'Y_pos' shows the y coordinate of the location point. Both x and y coordinates are local coordinates and assigned

with respect to the maps shown in Figures 11 to 12, where the top left corner indicates (1,1) for x, and y.

④ 'SSID' refers to the service set identifier to show the name of the Wi-Fi AP.

⑤ 'BSSID' refers to basic service set identifier which shows the unique address of a Wi-Fi AP. SSID may be the same for several APs, but each AP has a unique BSSID. For security, the original SSID and BSSID have been replaced.

⑥ 'RSS' shows the received signal strength for a particular Wi-Fi AP. In the dataset, the Wi-Fi records are sorted for the RSS value from highest to lowest.

B. DETAILS OF MAGNETIC FIELD DATA

Magnetic field data has 16 columns and contains the accelerometer, gyroscope, and barometer data, in addition to the magnetic field data.

① 'Time' shows the time stamp when the data are collected and stores the time stamp in 'yyyy.mm.dd hh.mm.ss' format.

② 'X_pos' refers to the local x coordinate of the location point where the data are collected.

③ 'Y_pos' refers to the local y coordinate of the location point. The x and y coordinates in the magnetic field file are available only for the static data where the data are collected while standing at a particular point. On the other hand, for the continuous data, since the user collects the data while walking, these coordinates are not recorded.

④, ⑤, ⑥ 'Mag_x', and 'Mag_y', 'Mag_z' represent the magnetic field x, y, and z components, respectively and indicates the North, West and vertical component of the magnetic field. Using x, y, and z, total magnetic field intensity can be calculated using

$$F = \sqrt{Mag_x^2 + Mag_y^2 + Mag_z^2} \qquad (1)$$

⑦, ⑧, ⑨ 'Acc_x', and 'Acc_y', 'Acc_z' represent the data for smartphone accelerometer and show acceleration of x, y, and z components, respectively. Total acceleration in m/s^2 can be calculated using

$$A = \sqrt{Acc_x^2 + Acc_y^2 + Acc_z^2} \qquad (2)$$

⑩, ⑪, ⑫ 'Gyro_x', and 'Gyro_y', 'Gyro_z' refer to the data from the gyroscope of the smartphone. The values are given for gyroscope x, y, and z components, respectively and the measuring unit is $radian/s$.

⑬, ⑭, ⑮ 'Orn_x', and 'Orn_y', 'Orn_z' shows the data for smartphone current orientation with x, y and z axes. Both static and continuous data record the orientation to track the orientation of the user smartphone. For the static data, it is used to find the user's phone orientation for data collection. On the other hand, for a continuous walk, it is used to track the user's orientation, as well as, the direction where the user is headed. The values for 'Gyro' and 'Orn' are different because they represent different items; the former refers to the data from the gyroscope while the latter is obtained using the rotation matrix, accelerometer data, and the magnetic field data at a particular location.
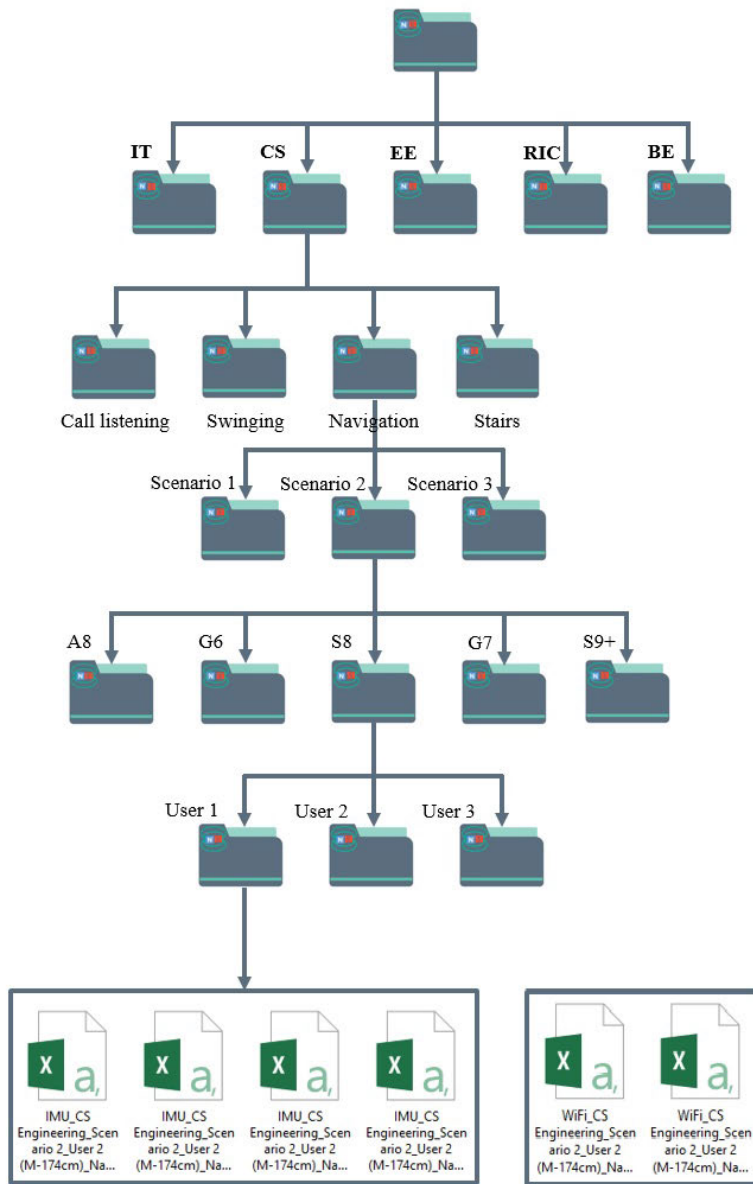
FIGURE 15. Structure of magnetic field and Wi-Fi data under MagWi dataset.

FIGURE 16. Detail of excel sheet for Wi-Fi data.

⑯ 'Pressure' shows the data for the atmospheric pressure received for the barometer of the smartphone. The data for the barometer is represented in hecto Pascal (hPa) and is equal to the millibar pressure unit.

VI. USING MAGWI FOR INDOOR POSITIONING

MagWi dataset contains the data from the magnetic field sensor, Wi-Fi, and IMU sensors. The data from such sensors is collected to be used for sensor fusion to enhance the positioning accuracy of indoor positioning approaches. Several experiments are performed to show the use of the MagWi hybrid dataset.

Initially, the positioning performance of the magnetic field data and Wi-Fi data is analyzed separately. For this purpose, the fingerprinting approach is utilized where a fingerprinting database is built for both the magnetic field and Wi-Fi data separately. Let x_i and y_i be the x and y coordinates for a given ground truth location point, the magnetic field fingerprint for that location is given as $\{mag_{x_i}, mag_{y_i}, mag_{z_i}\}$. Where mag_x, mag_y, and mag_z represents the normalized values for

① Time	② X-cord	③ Y-cord	④ Mag_x	⑤ Mag_y	⑥ Mag_z
2018.02.20 12:06:47	1	24	35.22	-7.92	-36.12
2018.02.20 12:06:47	1	24	35.4	-7.74	-35.52
2018.02.20 12:06:47	1	24	36	-7.5	-36.6
2018.02.20 12:06:47	1	24	36.54	-7.98	-37.8
2018.02.20 12:06:48	1	24	37.38	-8.34	-38.64
2018.02.20 12:06:48	1	24	37.86	-8.52	-39.66
2018.02.20 12:06:48	1	24	38.28	-9	-39.54
2018.02.20 12:06:48	1	24	38.34	-9.06	-39.12
2018.02.20 12:06:48	1	24	38.4	-9.48	-39.18
2018.02.20 12:06:48	1	24	38.16	-9.84	-39.66
2018.02.20 12:06:48	1	24	37.32	-9.6	-39.9

",,,"

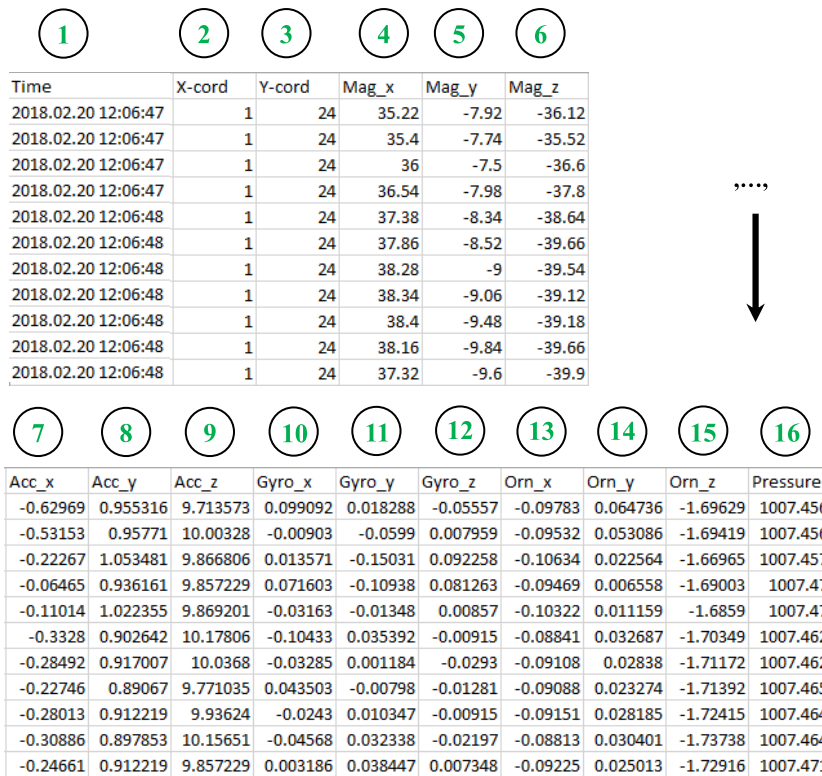| ⑦ Acc_x	⑧ Acc_y	⑨ Acc_z	⑩ Gyro_x	⑪ Gyro_y	⑫ Gyro_z	⑬ Orn_x	⑭ Orn_y	⑮ Orn_z	⑯ Pressure
-0.62969	0.955316	9.713573	0.099092	0.018288	-0.05557	-0.09783	0.064736	-1.69629	1007.456
-0.53153	0.95771	10.00328	-0.00903	-0.0599	0.007959	-0.09532	0.053086	-1.69419	1007.456
-0.22267	1.053481	9.866806	0.013571	-0.15031	0.092258	-0.10634	0.022564	-1.66965	1007.457
-0.06465	0.936161	9.857229	0.071603	-0.10938	0.081263	-0.09469	0.006558	-1.69003	1007.47
-0.11014	1.022355	9.869201	-0.03163	-0.01348	0.00857	-0.10322	0.011159	-1.6859	1007.47
-0.3328	0.902642	10.17806	-0.10433	0.035392	-0.00915	-0.08841	0.032687	-1.70349	1007.462
-0.28492	0.917007	10.0368	-0.03285	0.001184	-0.0293	-0.09108	0.02838	-1.71172	1007.462
-0.22746	0.89067	9.771035	0.043503	-0.00798	-0.01281	-0.09088	0.023274	-1.71392	1007.465
-0.28013	0.912219	9.93624	-0.0243	0.010347	-0.00915	-0.09151	0.028185	-1.72415	1007.464
-0.30886	0.897853	10.15651	-0.04568	0.032338	-0.02197	-0.08813	0.030401	-1.73738	1007.464
-0.24661	0.912219	9.857229	0.003186	0.038447	0.007348	-0.09225	0.025013	-1.72916	1007.471

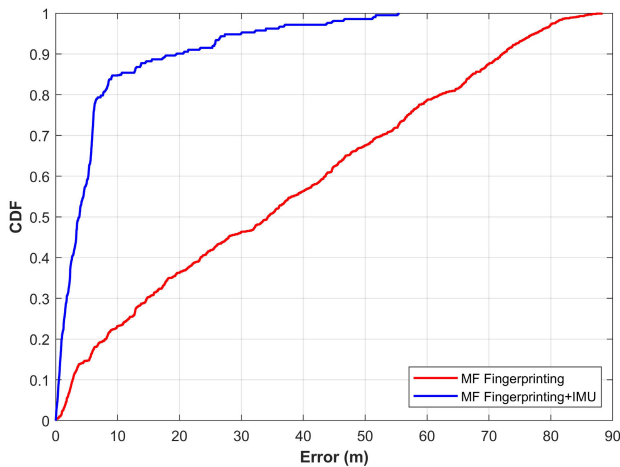**FIGURE 17.** Details of the magnetic field data records.

FIGURE 18. Positioning results using the magnetic field data alone and magnetic field data with IMU sensors.

magnetic field components. During the data collection, the magnetic field values are collected at a sampling rate of 10 Hz for 15 to 20 s. During the normalization, the average value of the collected data is calculated for the magnetic field data at a given point. This process is repeated for all the location points of an indoor area intended for positioning. For the current study, these points are separated by 1 m and points are marked by measuring the distance manually.

During the testing phase, the user collects the magnetic field data at an unknown location point and the current position is estimated by matching it with the pre-built fingerprint database. For matching the user data with the database, Euclidean distance is used. The position is estimated using k nearest neighbor (KNN) using $k = 7$, where the value of k is empirical. Positioning results using the magnetic field fingerprinting are shown in Figure 18. Results displayed in Figure 18 are obtained by running several experiments with a single user and single smartphone and the total number of estimated positions is 1130. As finding the optimal value of k is challenging, several experiments are performed to find a suitable value of k for increasing the positioning performance. A smaller value of k is sensitive to produce high positioning error while a large value increases the computation time. For this purpose, experiments are carried out using $k = 3, 5, 7$ to analyze the positioning accuracy. Although conventionally $k < 5$ is used, in the current case, $k = 7$ produces good results as compared to both $k = 3$ and $k = 5$. So, the adopted value of k for the current study is 7. For illustration, results for magnetic field-based positioning using KNN with $k = 3, 5, 7$ are shown in Figure 19.

Results are displayed for positioning using the magnetic field data alone, as well as, the magnetic field and IMU sensors data. For the second scenario, the data are collected while the user is walking along the dedicated path. For each 1 s, the data from the magnetic field sensor and IMU sensors are used for positioning. IMU sensors data are used to determine users' state of walking and the direction of walking. For matching the magnetic field data with the fingerprint
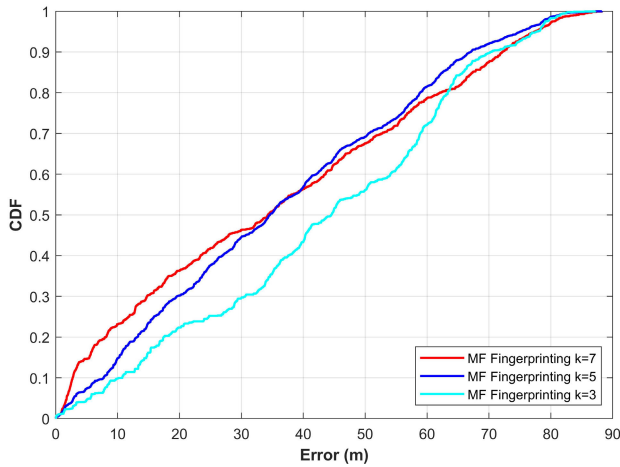
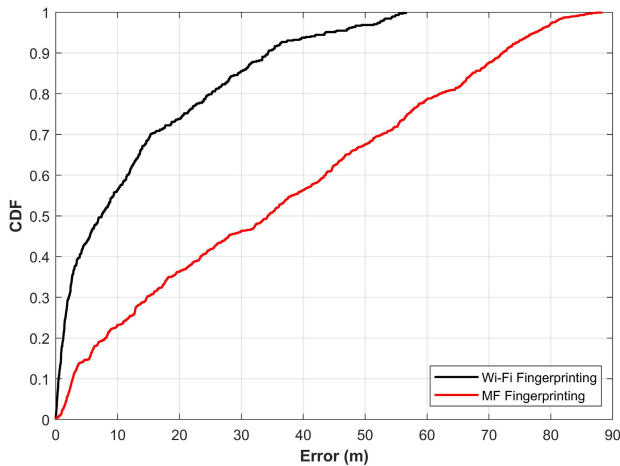**FIGURE 19.** Plot for positioning results with different values of *k*.

FIGURE 20. Positioning results using the Wi-Fi fingerprinting.

database, a second database is used where a 'spline' interpolation is used to generate the intermediate magnetic field data between two location points. In this method, instead of matching one value of mag_x, mag_y, and mag_z, a sequence of 10 values is matched which improves the positioning performance.

For Wi-Fi positioning, a fingerprinting approach similar to that of magnetic field positioning is adopted. Let x_i and y_i be the x and y coordinates for a given ground truth location point, Wi-Fi fingerprint for that location can be given as $\{(AP_{i_1}, RSS_{i_1}), (AP_{i_2}, RSS_{i_2}), \ldots, (AP_{i_n}, RSS_{i_n})\}$ where n represents total number of APs detected during the scan and RSS_{i_1} indicates the normalized RSS value for AP_1 at (x_i, y_i). During the scanning time of 15 to 20 scans, approximately 10 scans are carried out for detecting Wi-Fi APs and RSS values for each AP are normalized to make the fingerprint database. For positioning, the KNN approach with $k = 7$ is followed using the Euclidean distance measure. Results for Wi-Fi fingerprinting-based positioning are shown in Figure 20.

Results indicate that the positioning with the Wi-Fi data is superior to that of the magnetic field data. The primary reason for the higher performance of Wi-Fi is the uniqueness of

Wi-Fi fingerprints. The distribution of Wi-Fi signals is more unique than that of the magnetic field data. The magnetic field data distribution depends on the internal structure of a building which interferes with the earth's natural magnetic field and causes anomalies. Often, the magnetic field data fingerprints may be the same or very similar at different locations, especially in large indoor areas. Consequently, while using alone, the positioning accuracy of the magnetic field data is inferior to that of the Wi-Fi. Secondly, the fingerprint vector of the magnetic field data contains only mag_x, mag_y, and mag_z and may repeat for many locations. On the other hand, at any given location, many APs are visible making the Wi-Fi fingerprint vector several times larger which ultimately improves the uniqueness of the Wi-Fi fingerprint.

Due to the complementary nature of Wi-Fi and the magnetic field data, they are often combined to achieve better positioning performance. Several research works used Wi-Fi data with the magnetic field data to show that positioning accuracy is improved when magnetic field and Wi-Fi data are used together [55]–[57]. Similarly, the authors in [28] study the behavior of Wi-Fi and the magnetic field data for indoor positioning and confirm that Wi-Fi and magnetic field data are complementary and improve the positioning performance if used jointly.

To corroborate the findings of these research works, we performed fingerprinting positioning using Wi-Fi and magnetic field data. Initially, the Wi-Fi data are utilized to calculate a coarse position which is later used to restrict the search space in the magnetic field database. For this purpose, an extended Kalman filter (EKF) is leveraged for data fusion from IMU, WiFi, and the magnetic field. The purpose of these experiments is to show the practical use of MagWi and the feasibility of using the magnetic field and WiFi together to enhance the positioning accuracy. Positioning results are shown in Figure 21. Results prove that using Wi-Fi and the magnetic field data, positioning accuracy is enhanced substantially. For example, the mean value for the magnetic field data and Wi-Fi alone is 34.29 m and 12.86 m, respectively. However, when used together, the positioning performance is elevated with a mean value of 4.89 m; the maximum error is reduced as well. Similarly, when WiFi, magnetic field, and IMU data are fused, the mean error is further reduced to 4.31 m. Using the Wi-Fi and the magnetic field data with fingerprinting gives a positioning error of 10 m for 90%. The positioning error at 90% is approximately 9 m when IMU data are used along with the WiFi and magnetic field data. When used with state-of-the-art approaches, the positioning performance can be further improved.

VII. DISCUSSION AND CONCLUSION
The wide proliferation of modern smartphones accelerated the pace of positioning research to meet the needs of location-based services. Hence, a large number of positioning technologies and methods have been proposed to provide precise location information, both indoors and outdoors. Unlike the outdoor environment where GPS can be used to
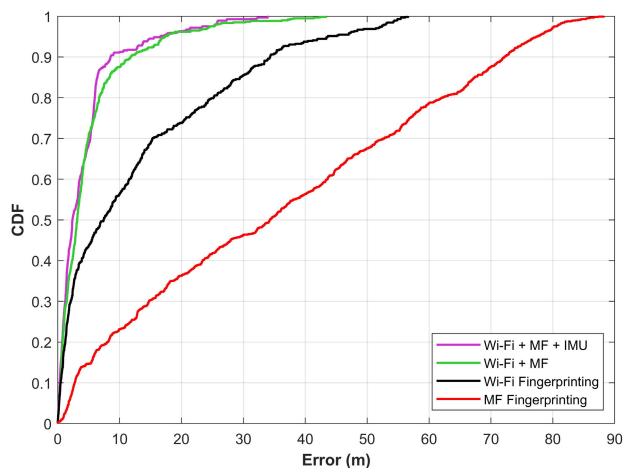
FIGURE 21. CDF graph for positioning using the Wi-Fi and magnetic field data.

estimate the accurate position of the user, indoor environments are complex and pose several extra challenges. Consequently, several indoor positioning technologies have been adopted like RFID, UWB, Wi-Fi, Bluetooth, and magnetic field. Contrary to RFID, UWB, and Bluetooth that require dedicated infrastructure for positioning, Wi-Fi and magnetic field do not need such infrastructure. Wide deployment of Wi-Fi APs provides the opportunity to use wireless signals for positioning while the magnetic field is the latest adoption of earth's natural magnetic field for positioning which is pervasive and does not require additional infrastructure.

Wi-Fi and the magnetic field data exhibit complementary characteristics and can be used together to overcome the limitations and enhance indoor positioning accuracy. However, the lack of publicly available datasets containing both Wi-Fi and magnetic field data limits the testing and evaluation of the state-of-the-art indoor positioning approaches. This research first discusses the important characteristics of Wi-Fi and magnetic field data and then designs the strategies to provide the data that can be used to extensively evaluate the important aspects of these positioning technologies.

The magnetic field data has shown great potential for indoor positioning and several important works can be found in the literature. The earth's magnetic field experiences disturbances in the indoor environment due to ferromagnetic materials like iron, nickel, and cobalt, etc. These disturbances are also called anomalies as they tend to interrupt the direction-finding using the magnetic field and studied for indoor positioning [13]. Due to various indoor settings of buildings, unique magnetic field anomalies are observed which can be used for indoor positioning, as well as, identification of particular buildings [58].

Despite possessing the potential for indoor positioning, magnetic field-based indoor positioning is still in its infancy and requires substantial research efforts for its practical deployment. For this purpose, several characteristics are to be investigated. For example, its long-term behavior is not studied very well. Although magnetic field data shows much

less mutation than that of Wi-Fi, smaller changes are observed over two years in the magnetic field intensity. To adjust such changes, the world magnetic model which determines the calculation of the magnetic field value is revised after five years to adjust for the mutation. Similarly, the magnetic field data is more tolerant to indoor infrastructural changes. Several experiments conducted to investigate the impact of furniture, human mobility, and dynamic environments indicate that the magnetic field data changes slightly. Similarly, human mobility slightly affects the magnetic field data. However, indoor changes involving the movement of items containing ferromagnetic materials like steel trolleys, or similar other objects influence the magnetic field data and degrades the performance of the magnetic field-based positioning approaches.

More challenging aspects of the magnetic field data are device heterogeneity, user's complex behavior involving multiple smartphone orientations, and its attitude in different kinds of buildings. Owing to a large number of different smartphone companies and models, the embedded magnetometer poses a huge challenge to devise a positioning framework that can provide similar positioning performance with all the smartphones. Manufactured from different vendors, these magnetometers possess different levels of sensitivity and noise tolerance which leads to different magnetic field data. MEMS sensors are inexpensive devices with limited accuracy and the data can easily get noisy and positioning erroneous. Device calibration can be used to overcome the change in the data for similar model devices [59], [60]. Similarly, the complex behavior of the user like calling, SMS sending, or phone in pocket, etc. is a huge problem for magnetic field-based indoor positioning. The magnetic field data change with the smartphone orientation which requires the transformation of the data. For transformation, smartphone attitude is to be tracked using the accelerometer and the gyroscope sensor which adds additional complexity. For fingerprint-based magnetic field positioning, data collection is laborious and time-consuming. However, it can be reduced using a crowdsourcing approach where multiple users collect the data over different times which can be combined into one database [61]–[63].

Wi-Fi-based indoor positioning is an extensively researched domain, yet, several challenges remain unresolved. For example, the direction of the smartphone, as well as, the orientation of the smartphone tend to change the RSS of APs. During this process, LOS APs may become NLOS which changes their RSS substantially. Smartphone diversity and antenna specifications are also a problem for Wi-Fi fingerprinting solutions. A more serious concern for Wi-Fi-based positioning solutions is the dynamic indoor environment. Doors opening and closing, temperature variations due to weather, and movements of the object all have a large influence on the measured RSS [37], [38], [42]. Human mobility can cause signal shadowing, absorption, and multipath with Wi-Fi-based positioning systems. Similarly, indoor infrastructural changes adversely affect the indoor positioning performance of Wi-Fi-based approaches. Exist-

ing publicly available datasets do not provide Wi-Fi data to investigate the above-mentioned issues.

Magnetic field data and Wi-Fi data possess several characteristics that can compensate for the limitations of each other and be used as a hybrid approach. For example, human mobility that tends to cause signal shadowing, absorption, and multipath for Wi-Fi positioning, seem to have almost no impact on the magnetic field data. In the same way, contrary to the adversely affected RSS due to indoor infrastructural changes, the magnetic field data is less/not affected by such changes. The low discernibility and the low number of elements of the magnetic field can be complemented by the Wi-Fi position vector where tens of APs' RSS values are available at a given point.

Magnetic field-based indoor positioning is an emerging paradigm that necessitates research from academia and industry to investigate its shortcomings and explore its potentials. However, several of its challenges can be eliminated or mitigated when used with the wi-Fi data. Due to the unavailability of publicly available hybrid datasets containing both Wi-Fi and magnetic field data, such aspects are not investigated properly. This study provides the dataset in the light of discussed important characteristics for Wi-Fi and the magnetic field data. The influence of device heterogeneity, spatial diversification, smartphone orientation, walking speed, and time-related mutation, can be studied using the presented dataset as it is collected using five different smartphones over almost five years. Similarly, the data collected from dynamic environments involving public exhibition halls serve to study the impact of human mobility on both wi-Fi and magnetic field data. Besides, the data are collected in a multi-floor environment and from uneven floor structures which is very helpful to determine the positioning performance of hybrid systems in multi-story buildings.

REFERENCES

[1] Saleforce Blog. *What are Conumsers Doing on Their Smartphones Anyway*. Accessed: May 19, 2020. [Online]. Available: https://www.salesforce.com/blog/2018/02/consumer-smartphone-use.html

[2] H. Falaki, R. Mahajan, S. Kandula, D. Lymberopoulos, R. Govindan, and D. Estrin, "Diversity in smartphone usage," in *Proc. 8th Int. Conf. Mobile Syst., Appl., Services*, 2010, pp. 179–194.

[3] L. M. Ni, Y. Liu, Y. C. Lau, and A. P. Patil, "LANDMARC: Indoor location sensing using active RFID," in *Proc. 1st IEEE Int. Conf. Pervas. Comput. Commun. (PerCom)*, Mar. 2003, pp. 407–415.

[4] E. M. Gorostiza, J. L. L. Galilea, F. J. M. Meca, D. S. Monzú, F. E. Zapata, and L. P. Puerto, "Infrared sensor system for mobile-robot positioning in intelligent spaces," *Sensors*, vol. 11, no. 5, pp. 5416–5438, May 2011.

[5] A. Alarifi, A. Al-Salman, M. Alsaleh, A. Alnafessah, S. Al-Hadhrami, M. Al-Ammar, and H. Al-Khalifa, "Ultra wideband indoor positioning technologies: Analysis and recent advances," *Sensors*, vol. 16, no. 5, p. 707, May 2016.

[6] Y. Liu, Q. Wang, J. Liu, and T. Wark, "MCMC-based indoor localization with a smart phone and sparse WiFi access points," in *Proc. IEEE Int. Conf. Pervas. Comput. Commun. Workshops*, Mar. 2012, pp. 247–252.

[7] P. Bahl and V. N. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system," in *Proc. IEEE Conf. Comput. Commun., 19th Annu. Joint Conf. IEEE Comput. Commun. Soc. (INFOCOM)*, vol. 2, Mar. 2000, pp. 775–784.

[8] M. Youssef and A. Agrawala, "The horus WLAN location determination system," in *Proc. 3rd Int. Conf. Mobile Syst., Appl., Services (MobiSys)*, 2005, pp. 205–218.

[9] R. Bruno and F. Delmastro, "Design and analysis of a Bluetooth-based indoor localization system," in *Proc. IFIP Int. Conf. Pers. Wireless Commun.* Berlin, Germany: Springer, 2003, pp. 711–725.

[10] A. Papapostolou and H. Chaouchi, "Orientation-based radio map extensions for improving positioning system accuracy," in *Proc. Int. Conf. Wireless Commun. Mobile Comput. Connecting World Wirelessly (IWCMC)*, 2009, pp. 947–951.

[11] H. Zou, X. Lu, H. Jiang, and L. Xie, "A fast and precise indoor localization algorithm based on an online sequential extreme learning machine," *Sensors*, vol. 15, no. 1, pp. 1804–1824, Jan. 2015.

[12] G. Lui, T. Gallagher, B. Li, A. G. Dempster, and C. Rizos, "Differences in RSSI readings made by different Wi-Fi chipsets: A limitation of WLAN localization," in *Proc. Int. Conf. Localization GNSS (ICL-GNSS)*, Jun. 2011, pp. 53–57.

[13] B. Li, T. Gallagher, A. G. Dempster, and C. Rizos, "How feasible is the use of magnetic field alone for indoor positioning?" in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, 2012, pp. 1–9.

[14] V. Pasku, A. De Angelis, G. De Angelis, D. D. Arumugam, M. Dionigi, P. Carbone, A. Moschitta, and D. S. Ricketts, "Magnetic field-based positioning systems," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 2003–2017, 3rd Quart., 2017.

[15] D. Hanley, A. B. Faustino, S. D. Zelman, D. A. Degenhardt, and T. Bretl, "MagPIE: A dataset for indoor positioning with magnetic anomalies," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Sep. 2017, pp. 1–8.

[16] J. Torres-Sospedra, D. Rambla, R. Montoliu, O. Belmonte, and J. Huerta, "UJIIndoorLoc-mag: A new database for magnetic field-based localization problems," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Oct. 2015, pp. 1–10.

[17] M. A. Nassar, M. Hasan, M. Khan, M. Sultana, M. Hasan, L. Luxford, P. Cole, G. Oatley, and P. Koutsakis, "WiFi-based localisation datasets for no-GPS open areas using smart bins," *Comput. Netw.*, vol. 180, Oct. 2020, Art. no. 107422.

[18] G. Mendoza-Silva, P. Richter, J. Torres-Sospedra, E. Lohan, and J. Huerta, "Long-term WiFi fingerprinting dataset for research on robust indoor positioning," *Data*, vol. 3, no. 1, p. 3, Jan. 2018.

[19] X. Song, X. Fan, C. Xiang, Q. Ye, L. Liu, Z. Wang, X. He, N. Yang, and G. Fang, "A novel convolutional neural network based indoor localization framework with WiFi fingerprinting," *IEEE Access*, vol. 7, pp. 110698–110709, 2019.

[20] E. Lohan, J. Torres-Sospedra, H. Leppäkoski, P. Richter, Z. Peng, and J. Huerta, "Wi-Fi crowdsourced fingerprinting dataset for indoor positioning," *Data*, vol. 2, no. 4, p. 32, Oct. 2017.

[21] P. Barsocchi, A. Crivello, D. La Rosa, and F. Palumbo, "A multisource and multivariate dataset for indoor localization methods based on WLAN and geo-magnetic field fingerprinting," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Oct. 2016, pp. 1–8.

[22] Z. Tóth and J. Tamás, "Miskolc IIS hybrid IPS: Dataset for hybrid indoor positioning," in *Proc. 26th Int. Conf. Radioelektronika (RADIOELEKTRONIKA)*, Apr. 2016, pp. 408–412.

[23] J. Torres-Sospedra, R. Montoliu, G. M. Mendoza-Silva, O. Belmonte, D. Rambla, and J. Huerta, "Providing databases for different indoor positioning technologies: Pros and cons of magnetic field and Wi-Fi based positioning," *Mobile Inf. Syst.*, vol. 2016, pp. 1–22, Jan. 2016.

[24] J. Talvitie, E. S. Lohan, and M. Renfors, "The effect of coverage gaps and measurement inaccuracies in fingerprinting based indoor localization," in *Proc. Int. Conf. Localization GNSS (ICL-GNSS)*, Jun. 2014, pp. 1–6.

[25] E. L. Gunnarsdóttir, "The earth's magnetic field," Ph.D. dissertation, 2012.

[26] A. Chulliat, S. Macmillan, P. Alken, C. Beggan, M. Nair, B. Hamilton, A. Woods, V. Ridley, S. Maus, and A. Thomson, "The US/UK world magnetic model for 2015–2020," Nat. Ocean. Atmos. Admin., Washington, DC, USA, Tech. Rep., 2015, doi: 10.7289/V5TB14V7.

[27] I. Ashraf, M. Kang, S. Hur, and Y. Park, "MINLOC: Magnetic field patterns-based indoor localization using convolutional neural networks," *IEEE Access*, vol. 8, pp. 66213–66227, 2020.

[28] Y. Shu, C. Bo, G. Shen, C. Zhao, L. Li, and F. Zhao, "Magicol: Indoor localization using pervasive magnetic field and opportunistic WiFi sensing," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 7, pp. 1443–1457, Jul. 2015.

[29] I. Ashraf, S. Hur, and Y. Park, "Enhancing performance of magnetic field based indoor localization using magnetic patterns from multiple smartphones," *Sensors*, vol. 20, no. 9, 2704, May 2020.

[30] K. P. Subbu, B. Gozick, and R. Dantu, "LocateMe: Magnetic-fields-based indoor localization using smartphones," *ACM Trans. Intell. Syst. Technol.*, vol. 4, no. 4, pp. 1–27, Sep. 2013.

[31] *Developer Library. Managing the Keyboard*, I Apple, Cupertino, CA, USA, 2016, p. 9, vol. 26.

[32] Y. Zheng, G. Shen, L. Li, C. Zhao, M. Li, and F. Zhao, "Travi-navi: Self-deployable indoor navigation system," *IEEE/ACM Trans. Netw.*, vol. 25, no. 5, pp. 2655–2669, Oct. 2017.

[33] H. Xie, T. Gu, X. Tao, H. Ye, and J. Lu, "A reliability-augmented particle filter for magnetic fingerprinting based indoor localization on smartphone," *IEEE Trans. Mobile Comput.*, vol. 15, no. 8, pp. 1877–1892, Aug. 2016.

[34] Farproc. *WiFi Analyzer*. Accessed: Feb. 22, 2021. [Online]. Available: https://play.google.com/store/apps/details?id=com.farproc.wifi.analyzer&hl=en&gl=US

[35] J.-G. Park, D. Curtis, S. Teller, and J. Ledlie, "Implications of device diversity for organic localization," in *Proc. IEEE INFOCOM*, Shanghai, China, Apr. 2011, pp. 3182–3190, doi: 10.1109/INFCOM.2011.5935166.

[36] I. Ashraf, S. Hur, and Y. Park, "Indoor positioning on disparate commercial smartphones using Wi-Fi access points coverage area," *Sensors*, vol. 19, no. 19, p. 4351, Oct. 2019.

[37] M. A. Bitew, R.-S. Hsiao, H.-P. Lin, and D.-B. Lin, "Hybrid indoor human localization system for addressing the issue of RSS variation in fingerprinting," *Int. J. Distrib. Sensor Netw.*, vol. 11, no. 3, Mar. 2015, Art. no. 831423.

[38] Z. Zheng, Y. Chen, T. He, F. Li, and D. Chen, "Weight-RSS: A calibration-free and robust method for WLAN-based indoor positioning," *Int. J. Distrib. Sensor Netw.*, vol. 11, no. 4, Apr. 2015, Art. no. 573582.

[39] J. Torres-Sospedra, A. Jiménez, A. Moreira, T. Lungenstrass, W.-C. Lu, S. Knauth, G. Mendoza-Silva, F. Seco, A. Pérez-Navarro, M. Nicolau, A. Costa, F. Meneses, J. Farina, J. Morales, W.-C. Lu, H.-T. Cheng, S.-S. Yang, S.-H. Fang, Y.-R. Chien, and Y. Tsao, "Off-line evaluation of mobile-centric indoor positioning systems: The experiences from the 2017 IPIN competition," *Sensors*, vol. 18, no. 2, p. 487, Feb. 2018.

[40] L. Xu, C. Feng, Y. Wang, Y. Yao, and M. Q.-H. Meng, "Variation of exterior telemetry links of capsule antenna ingested in human body," in *Proc. 8th World Congr. Intell. Control Automat.*, 2010, pp. 2269–2272.

[41] K. Subaashini, G. Dhivya, and R. Pitchiah, "ZigBee RF signal strength for indoor location sensing-experiments and results," in *Proc. 15th Int. Conf. Adv. Commun. Technol. (ICACT)*, 2013, pp. 50–57.

[42] S.-H. Fang, Y.-C. Cheng, and Y.-R. Chien, "Exploiting sensed radio strength and precipitation for improved distance estimation," *IEEE Sensors J.*, vol. 18, no. 16, pp. 6863–6873, Aug. 2018.

[43] L. Xu, F. Yang, Y. Jiang, L. Zhang, C. Feng, and N. Bao, "Variation of received signal strength in wireless sensor network," in *Proc. 3rd Int. Conf. Adv. Comput. Control*, Jan. 2011, pp. 151–154.

[44] A. Haeberlen, E. Flannery, A. M. Ladd, A. Rudys, D. S. Wallach, and L. E. Kavraki, "Practical robust localization over large-scale 802.11 wireless networks," in *Proc. 10th Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, 2004, pp. 70–84.

[45] S. P. Tarzia, P. A. Dinda, R. P. Dick, and G. Memik, "Indoor localization without infrastructure using the acoustic background spectrum," in *Proc. 9th Int. Conf. Mobile Syst., Appl., Services (MobiSys)*, 2011, pp. 155–168.

[46] Asahi Kasei Microdevices Corporation. *AK09918C*. Accessed: Apr. 19, 2020. [Online]. Available: https://www.akm.com/global/en/products/electronic-compass/ak09918c/

[47] STMicroelectronics. *LSM6dsl*. Accessed: Apr. 21, 2020. [Online]. Available: https://www.st.com/resource/en/datasheet/lsm6dsl.pdf

[48] Asahi Kasei Microdevices Corporation. *AK09916C*. Accessed: Apr. 20, 2020. [Online]. Available: https://www.akm.com/akm/en/file/datasheet/AK09916C.pdf

[49] Asahi Kasei Microdevices Corporation. *AK09915C*. Accessed: Apr. 19, 2020. [Online]. Available: https://www.akm.com/global/en/products/electronic-compass/ak09915c/

[50] BOSCH. *BMI-160*. Accessed: Apr. 18, 2020. [Online]. Available: https://www.bosch-sensortec.com/media/boschsensortec/downloads/datasheets/bst-bmi160-ds000.pdf

[51] TDK Corporation. *TDK InvenSense*. Accessed: Apr. 19, 2020. [Online]. Available: https://invensense.tdk.com/wp-content/uploads/2017/11/DS-000196-IAM-20680-v1.1-Typ.pdf

[52] F. D. Rosa, L. Xu, J. Nurmi, C. Laoudias, M. Pelosi, and A. Terrezza, "Hand-grip and body-loss impact on RSS measurements for localization of mass market devices," in *Proc. Int. Conf. Localization GNSS (ICL-GNSS)*, Jun. 2011, pp. 58–63.

[53] K. Kaemarungsi and P. Krishnamurthy, "Properties of indoor received signal strength for WLAN location fingerprinting," in *Proc. 1st Annu. Int. Conf. Mobile Ubiquitous Syst., Netw. Services (MOBIQUITOUS)*, 2004, pp. 14–23.

[54] Y. Chen, D. Lymberopoulos, J. Liu, and B. Priyantha, "FM-based indoor localization," in *Proc. 10th Int. Conf. Mobile Syst., Appl., Services (MobiSys)*, 2012, pp. 169–182.

[55] W. Zhang, R. Sengupta, J. Fodero, and X. Li, "DeepPositioning: Intelligent fusion of pervasive magnetic field and WiFi fingerprinting for smartphone indoor localization via deep learning," in *Proc. 16th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2017, pp. 7–13.

[56] R. Ban, K. Kaji, K. Hiroi, and N. Kawaguchi, "Indoor positioning method integrating pedestrian dead reckoning with magnetic field and WiFi fingerprints," in *Proc. 8th Int. Conf. Mobile Comput. Ubiquitous Netw. (ICMU)*, Jan. 2015, pp. 167–172.

[57] E. Wang, M. Wang, Z. Meng, and X. Xu, "A study of WiFi-aided magnetic matching indoor positioning algorithm," *J. Comput. Commun.*, vol. 5, no. 3, pp. 91–101, 2017.

[58] I. Ashraf, S. Hur, and Y. Park, "BLocate: A building identification scheme in GPS denied environments using smartphone sensors," *Sensors*, vol. 18, no. 11, p. 3862, Nov. 2018.

[59] N. Roy, H. Wang, and R. R. Choudhury, "I am a smartphone and I can tell my user's walking direction," in *Proc. 12th Annu. Int. Conf. Mobile Syst., Appl., Services*, Jun. 2014, pp. 329–342.

[60] P. Zhou, M. Li, and G. Shen, "Use it free: Instantly knowing your phone attitude," in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2014, pp. 605–616.

[61] N. Wahlström, M. Kok, T. B. Schön, and F. Gustafsson, "Modeling magnetic fields using Gaussian processes," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 3522–3526.

[62] A. Solin, M. Kok, N. Wahlstrom, T. B. Schön, and S. Sarkka, "Modeling and interpolation of the ambient magnetic field by Gaussian processes," *IEEE Trans. Robot.*, vol. 34, no. 4, pp. 1112–1127, Aug. 2018.

[63] S. Wang, H. Wen, R. Clark, and N. Trigoni, "Keyframe based large-scale indoor localisation using geomagnetic field and motion pattern," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2016, pp. 1910–1917.

IMRAN ASHRAF received the M.S. degree in computer science from the Blekinge Institute of Technology, Karlskrona, Sweden, in 2010, and the Ph.D. degree in information and communication engineering from Yeungnam University, Gyeongsan, South Korea, in 2018. He has worked as a Postdoctoral Fellow with Yeungnam University. He is currently working as an Assistant Professor with the Department of Information and Communication Engineering, Yeungnam University. His research interests include indoor positioning and localization, indoor location-based services in wireless communication, and sentiment analysis.

SADIA DIN was working as a Postdoctoral Researcher with Kyungpook National University, South Korea, from March 2020 to August 2020. She is currently working as an Assistant Professor with the Department of Information and Communication Engineering, Yeungnam University, South Korea. Her research interests include demosaicking and denoising using machine/deep learning, artificial learning, big data analytics, 5G, and the IoT.

MUHAMMAD USMAN ALI received the M.S. degree in computer network engineering from the University of Engineering and Technology at Taxila, Pakistan, and the Ph.D. degree from Yeungnam University, South Korea, in 2008 and 2018, respectively. He is currently working with the Department of Computer Science, University of Gujrat, Pakistan. His current research interests include multisensor fusion-based indoor positioning systems, Wi-Fi fingerprinting, indoor navigation and mapping, and computer vision technologies.

YOUSAF BIN ZIKRIA (Senior Member, IEEE) received the Ph.D. degree from the Department of Information and Communication Engineering, Yeungnam University, South Korea, in 2016. He has more than ten years of experience in research, academia, and industry in the field of information and communication engineering and computer science. He is currently working as an Assistant Professor with the Department of Information and Communication Engineering, College of Engineering, Yeungnam University, Gyeongsan, South Korea. He has authored more than 60 scientific peer-reviewed journals, conferences, patents, and book chapters.

YONGWAN PARK received the B.E. and M.E. degrees in electrical engineering from Kyungpook National University, Daegu, South Korea, in 1982 and 1984, respectively, and the M.S. and Ph.D. degrees in electrical engineering from the State University of New York at Buffalo, USA, in 1989 and 1992, respectively. From 1992 to 1993, he worked with the California Institute of Technology, as a Research Fellow. From 1994 to 1996, he served as a Chief Researcher for developing the IMT-2000 system at SK Telecom, South Korea. Since 1996, he has been a Professor of information and communication engineering with Yeungnam University, South Korea. From January 2000 to February 2000, he was an Invited Professor with the NTT DoCoMo Wireless Laboratories, Japan. In 2003, he was also a Visiting Professor with UC Irvine, USA. From 2009 to March 2017, he also served as the President for the Gyeongbuk Institute of IT Convergence Industry Technology (GITC), South Korea. His current research interests include 5G systems in communication, OFDM, PAPR reduction, indoor location-based services in wireless communication, and smart sensors (LIDAR) for smart cars. From 2008 to 2009, he served as the Director for the Technology Innovation Center for wireless multimedia by the Korean Government. He is serving as the Chairman for the 5G Forum Convergence Service Committee, South Korea.

SOOJUNG HUR received the B.S. degree from Daegu University, South Korea, in 2001, the M.S. degree in electrical engineering from San Diego State University, San Diego, in 2004, and the M.S. and Ph.D. degrees in information and communication engineering from Yeungnam University, South Korea, in 2007 and 2012, respectively. She is currently working as a Research Professor with the Mobile Communication Laboratory, Yeungnam University. Her current research interests include the performance of mobile communication, indoor/outdoor location, and unnamed vehicle.

• • •