SCIENTIFIC DATA

Check for updates

OPEN

DATA DESCRIPTOR

OutFin, a multi-device and multi-modal dataset for outdoor localization based on the fingerprinting approach

Fahad Alhomayani & Mohammad H. Mahoor ✉

In recent years, fingerprint-based positioning has gained researchers' attention since it is a promising alternative to the Global Navigation Satellite System and cellular network-based localization in urban areas. Despite this, the lack of publicly available datasets that researchers can use to develop, evaluate, and compare fingerprint-based positioning solutions constitutes a high entry barrier for studies. As an effort to overcome this barrier and foster new research efforts, this paper presents OutFin, a novel dataset of outdoor location fingerprints that were collected using two different smartphones. OutFin is comprised of diverse data types such as WiFi, Bluetooth, and cellular signal strengths, in addition to measurements from various sensors including the magnetometer, accelerometer, gyroscope, barometer, and ambient light sensor. The collection area spanned four dispersed sites with a total of 122 reference points. Each site is different in terms of its visibility to the Global Navigation Satellite System and reference points' number, arrangement, and spacing. Before OutFin was made available to the public, several experiments were conducted to validate its technical quality.

Background & Summary

Location-Based Services (LBS) has become a multibillion-dollar industry that is expected to continue to steadily grow over the upcoming years[1]. Some of these services include location-based marketing[2], authentication[3], gaming[4], and social networking[5], among others. A key enabling technology at the heart of such services is positioning[6]. However, the de facto standard for positioning, the Global Navigation Satellite System (GNSS), has two major issues that limit the use of LBS. First, the availability and accuracy of GNSS are severely degraded in urban areas due to shadowing and multipath effects[7]. Second, GNSS chipsets are notorious for being power-hungry, which is problematic for power-constrained devices such as smartphones and smartwatches[8]. A more energy-efficient approach for positioning is achieved using cellular networks. Yet, the offered accuracy, which is in the order of tens[9] to hundreds[10] of meters, fails to satisfy the accuracy requirements imposed by many services and applications.

Recently, in an attempt to devise positioning solutions that can yield better performance, researchers have turned their attention to *fingerprinting*, a positioning technique that has achieved great success in the indoor positioning domain, a domain where GNSS signals are generally unavailable[11]. Fingerprinting is used to identify spatial locations based on location-dependent measurable features (location fingerprints). These fingerprints can be of different types such as WiFi fingerprints[12], Bluetooth fingerprints[13], cellular fingerprints[14], and magnetic field fingerprints[15]. From an implementation perspective, the fingerprinting approach is a two-phase process that consists of an *offline phase* and an *online phase*. During the offline phase, *site surveying* is performed by sampling fingerprints of an area of interest at predefined *reference points* (RPs). Fingerprints are often sampled using a smartphone or a dedicated data acquisition platform. Fingerprints, along with the coordinates at which they were sampled, are stored in a database. The data is then used to train a machine learning algorithm to learn a function that best maps sampled fingerprints to their ground truth coordinates. Afterward, the learned function is utilized during the online phase to infer a user's coordinates given the fingerprints measured at the user's location. The process of fingerprinting is visually depicted in Fig. 1.

Department of Electrical and Computer Engineering, Ritchie School of Engineering and Computer Science, University of Denver, Denver, CO, 80208, USA. ✉e-mail: mmahoor@du.edu
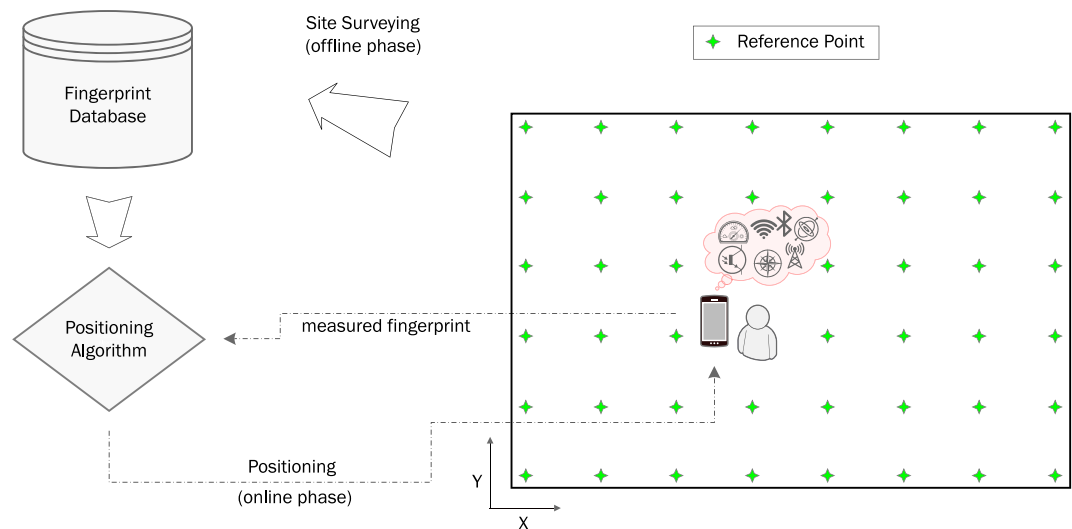
Fig. 1 A graphical representation of the fingerprinting approach for positioning.

Despite its low complexity and ability to produce accurate location estimates, the main drawback of finger-printing is the laborious and time-consuming site surveying task. This drawback has led many studies to resort to either simulated[16] or crowdsourced data[17], where the former never fully reflects the real world and the latter may suffer from integrity and consistency problems. The proposal of OutFin aims at addressing these drawbacks by making real-world measurements and reliable ground truth coordinates publicly available. Table 1 summarizes the main aspects of publicly available fingerprinting datasets published since 2014. Compared to these datasets, OutFin combines several features that place it in a unique position:

- To the best of our knowledge, OutFin is the first multi-modal, outdoor fingerprints dataset to be publicly available.
- The data was collected using two contemporary smartphones rather than outdated smartphones or custom-built platforms.
- The data was collected at highly granular RPs with 61 to 183 centimeters (cm) spacing.
- OutFin not only provides location fingerprints, but it also provides information about the devices that generated them (e.g., the service set identifier of an access point, the communication protocol of a Bluetooth device, and the number of neighboring cells of a serving cell).
- OutFin is accompanied by an interactive map that provides various information about the collection environment, such as RP coordinates (both ground truth and Global Positioning System (GPS) estimates) and building ground elevations and heights.

In addition to facilitating the research and development of outdoor positioning solutions that are based on the fingerprinting approach, OutFin might spur innovation in other research realms, including but not limited to: machine learning[18], Bayesian optimization[19], simultaneous localization and mapping[20], and map-matching[21].

Methods

Data acquisition platform. OutFin was created using two smartphones for data acquisition: Samsung's Galaxy S10+ (Phone 1) and Google's Pixel 4 (Phone 2). The former was released in the U.S. market on March 8, 2019, while the latter was released on October 24, 2019. Both smartphones ran on Android 10, released on September 3, 2019. The motivation behind choosing Android-powered smartphones was twofold. First, Android provides application programming interfaces (APIs) that allow for acquiring raw data at the hardware level. Second, Android-powered smartphones account for over 74 of the market share worldwide[22]. The two smartphones were attached to a tripod head using a dual mount that horizontally separated them by 10 (see Fig. 2 (Site 1)). Both smartphones were in portrait mode. The tripod kept them at a fixed height of 132. The tripod head was adjusted to tilt the smartphones at a ~40 degree (°) angle to the vertical plane. The same set of third-party apps used for data collection were installed on both smartphones. These apps, which can be downloaded from the Google Play Store, included: WiFi Analyzer Pro (App 1)[23], Bluetooth Scanner Extreme Edition (App 2)[24], NetMonitor Pro (App 3)[25], and Physics Toolbox Sensor Suite Pro (App 4)[26]. The apps allowed for conveniently collecting and exporting WiFi, Bluetooth, cellular, and sensor data, respectively.

Data collection environment. Data collection was performed at the University of Denver's campus where four separate sites were considered. The motivation behind collecting data at separate sites was to offer diversity. For instance, each site is different in terms of its reference points' number, arrangement, and spacing. Also, due to different ground elevations and heights of surrounding buildings, each site has different visibility to the GNSS. This is reflected by GPS errors produced at a given site. The mean GPS error was 12.1 meters (m), 11.4 m, 4.3 m,

Dataset	Year	Category	Environment	Data type(s)	Device type(s)	# of samples	Granularity
UJIIndoorLoc[45]	2014	Indoor	Three university buildings	WiFi	Smartphone, Tablet	Tens of thousands	Medium
UJIIndoorLoc-Mag[46]	2015	Indoor	A research lab	sensor	Smartphone	Tens of thousands	Medium
Dataset described in[47]	2016	Indoor	A research facility	WiFi, sensor	Smartphone, Smartwatch	Tens of thousands	High
Dataset described in[48]	2016	Indoor	A university building	WiFi, Bluetooth, sensor	Smartphone	Thousands	High
PerfLoc[49]	2016	Indoor	An office building, two industrial warehouses, and a subterranean structure	WiFi, cellular, sensor	Smartphone	Millions	Medium
AmbiLoc[50]	2017	Indoor	An apartment and two university buildings	TV, FM, cellular	Dedicated data acquisition platform	Thousands	Medium
MagPIE[51]	2017	Indoor	Three university buildings	sensor	Smartphone	Hundreds of thousands	High
Dataset described in[52]	2018	Indoor	A university library	WiFi	Smartphone	Hundreds of thousands	High
Dataset described in[53]	2018	Indoor	Four residential homes	Bluetooth, sensor	Dedicated data acquisition platform	Hundreds of thousands	High
Dataset described in[54]	2018	Indoor	A university library	Bluetooth	Smartphone	Thousands	Medium
Dataset described in[55]	2018	Indoor	A research facility	Bluetooth	Smartphone, Dedicated data acquisition platform	Millions	High
Dataset described in[56]	2018	Outdoor	A large-scale urban area and a large-scale rural area	Sigfox, LoRaWAN	Dedicated data acquisition platform	Hundreds of thousands	Low
Dataset described in[57]	2019	Indoor	Two university buildings	Bluetooth	Smartphone	Thousands	High
Dataset described in[45]	2019	Indoor, Outdoor	Worldwide	Cellular	Smartphone	Millions	Low
OutFin[33]	2020	Outdoor	A university campus	WiFi, Bluetooth, cellular, sensor	Smartphone	Hundreds of thousands	High

Table 1. A comparison of the main aspects of publicly available fingerprinting datasets published since 2014. **Dataset**: the name of the dataset (if indicated) and a reference to its description. **Year**: the year the dataset was made available. **Category**: indicates whether the data was collected indoors or outdoors. **Environment**: a brief description of the collection environment. **Data type(s)**: the type(s) of data that was collected. **Device type(s)**: the type(s) of devices used to collect the data. **# of samples**: the highest place value of the number of samples in the dataset. **Granularity**: a descriptor indicating how close the RPs were to each other; High: indicates a spacing of fewer than 2 meters, Medium: indicates a spacing between 2 and 8 meters, and Low: indicates a spacing of greater than 8 meters.
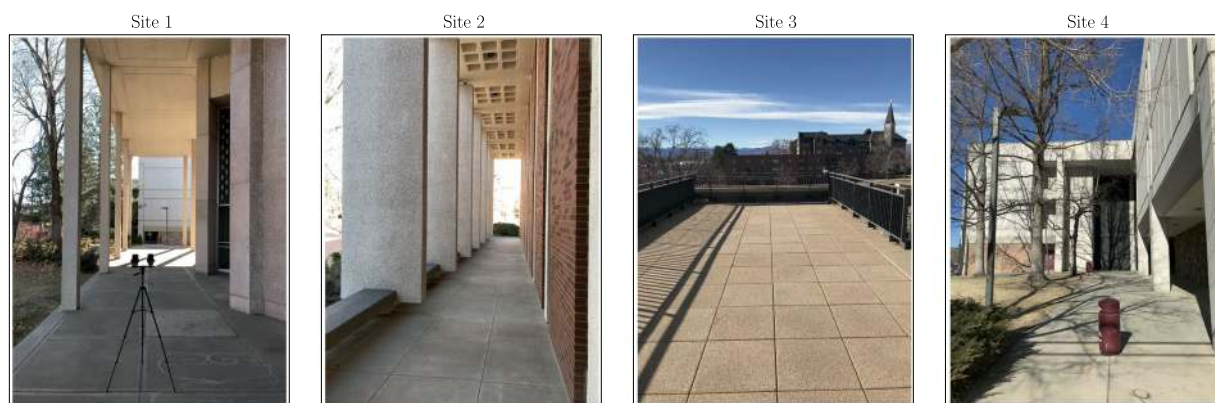

Fig. 2 Pictures of the four sites where data was collected.

and 12.7 m for the first, second, third, and fourth site, respectively. GPS estimates are provided in OutFin to help researches compare their system's performance to that obtained by GPS. A description of the data collection sites is provided below:

Site 1: Site 1 represents a portion of a covered sidewalk next to the east side of the 11.8 high Boettcher Auditorium (see Fig. 2). Site 1 contained 31 RPs arranged in three north-to-south lines (see Fig. 3). The spacing between RPs in each line was fixed at 152.5 and the distance between lines was fixed at 76.25.

Site 2: Site 2 is ∼245 north of Site 1 and represents a portion of a covered sidewalk next to the north side of the 11.5 high Sie International Relations Complex (see Fig. 2). Site 2 contained 23 RPs arranged in a single east-to-west line (see Fig. 3). The spacing between RPs was fixed at 101.5.
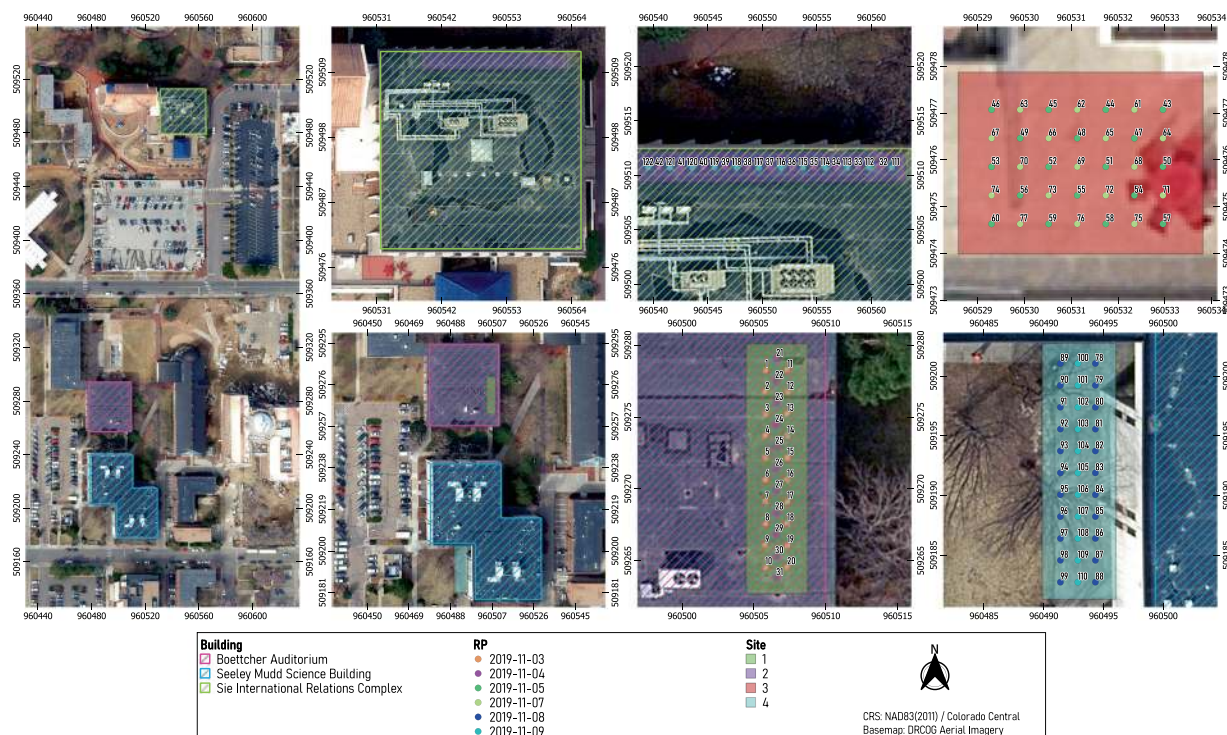
Fig. 3 An aerial map of the collection environment showing the four collection sites and the 122 RPs. RPs are color-coded according to the date of collection.

Site 3: Site 3 is ∼40 south of Site 2 and represents a portion of an open terrace next to the south side of the Sie International Relations Complex (see Fig. 2). Site 3 contains 35 RPs arranged in a seven-column and five-row grid (see Fig. 3). The spacing between column RPs and row RPs were fixed at 61.

Site 4: Site 4 is ∼288 south of Site 3 and represents a portion of an open sidewalk by the south and west sides of the 13.4 high Seeley Mudd Science Building (see Fig. 2). Site 4 contains 33 RPs arranged in a three-column and eleven-row grid (see Fig. 3). The spacing between column RPs was fixed at 183, while the spacing between row RPs was fixed at 146.5.

Each RP is uniquely identified by an integer (an ID number) that symbolizes its order in the collection campaign. For example, data collection started with RP 1 on November 3, 2019, and ended with RP 122 on November 9, 2019. The ground truth locations of RPs belonging to a site are expressed with respect to a local frame of reference. Additionally, the easting and northing (X,Y) coordinates of all RPs were provided with respect to a global coordinate system (i.e., NAD83(2011)/Colorado Central). This was accomplished with help from the university's Department of Geography & the Environment and by using a geographic information system software[27].

Procedure. Data collection spanned six days (3–5/11/2019 and 7–9/11/2019) and involved four sites with a total of 122 RPs. Due to the fact that rain could severely affect wireless signal measurements, we did not collect any data on rainy days. The RPs surveyed each day are indicated in Fig. 3. The sequence of steps performed during a day of data collection are described below:

Step 1: Before mounting the smartphones to the tripod, App 4 was launched to collect magnetic field measurements by rotating the smartphones around their X, Y, and Z axes multiple times (see Fig. 4). This process was performed for at least two minutes at a sampling rate of 1 Hertz (Hz). The resultant data was exported as a comma-separated values (CSV) file, named with the smartphone's name and date (e.g., `Phone1_051119.csv`). Such data can be used to offset the hard-iron distortion caused by placing the smartphones close to each other. After this process, the smartphones were mounted to the tripod and placed at the RP where data was to be collected.

Step 2: App 1 was launched to collect WiFi data, ensuring that at least two WiFi scans were performed along the four cardinal directions by routing the tripod head counterclockwise, ∼90 at a time. A WiFi scan recorded the received signal strength (RSS) from all access points (APs) in range in addition to information about the APs themselves. Android only supports passive scanning, and the duration of a scan varies depending on the smartphone's WiFi hardware and firmware. However, Google recently released a restriction that limits the frequency of scans that an app can perform to only four times in a 2-minute period[28]. This restriction applies to Android 9 and higher. The app reported scan results approximately every 30 seconds for Phone 1 and every 25 seconds for Phone 2. For Site 1 and 4's RPs, data collection started facing south and ended facing west. For Site 2 and 3's RPs, data collection started facing west and ended facing north.
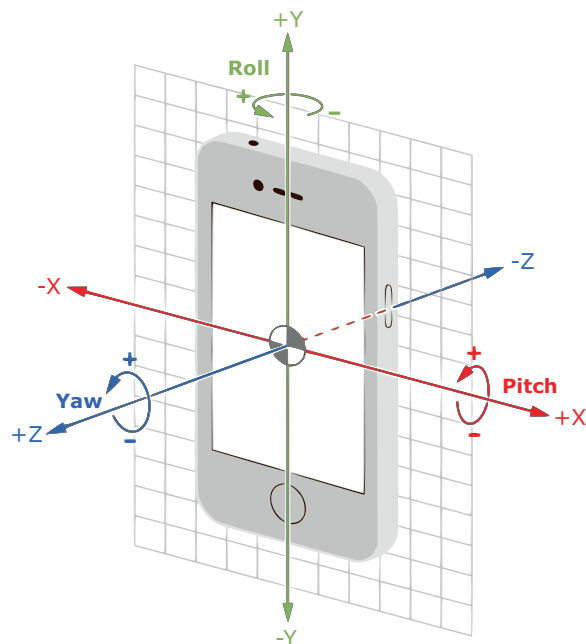
Fig. 4 Illustration of the X, Y, and Z axes relative to a typical smartphone. Figure reproduced from[44].

Collecting data along four directions mitigates the shadowing effect caused by the body of the data collector who is constantly facing the smartphone screens. Scan outcomes were exported as a CSV file, named with the smartphone's model as a prefix and the RP's ID as a suffix (e.g., `Phone2_WiFi_73.csv`).

Step 3: App 2 was launched to collect Bluetooth data. Android allows active Bluetooth scanning; thus, scans can be triggered by a user-level app. A Bluetooth scan involves an inquiry scan of approximately 12 seconds, followed by a page scan for each discovered device to retrieve its information and the RSS[29]. The duration of a scan, for both smartphones, took anywhere between 15 and 30 seconds, primarily depending on the number of discoverable devices in the area. As in Step 2, the shadowing effect was accounted for by performing two scans along each cardinal direction. Scan results were exported as a CSV file with a naming convention like that described in Step 2 (e.g., `Phone1_Bluetooth_29.csv`).

Step 4: App 3 was launched to collect cellular data. A smartphone's cellular modem constantly scans the cellular network for cell selection/reselection and handover purposes. Android provides APIs to extract information associated with scans such as Reference Signal Received Power (RSRP) and cell identity information[30]. The sampling frequency can be set manually and was fixed to 1. As noted in Step 2, the shadowing effect was accounted for by collecting at least fifteen samples along each cardinal direction. Collected data was exported as a CSV file with a naming convention like that described previously (e.g., `Phone2_Cellular_14.csv`). Moreover, App 3 allowed for collecting GPS data as part of the data record. The GPS readings corresponding to RPs belonging to the same site were extracted and stored under a CSV file named with the site's name as a prefix and the smartphone's model and app name as a suffix (e.g., Site1_GPS_Phone1_App3.csv).

Step 5: App 4 was launched to collect sensor data. A smartphone's built-in sensors can be classified as either hardware-based, such as the magnetometer and gyroscope, or software-based, such as the gravity and linear acceleration sensors. Android provides APIs for accessing and acquiring raw sensor data at defined rates[31]. The sampling frequency was set to 1. Although sensor measurements are not subject to the shadowing effect, data was collected along the four cardinal directions to both conform with the survey pattern established above and diversify the dataset since magnetic field strength can vary greatly even within a small area (in the orders of a few centimeters or less)[32]. At least fifteen samples were collected along each direction, following the same directions described in Step 2. Sensor data was exported as a CSV file with a naming convention like that described previously (e.g., `Phone1_Sensors_58.csv`). App 4 also allowed for collecting GPS data as part of the data record. As in Step 4, the GPS readings corresponding to RPs belonging to the same site were extracted and stored under a CSV file with a naming convention like that described in Step 4 (e.g., `Site3_GPS_Phone2_App4.csv`).

Step 6: The tripod was moved to the next RP and Steps 2–5 were repeated. This process continued until all RPs designated for a given day were surveyed.

Data Records

On April 2, 2020, the OutFin dataset was made publicly available on figshare[33]. Figure 5 shows the dataset's file structure and presents an overview of all CSV file types, their field labels, and a data record example. A description of the CSV file types and their field labels is provided below:
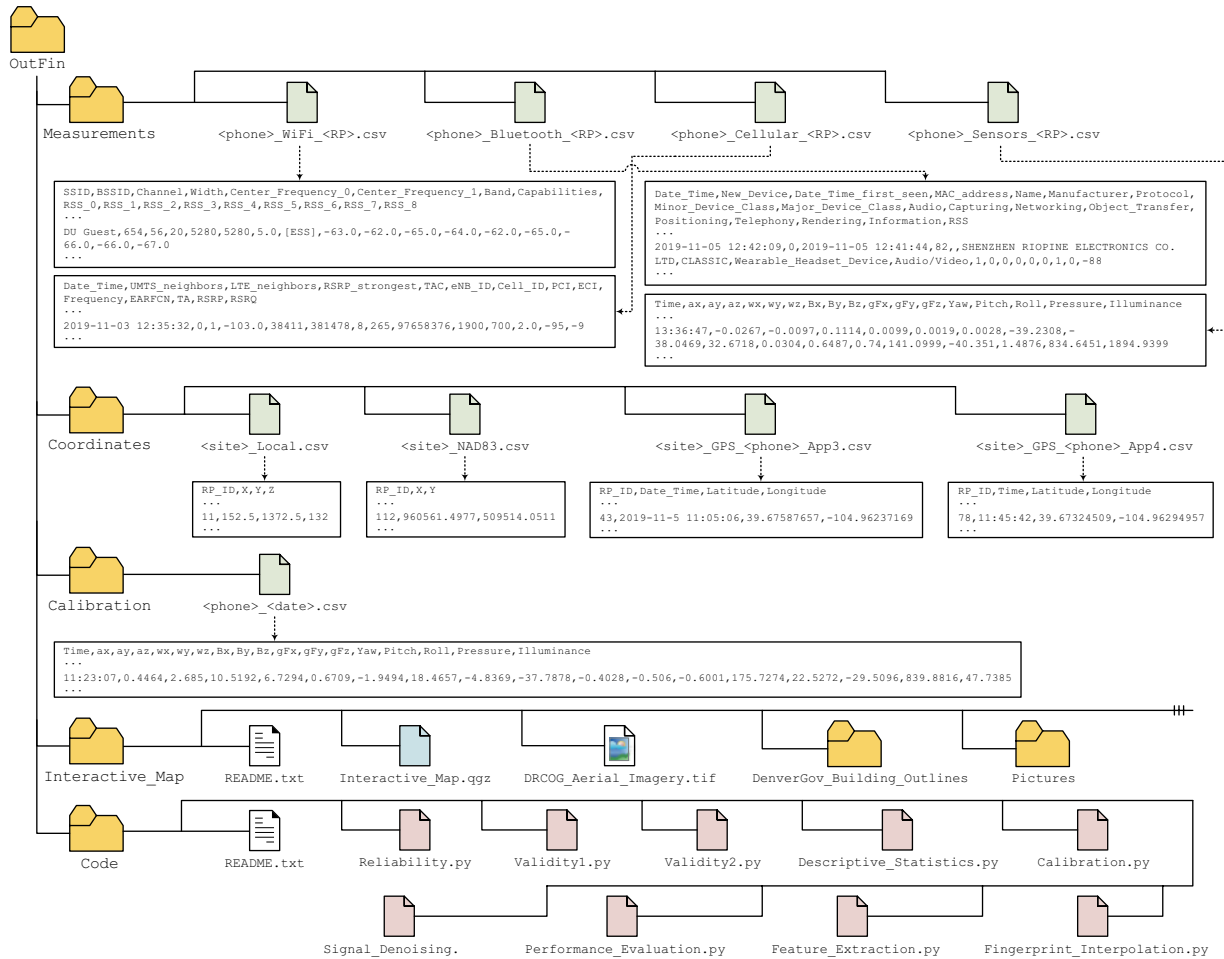
Fig. 5 Directory tree of the OutFin dataset along with CSV file types and example data records. `<phone>` ∈ {Phone1,Phone2}, `<RP>` ∈ {1,2,...,122}, `<site>` ∈ {Site1,Site2,Site3,Site4}, and `<date>` ∈ {031119,041119,051119,071119,081119,091119}.

I. `<phone>_WiFi_<RP>.csv` contains WiFi data collected by a smartphone via App 1:

1. `SSID`: The Service Set IDentifier (i.e., the AP's network name).
2. `BSSID`: The Basic Service Set IDentifier (i.e., the AP's media access control address (MAC address)) encoded as an integer.
3. `Channel`: The channel number that the AP uses for communication.
4. `Width`: The bandwidth of the channel in megahertz (MHz); can be 20, 40, or 80 MHz.
5. `Center_Frequency_0`: The center frequency of the primary channel in MHz.
6. `Center_Frequency_1`: The center frequency of the 40 or 80 MHz-wide channel in MHz. If a 20-MHz channel is used, then `Center_Frequency_1` ≡ `Center_Frequency_0`.
7. `Band`: The AP's frequency band in gigahertz (GHz); can be either 2.4 or 5 GHz.
8. `Capabilities`: Describes the authentication, key management, and encryption schemes supported by the AP.
9–17. `RSS_0-RSS_8`: The Received Signal Strengths in decibel-milliwatts (dBm), with respect to the back-to-back scans.

II. `<phone>_Bluetooth_<RP>.csv` contains Bluetooth data collected by a smartphone via App 2:

1. `Date_Time`: The date and time the scan was triggered as `YYYY-MM-DD` and `hh:mm:ss`. Denver, Colorado is in the Mountain Time Zone, which is seven hours behind Coordinated Universal Time (UTC-07:00).
2. `New_Device`: A binary flag that is set to 1 if the remote Bluetooth device is discovered for the first time at the current RP.
3. `Date_Time_first_seen`: The date and time the device was first discovered at the current RP. The date and time formats are as described above.
4. `MAC_address`: The device's MAC address encoded as an integer.
5. `Name`: The device's friendly name.
6. `Manufacturer`: The device's manufacturer name.

7. `Protocol`: The Bluetooth protocol that the device uses for communication; can be CLASSIC (Basic Rate/Enhanced Data Rate (BR/EDR)), BLE (Bluetooth Low Energy), or DUAL (BR/EDR + BLE).

8, 9. `Minor_Device_Class`, `Major_Device_Class`: Indicates the device's minor and major classes, respectively, as specified by the Bluetooth Special Interest Group (SIG)[34].

10–17. `Audio`, `Capturing`, `Networking`, `Object_Transfer`, `Positioning`, `Telephony`, `Rendering`, `Information`: Binary flags that are set to 1 if the device is associated with any of the eight service classes specified by the Bluetooth SIG[34].

18. `RSS`: The Received Signal Strength in dBm.

III. `<phone>_Cellular_<RP>.csv` contains cellular data collected by a smartphone via App 3. It should be noted that the entire collection environment was covered by Long-Term Evolution (LTE) cells. The Public Land Mobile Network (PLMN) identifier is 310410:

1. `Date_Time`: The date and time the sample was captured. The date and time formats are as described above.
2. `UMTS_neighbors`: The number of neighboring Universal Mobile Telecommunications Service (UMTS) cells.
3. `LTE_neighbors`: The number of neighboring LTE cells.
4. `RSRP_strongest`: The Reference Signal Received Power, in dBm, corresponding to the strongest neighboring cell, which employs the same technology as the serving cell.
5. `TAC`: The Tracking Area Code, which uniquely defines a group of cells within a PLMN.
6. `eNB_ID`: The E-UTRAN (Evolved-UMTS Terrestrial Radio Access Network) NodeB IDentifier that is used to uniquely identify an eNB (i.e., a base station in LTE) within a PLMN.
7. `Cell_ID`: The Cell IDentifier, which is an internal descriptor for a cell. It can take any value between 0 and 255.
8. `PCI`: The Physical Cell Identifier that is used to indicate the physical layer identity of a cell. It can take any value between 0 and 503.
9. `ECI`: The E-UTRAN Cell Identifier that is used to uniquely identify a cell within a PLMN. ECI = 256 × eNB_ID + Cell_ID.
10. `Frequency`: The downlink frequency band in MHz.
11. `EARFCN`: The downlink E-UTRAN Absolute Radio Frequency Channel Number.
12. `TA`: The Timing Advance value which ranges from 0 to 1282. A change of 1 in TA corresponds to a 156m round-trip distance[35]. For example, if TA = 7, then the eNB is located within a 546 radius from the smartphone.
13. `RSRP`: The Reference Signal Received Power in dBm.
14. `RSRQ`: The Reference Signal Received Quality in decibel (dB).

IV. `<phone>_Sensors_<RP>.csv` contains sensor data collected by a smartphone via App 4:

1. `Time`: The time the sample was captured. The time format is as described above.
2–4. `ax, ay, az`: The linear acceleration, in meters per second squared (m/s^2), along the smartphone's X, Y, and Z axes, respectively.
5–7. `wx, wy, wz`: The angular velocity, in radian per second (rad/s), around the smartphone's X, Y, and Z axes, respectively.
8–10. `Bx, By, Bz`: The magnetic field strength, in microtesla (μT), along the smartphone's X, Y, and Z axes, respectively.
11–13. `gFx, gFy, gFz`: The g-force measured as the ratio of normal force to gravitational force (FN/Fg), along the smartphone's X, Y, and Z axes, respectively.
14–16. `Yaw, Pitch, Roll`: The angle of rotation, in degrees (°), around the smartphone's X, Y, and Z axes, respectively.
17. `Pressure`: The atmospheric pressure in hectopascal (hPa).
18. `Illuminance`: The illuminance in lux (lx).

V. `<site>_Local.csv` contains the local coordinates of RPs belonging to a site. Each site has its own frame of reference and the origins are at RPs 10, 122, 60, and 99 for Sites 1, 2, 3, and 4, respectively.

1. `RP_ID`: The Reference Point IDentifier.
2–4. `X, Y, Z`: The X, Y, and Z coordinates of the RP in centimeters (cm).

VI. `<site>_NAD83.csv` contains the global coordinates of RPs belonging to a site with respect to the NAD83(2011)/Colorado Central coordinate system.

1. `RP_ID`: The Reference Point IDentifier.
2. `X, Y`: The X and Y coordinates of the RP in meters (m).

VII. `<site>_GPS_<phone>_App3.csv` contains the GPS coordinates of RPs belonging to a site as computed by the smartphone's GPS chipset and reported by App 3.

1. `RP_ID`: The Reference Point IDentifier.
2. `Date_Time`: The date and time the sample was captured. The date and time formats are as described above.
3,4. `Latitude, Longitude`: The latitude and longitude coordinates of the RP.

	Phone 1			Phone 2		
	$\{day_1, day_2\}$	$\{day_2, day_3\}$	$\{day_1, day_3\}$	$\{day_1, day_2\}$	$\{day_2, day_3\}$	$\{day_1, day_3\}$
WiFi						
Spearman's ρ	0.960	0.949	0.946	0.952	0.968	0.936
Kendall's τ	0.837	0.826	0.815	0.828	0.877	0.796
Bluetooth						
Spearman's ρ	0.575	0.736	0.700	0.716	0.889	0.790
Kendall's τ	0.454	0.609	0.578	0.584	0.786	0.683
Cellular						
Spearman's ρ	0.964	0.964	1.0	0.964	0.964	1.0
Kendall's τ	0.904	0.904	1.0	0.904	0.904	1.0
Sensors						
Spearman's ρ	0.928	0.970	0.933	0.960	0.990	0.943
Kendall's τ	0.823	0.911	0.852	0.897	0.955	0.852

Table 2. Results of the correlation analysis between the measurements obtained on three different days for Phone 1 and Phone 2. Spearman's ρ varies between -1 and $+1$ with 0 implying no correlation, while values of -1 or $+1$ imply an exact monotonic relationship. Kendall's τ varies between -1 and $+1$. Values close to $+1$ indicate strong agreement, while values close to -1 indicate strong disagreement. For WiFi, the results were generated using averaged RSS readings of fifty randomly selected APs that were observed over the three days. For Bluetooth, the results were generated using averaged RSS readings of fifteen randomly selected devices that were observed over the three days. The relatively lower correlation results obtained for Bluetooth is attributed to the fact that Bluetooth signals are more vulnerable to channel gain and fast fading than WiFi signals, causing measurements to fluctuate severely over time[13]. For Cellular, the results were generated using averaged readings of UMTS neighbors, LTE neighbors, RSRP strongest, frequency, EARFCN, RSRP, and RSRQ from a cellular base station that a phone connected to over the three days. For Sensors, the results were generated using the averaged readings of linear acceleration, angular velocity, magnetic field strength, g-force, angle of rotation, atmospheric pressure, and illuminance. The p-value of all results ranged between 0.0 and 0.02.

	day_1	day_2	day_3
WiFi			
Spearman's ρ	0.920	0.925	0.893
Kendall's τ	0.773	0.796	0.728
Bluetooth			
Spearman's ρ	0.763	0.706	0.843
Kendall's τ	0.657	0.535	0.703
Cellular			
Spearman's ρ	1.0	1.0	1.0
Kendall's τ	1.0	1.0	1.0
Sensors			
Spearman's ρ	0.725	0.774	0.752
Kendall's τ	0.617	0.720	0.676

Table 3. Results of the correlation analysis between the measurements obtained from Phone 1 and Phone 2 for three different days. Spearman's ρ varies between -1 and $+1$ with 0 implying no correlation, while values of -1 or $+1$ imply an exact monotonic relationship. Kendall's τ varies between -1 and $+1$. Values close to $+1$ indicate strong agreement, while values close to -1 indicate strong disagreement. For WiFi, the results were generated using the averaged RSS readings of fifty randomly selected APs that were observed by both phones for a given day. For Bluetooth, the results were generated using the averaged RSS readings of fifteen randomly selected devices that were observed by both phones for a given day. For Cellular, the results were generated using averaged readings of UMTS neighbors, LTE neighbors, RSRP strongest, frequency, EARFCN, RSRP, and RSRQ of a cellular base station that both phones connected to for a given day. For Sensors, the results were generated using the averaged readings of linear acceleration, angular velocity, magnetic field strength, g-force, angle of rotation, atmospheric pressure, and illuminance for a given day. The p-value of all results ranged between 0.0 and 0.01.

VIII. `<site>_GPS_<phone>_App4.csv` contains the GPS coordinates of RPs belonging to a site as computed by the smartphone's GPS chipset and reported by App 4.

1. `RP_ID`: The Reference Point IDentifier.
2. `Time`: The time the sample was captured. The time format is as described above.
3,4. `Latitude, Longitude`: The latitude and longitude coordinates of the RP.
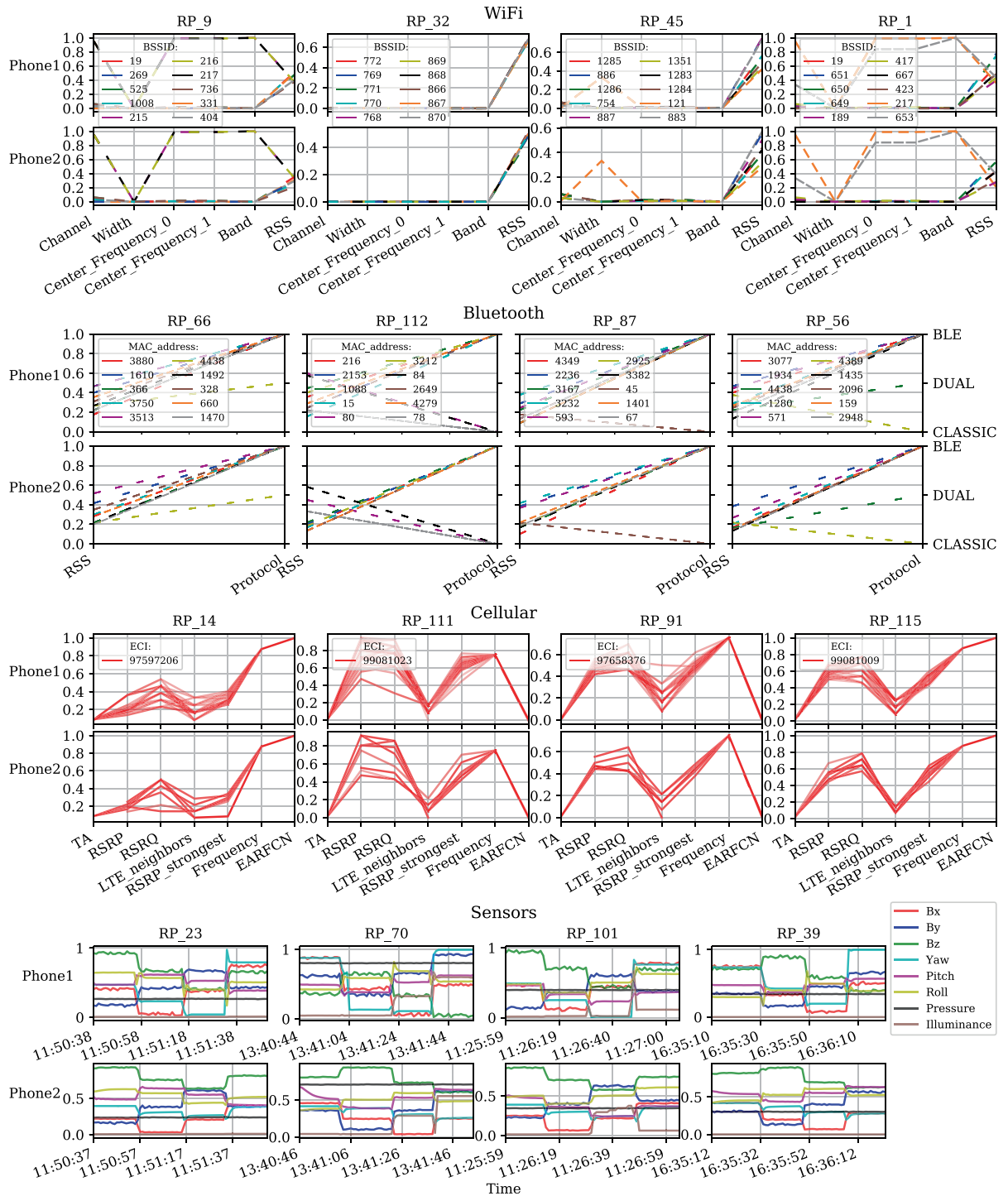
Fig. 6 Visualization of the data collected by Phone 1 and Phone 2 over randomly selected RPs. WiFi, Bluetooth, and cellular data are represented using parallel coordinate plots of the most important features, while sensor data are represented using time plots of magnetic field strength, angle of rotation, atmospheric pressure, and illuminance. All features are normalized between 0 and 1.

IX. `<phone>_<date>.csv` contains sensors data collected by a smartphone via App 3 before the smartphone is mounted to the tripod. Field labels are identical to that described in IV (`<phone>_Sensors_<RP>.csv`).

Technical Validation

The technical quality of the OutFin dataset was evaluated using experiments that consider two basic requirements that any high-quality dataset should satisfy, i.e., reliability and validity. Additionally, as a demonstration of the dataset's potential for positioning applications, a number of practical usage examples are presented.

	Phone 1				Phone 2				Reference values
	Min	Max	Mean	SD	Min	Max	Mean	SD	
WiFi									
Detected SSIDs	12	51	26.09	8.95	9	40	21.29	6.80	—
Detected BSSIDs	98	223	159.32	31.68	67	168	114.97	23.92	—
RSS (dBm)	−97	−53.33	−85.82	6.86	−99	−38	−84.20	6.88	≈[−102, −34][58]
Bluetooth									
Detected MAC addresses	5	205	59.50	47.46	4	168	45.45	35.99	—
RSS (dBm)	−98	−53	−86.28	4.69	−113	−65	−99.40	5.35	≈[−110, −48][55]
Cellular									
Detected ECIs	1	5	1.45	0.91	1	4	1.35	0.73	—
LTE neighbors	0	12	2.36	1.53	0	14	2.45	1.79	—
RSRP strongest (dBm)	−128	−81	−103.32	6.90	−127	−82	−105.18	8.26	—
RSRP (dBm)	−118	−82	−99.86	6.28	−118	−82	−100.89	6.98	≈[−120, −70][59]
RSRQ (dB)	−20	−7	−12.83	2.33	−20	−6	−12.87	2.48	≈[−24, −5][59]
Sensors									
Magnitude of magnetic field (μT)	38.52	51.07	44.49	3.51	29.45	73.03	51.90	13.40	≈51[60]
Atmospheric pressure (hPa)	833.14	845.02	837.93	3.13	831.67	843.52	836.37	3.12	≈[829.66, 843.21, 836][61]
Illuminance (μlx)	1e-6	0.1508	0.0138	0.0271	2e-7	0.1243	0.0104	0.0207	≈[0.1, 0.01e−6][62]

Table 4. Descriptive statistics of the OutFin dataset. These include the minimum, maximum, mean, and standard deviation of the most important variables. Reference values are provided where applicable. Small variations in results between the phones are mainly attributed to *device heterogeneity*[63] (e.g., the sensitivity of the radio receiver or sensor). The reference value for the magnitude of the magnetic field represents the Earth's magnetic field around Denver, Colorado. The reference values for atmospheric pressure represent, respectively, the minimum, maximum, and mean recorded atmospheric pressure in Denver, Colorado, during the data collection period. The reference values for illuminance represent the light intensity for sunlight, daylight, and twilight, respectively. An hour-by-hour description of other weather conditions, such as temperature, humidity, and visibility at the time of data collection can be retrieved from[64].
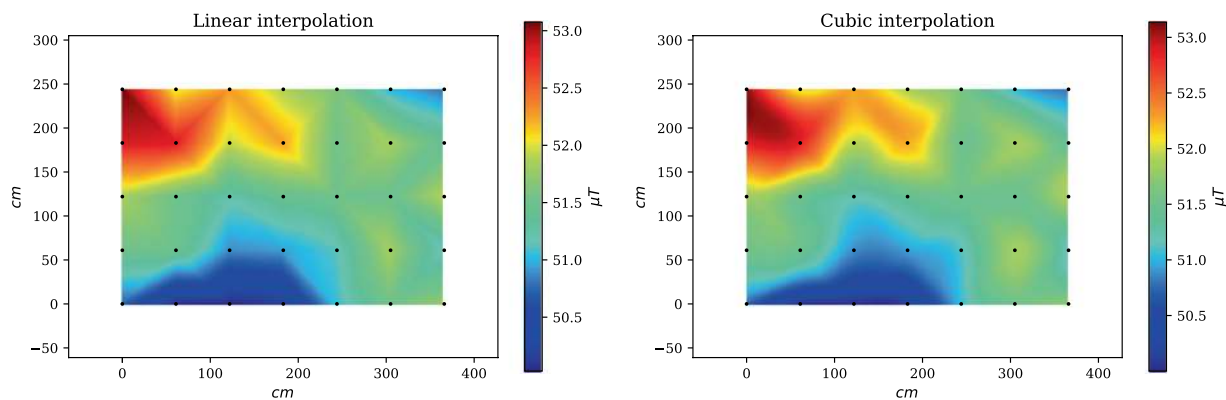

Fig. 7 Interpolated magnetic field magnitude of Site 3 using linear interpolation (left) and cubic interpolation (right). The maps were generated using calibrated magnetic field measurements from Phone 1 and Phone 2.

Measurement reliability. A data acquisition platform is said to be reliable if it provides consistent measurements at different points in time. To this end, before the collection campaign, WiFi, Bluetooth, cellular, and sensor data was captured over three different days at the same location. Spearman's and Kendall's correlation coefficients were then used to quantify the degree of consistency between temporal measurements for a given phone. Table 2 shows Spearman's and Kendall's correlation coefficients for the two smartphones for all possible pairs of days. Given that correlation results are high (i.e., close to the maximum value of 1.0), it can be concluded that the dataset possesses a high degree of reliability.

Measurement validity. A data acquisition platform is said to be valid if it accurately measures what it is intended to measure. In some cases, this requires the presence of theoretically-derived data to compare experimental data against. For example, WiFi RSS values can be computed using a path loss model. An input to the model is the distance between the transmitter and receiver. However, obtaining such inputs is not feasible since the exact location of all APs in the environment needs to be known. In the absence of theoretically-derived data, validity can be assessed by comparing data generated by different sources and checking for consistency. Accordingly, for a given
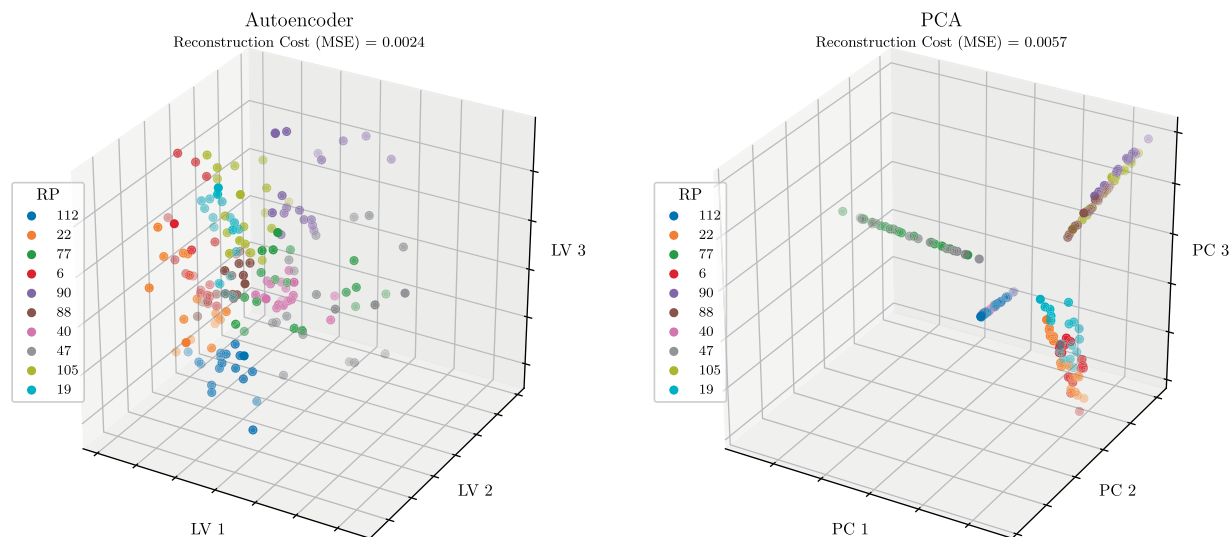
Fig. 8 The 3D codes for 18 WiFi RSS measurements (9 measurements per phone) for 10 randomly selected RPs produced by the autoencoder (left) and PCA (right). MSE: mean squared error; PC: principal component; LV: latent variable.

day, Spearman's and Kendall's correlation coefficients were used to quantify the degree of consistency between the measurements obtained by the phones. The correlation results for the foregoing three days are shown in Table 3. These results demonstrate high levels of consistency, which attests to the validity of the dataset.

As graphical evidence of measurement validity, Fig. 6 compares some of the data generated by the smartphones at randomly selected RPs side-by-side. Plots of the same data type exhibit the same profile despite corresponding to two different smartphones. Table 4 reports descriptive statistics of the data collected by each phone with respect to various variables. These statistics are compared against previously reported reference values, where applicable. The statistics displayed in Table 4 further support the validity of the dataset by ruling out the possibility that the dataset contains unrealistic, erratic, or random data.

Usage Examples

This subsection provides a brief demonstration of some of the application domains that OutFin can be used for. These include *fingerprint interpolation*, *feature extraction*, *performance evaluation*, and *signal denoising*.

Fingerprint interpolation.   Building a fingerprint map is usually required to provide positioning in a continuous fashion. The resolution of a map depends highly on the RP granularity (the higher the RP granularity, the better the map resolution). However, collecting fingerprints at highly granular RPs is time-consuming and labor intensive. Thus, interpolation methods are often employed to calculate the fingerprints between the locations of known fingerprints[36]. The choice of an interpolation technique is pivotal to the resulting map. For example, Fig. 7 compares the magnetic field maps created for Site 3 by two different interpolation techniques, namely linear and cubic interpolation. Clearly, the resulting maps are not identical, which suggests that a positioning algorithm would exhibit a difference in performance depending on the employed map.

Feature extraction.   A WiFi fingerprint has entries for all APs detected in an entire environment, but only a subset of these APs is observed at different locations. This is especially true for large-scale environments. For example, OutFin contains measurements from 1,379 unique APs; however, on average, only 10 of these APs are observed at any given RP. Consequently, feature extraction techniques are often utilized to reduce the dimensionality of the fingerprint space in order to achieve efficient and robust positioning[37]. Figure 8 compares two dimensionality reduction methods, i.e., the autoencoder and principal component analysis (PCA). The reconstruction cost obtained by the autoencoder is lower than that obtained by PCA. This suggests that the autoencoder is better at compressing the fingerprint space into a lower dimensional representation that comprises the informative content of the fingerprint space.

Performance evaluation.   When proposing a new positioning method, the performance of the proposed method is often evaluated against the performance of previously proposed methods. It is often the case that at the heart of many of the methods benchmarked against is a machine learning algorithm, such as k-Nearest Neighbors (k-NN), Support Vector Machine (SVM), Decision Tree, or Naive Bayes[38]. Therefore, with the purpose of comparing the performance of such algorithms, the positioning problem was casted as a classification task where each RP is treated as a unique class. Various performance metrics were considered, including classification metrics, positioning error, and computational complexity. For the sake of fair comparison, the parameters of each algorithm were fine-tuned using grid search and cross-validation. Evaluation results, shown in Table 5, are reported on the Bluetooth measurements collected from Site 4. The results demonstrate that different algorithms can be

	Classification Metric				Positioning Error (cm)				Computational Complexity[65]	
	Accuracy	Precision	Recall	F1	Min	Max	Mean	SD	Training	Prediction
Algorithm										
k-NN	0.948	0.964	0.948	0.945	0.0	366.0	11.46	51.52	—	$\mathcal{O}(np)$
RBF kernel SVM	0.962	0.970	0.962	0.961	0.0	1098.0	18.81	121.46	$\mathcal{O}(n^2p + n^3)$	$\mathcal{O}(n_{sv}p)$
Decision Tree	0.957	0.967	0.957	0.956	0.0	732.0	15.19	83.19	$\mathcal{O}(n^2p)$	$\mathcal{O}(p)$
Naive Bayes	0.910	0.956	0.910	0.911	0.0	549.0	23.82	82.38	$\mathcal{O}(np)$	$\mathcal{O}(p)$

Table 5. Performance evaluation of commonly used algorithms for positioning with respect to various metrics. The results were generated using 530 Bluetooth samples (60 training and 40 testing) collected by both phones from Site 4. RBF: radial basis function; n: number of training samples; p: number of features; n_{sv}: number of support vectors.
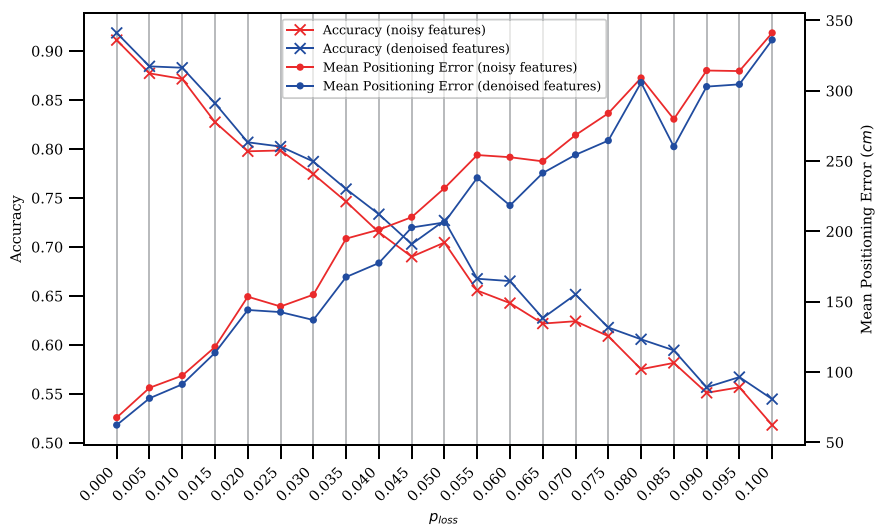
Fig. 9 Noisy vs. denoised features for positioning. For a given p_{loss} value, the results were generated using 3,111 cellular samples collected by both phones from Site 2. A k-NN algorithm is used for comparison where ~60 of the samples were used for training and the remaining ~40 for testing.

ranked differently depending on the chosen performance metric. For example, the best classification accuracy was achieved by RBF SVM, while the lowest mean positioning error was achieved by k-NN.

Signal denoising. Signal loss can negatively impact the performance of a positioning system. Thus, denoising techniques are often integrated as a preprocessing step to enhance positioning[39]. As an example, a denoising autoencoder was utilized as a denoising agent where the feature vector of a cellular fingerprint is corrupted to emulate randomized loss of data. The degree of corruption is controlled by a predefined probability (p_{loss}) where, for example, a p_{loss} of 0.03 indicates a 3 chance of setting a feature to zero. Figure 9 demonstrates the differences in performance between using noisy cellular features and their denoised versions for positioning in Site 2. On average, the use of the denoising step resulted in a 1.43 improvement in accuracy and a 13.25 reduction in positioning error.

Code availability

Well-documented scripts, written in `Python 3.6.4`[40], are present alongside the dataset (also available on GitHub[41]). These include the scripts used to generate the results described in the Technical Validation section as well as a script to calibrate magnetic field measurements against hard/soft-iron distortions. The data required to replicate the experiments reside in `OutFin/Code/temporal_data`. Depending on the script, some of the following libraries may be required: `os`, `pandas`, `scipy`, `random`, `sklearn`, `matplotlib`, `numpy`, `statistics`, `keras`, `math`. Additionally, a thorough description of the collection environment in the form of an interactive map (developed using `QGIS 3.10`[27]) is provided. The map is composed of several layers that display information such as RP coordinates (both ground truth and smartphone estimated), pictures of the collection sites, and building height and ground elevation (as provided by the City and County of Denver[42]). High-resolution aerial imagery (3-inch), provided by the Denver Regional Council of Governments[43], are used as the basemap.

References

1. Location-based services (lbs) market statistics and forecast-2026. *Allied Market Research* https://www.alliedmarketresearch.com/location-based-services-market (2019)
2. Hopkins, J. & Turner, J. Go mobile: location-based marketing, apps, mobile optimized ad campaigns, 2D codes and other mobile strategies to grow your business. (John Wiley & Sons, 2012).
3. Hammad, A. & Faith, P. Location based authentication, US Patent 9,721,250. (2017).
4. Leorke, D. Location-based gaming apps and the commercialization of locative media. *Mobility and Locative Media: Mobile Communication in Hybrid Spaces*, **132** (2014).
5. Zheng, Y. Location-based social networks: Users. In *Computing with spatial trajectories* 243–276. (Springer, 2011).
6. Huang, H., Gartner, G., Krisp, J. M., Raubal, M. & Weghe, V. D. Location based services: ongoing evolution and research agenda. *Journal of Location Based Services* **12**(2), 63–93 (2018).
7. Ranacher, P., Brunauer, R., Trutschnig, W., Spek, S. V. D. & Reich, S. Why gps makes distances bigger than they are. *International Journal of Geographical Information Science* **30**(2), 316–333 (2016).
8. Carroll, A. *et al.* An analysis of power consumption in a smartphone. In *USENIX annual technical conference*, **14** (2010)
9. Paek, J., Kim, K.-H., Singh, J. P. & Govindan R. Energy-efficient positioning for smartphones using cell-id sequence matching. (Association for Computing Machinery, Bethesda, Maryland, USA, 2011)
10. Paul, A. Z. Accuracy of iphone locations: A comparison of assisted gps, wifi and cellular positioning. *Transactions in GIS* **13**, 5–25 (2009).
11. Vo, Q. & De, P. A survey of _fingerprint-based outdoor localization. *IEEE Communications Surveys & Tutorials* **18**(1), 491–506 (2015).
12. Bahl, P. & Padmanabhan, V. N. Radar: An in-building rfbased user location and tracking system. In *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, **2**, 775-784 (2000).
13. Faragher, R. & Harle, R. Location _ngerprinting with bluetooth low energy beacons. *IEEE Journal on Selected Areas in Communications* **33**(11), 2418–2428 (2015).
14. Rizk, H., Torki, M. & Youssef, M. Cellindeep: Robust and accurate cellular-based indoor localization via deep learning. *IEEE Sensors Journal* **19**(6), 2305–2312 (2019).
15. Al-homayani, F. & Mahoor, M. Improved indoor geomagnetic field fingerprinting for smartwatch localization using deep learning. In *2018 Inter-national Conference on Indoor Positioning and Indoor Navigation (IPIN)*, (2018).
16. Luo, J. & Gao, H. Deep belief networks for fingerprinting indoor localization using ultrawideband technology. *International Journal of Distributed Sensor Networks* **12**(1), 5840916 (2016).
17. Wang, B., Chen, Q., Yang, L. T. & Chao. H.-C. Indoor smartphone localization via fingerprint crowdsourcing: Challenges and approaches. *IEEE Wireless Communications* **23**(3), 82–89 (2016).
18. Vepakomma, P. Tonde, C., Elgammal, A. Supervised dimensionality reduction via distance correlation maximization. *Electron. J. Statist.* **12**(1), 960–984 (2018).
19. Kandasamy, K., Neiswanger, W., Schneider, J., Poczos, B. & Xing, E. P. Neural architecture search with bayesian optimisation and optimal transport. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K. Cesa-Bianchi, N. & Garnett, R. editors. *Advances in Neural Information Processing Systems 31*. (2018).
20. Kudo, T. & Miura, J. Utilizing wifi signals for improving slam and person localization. In *2017 IEEE/SICE International Symposium on System Integration (SII)*, (2017).
21. Viel, A. *et al.* Map matching with sparse cellular fingerprint observations. In *2018 Ubiquitous Positioning, Indoor Navigation and Location-Based Services (UPINLBS)*, (2018).
22. Mobile operating system market share worldwide https://gs.statcounter.com/os-market-share/mobile/worldwide. *StatCounter* (2019).
23. Webprovider. Wifi analyzer pro. *Google Play* https://play.google.com/store/apps/details?id=info.wifianalyzer.pro&hl=en_US (2019).
24. Zoltán Pallagi. Bluetooth scanner. *Google Play* https://play.google.com/store/apps/details?id=com.pzolee.bluetoothscanner&hl=en_US.
25. Vitaly V. Netmonitor pro. *Google Play* https://play.google.com/store/apps/details?id=ru.v_a_v.netmonitorpro&hl=en_US.
26. Vieyra Software. Physics toolbox sensor suite pro. *Google Play* https://play.google.com/store/apps/details?id=net.vieyrasoftware.physicstoolboxsuitepro&hl=en_US.
27. A Free and Open Source Geographic Information System *Qgis 3.10 a coruña* https://qgis.org/en/site/.
28. Wi-fi scanning overview. *Android Developers* https://developer.android.com/guide/topics/connectivity/wifi-scan#wifi-scan-throttling (2020).
29. Bluetoothadapter. *Android Developers* https://developer.android.com/reference/android/bluetooth/BluetoothAdapter#ACTION_DISCOVERY_STARTED (2020).
30. Android.telephony. *Android Developers* https://developer.android.com/reference/android/telephony/package-summary (2020).
31. Sensors overview. *Android Developers* https://developer.android.com/guide/topics/sensors/sensors_overview (2020).
32. Angermann, M., Frassl, M., Doniec, M., Julian, B. J. & Robertson, P. Characterization of the indoor magnetic field for applications in localization and mapping. In *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, (2012).
33. Alhomayani, F. & Mahoor, M. H. OutFin, a multi-device and multi-modal dataset for outdoor localization based on the fingerprinting. *figshare* approach. https://doi.org/10.6084/m9.figshare.12069993 (2020).
34. Assigned numbers for baseband. *Bluetooth Special Interest Group.* https://www.bluetooth.com/specifications/assigned-numbers/baseband/.
35. Specification #: 36.213. *The 3rd Generation Partnership Project.* https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2427.
36. Solin, A., Kok, M., Wahlström, N., Schön, T. B. & Särkkä, S. Modeling and interpolation of the ambient magnetic _eld by gaussian processes. *IEEE Transactions on Robotics* **34**(4), 1112–1127 (2018).
37. Nowicki, M. & Wietrzykowski, J. Low-effort place recognition with wifi fingerprints using deep learning. *Automation 2017* (Springer International Publishing.Cham, 2017).
38. Alhomayani, F. & Mahoor, M. H. Deep learning methods for fingerprint-based indoor positioning: a review. *Journal of Location Based Services* **14**(3), 129–200 (2020).
39. Alhomayani, F. & Mahoor, M. Deep learning-based symbolic indoor positioning using the serving enodeb, arXiv: (2009).
40. Python 3.6.4. https://www.python.org/downloads/release/python-364/ (2017).
41. Outfin. *GitHub* https://github.com/alhomayani/OutFin (2020).
42. City of denver open data catalog. licensed under the Creative Commons Attribution 3.0 (http://creativecommons.org/licenses/by/3.0/).
43. Drcog regional data catalog. licensed under the Creative Commons Attribution 3.0 (http://creativecommons.org/licenses/by/3.0/).
44. Understanding reference frames and device attitude. *Apple developers* https://developer.apple.com/documentation/coremotion/getting_processed_device-motion_data/understanding_reference_frames_and_device_attitude#2875084.
45. Gubiani, G., Gallo, P., Viel, A., Torre, A. D. & Montanari, A. A cellular network database for fingerprint positioning systems. In Welzer, T. *et al.* editors, *New Trends in Databases and Information Systems*, (Springer International Publishing, Cham, 2019).
46. Torres-Sospedra, J., Rambla, D., Montoliu, R., Belmonte, O. & Huerta, J. Ujiindoorloc-mag: A new database for magnetic _eld-based localization problems. In *2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, (2015).
47. Barsocchi, P., Crivello, A., Rosa, D. L. & Palumbo, F. A multisource and multivariate dataset for indoor localization methods based on wlan and geo-magnetic field fingerprinting. In *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)* (2016).

48. Tóth, Z. & Tamás, J. Miskolc iis hybrid ips: Dataset for hybrid indoor positioning. In *2016 26th International Conference Radioelektronika (RADIOELEKTRONIKA)*, (2016).

49. Moayeri, N., Ergin, M. O., Lemic, F., Handziski, V. & Wolisz, A. Perfloc (part 1): An extensive data repository for development of smartphone indoor localization apps. In *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)* (2016).

50. Popleteev, A. Ambiloc: A year-long dataset of fm, tv and gsm fingerprints for ambient indoor localization. In *8th International Conference on Indoor Positioning and Indoor Navigation (IPIN)* (2017).

51. Hanley, D., Faustino, A. B., Zelman, S. D., Degenhardt, D. A. & Bretl, T. Magpie: A dataset for indoor positioning with magnetic anomalies. In *2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN)* (2017).

52. Mendoza-Silva, G., Richter, P. Torres-Sospedra, J., Lohan, E. & Huerta. J. Long-term wifi fingerprinting dataset for research on robust indoor positioning. *Data* **3**(1), 3 (2018).

53. Byrne, D., Kozlowski, M., Santos-Rodriguez, R., Piechocki, R. & Craddock, I. Residential wearable rssi and accelerometer measurements with detailed location annotations. *Sci. Data* **5**, 180168 (2018).

54. Mohammadi, M., Al-Fuqaha, A., Guizani, M. & Oh, J. Semisupervised deep reinforcement learning in support of iot and smart city services. *IEEE Internet of Things Journal* **5**(2), 624–635 (2018).

55. Baronti, P. Barsocchi, P., Chessa, S., Mavilia, F. & Palumbo, F. Indoor bluetooth low energy dataset for localization, tracking, occupancy, and social interaction. *Sensors* **18**(12), 4462 (2018).

56. Aernouts, M., Berkvens, R., Vlaenderen, K. V. & Weyn, M. Sigfox and lorawan datasets for figerprint localization in large urban and rural areas. *Data* **3**(2), 13 (2018).

57. Mendoza-Silva, G. M., Matey-Sanz, M., Torres-Sospedra, J. & Huerta, J. Blerss measurements dataset for research on accurate indoor positioning. *Data* **4**(1), 12 (2019).

58. Torres-Sospedra, J. *et al*. Ujiindoorloc: A new multibuilding and multi-floor database for wlan fingerprint-based indoor localization problems. In *2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN)* (2014).

59. Sevindik, V., Wang, J., Bayat, O. & Weitzen, J. Performance evaluation of a real long term evolution (lte) network. In *37th Annual IEEE Conference on Local Computer Networks - Workshops*, (2012).

60. Magnetic field calculators. *NOAA National Centers for Environmental Information.* https://www.ngdc.noaa.gov/geomag/calculators/magcalc.shtml?useFullSite=true.

61. Denver, co weather history. *Weather Underground.* https://www.wunderground.com/history/monthly/us/co/denver/KDEN/date/2019-11 (2019).

62. Recommended light levels (illuminance) for outdoor and indoor venues. *National Optical Astronomy Observatory* https://www.noao.edu/education/QLTkit/ACTIVITY_Documents/Safety/LightLevels_outdoor+indoor.pdf.

63. Chen, L., Wu, E. H., Jin, M. & Chen, G. Homogeneous features utilization to address the device heterogeneity problem in fingerprint localization. *IEEE Sensors Journal* **14**(4), 998–1005 (2014).

64. Past weather in Denver, Colorado, USA. *Time and Date AS.* https://www.timeanddate.com/weather/usa/denver/historic (2020).

65. Computational complexity of machine learning algorithms. https://www.thekerneltrip.com/machine/learning/computational-complexity-learning-algorithms/ (2018).

Acknowledgements

Author contributions

All authors equally contributed to all aspects of the presented work.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.H.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.