Designing Datasets for Smartphone Sensing Based Localization

This thesis is submitted to fulfil the requirement of the degree

Master in Computer Science & Engineering

Submitted by

Sourav Pramanik

Examination Roll no. – M4CSE22037

Registration no. – 154161 of 2020-2021

Year – **2**nd, Semester – **4**th

Under the guidance of

Dr. Chandreyee Chowdhury

Department of Computer Science & Engineering

Jadavpur University

Kolkata – 700032

August 2022

Faculty of Engineering and Technology
Jadavpur University, Kolkata - 700032

Certificate

This is to certify that the thesis entitled "Designing Datasets for Smartphone Sensing Based Localization", submitted by Mr. Sourav Pramanik (Examination Roll No. M4CSE22037), under our supervision and guidance during the session of 2020-22 in the department of Computer Science & Engineering, Jadavpur University. We are satisfied with his work, which is being presented for the partial fulfilment of the degree of Master in Computer Science & Engineering from Jadavpur University, Kolkata-700032.

……………………………………..

Dr. Chandreyee Chowdhury
Associate Professor
Department of
Computer Science & Engineering
Jadavpur University
Kolkata, 700032

…………………………………………

Prof. Chandan Mazumdar
Dean of Faculty Council of
Engineering and Technology
Jadavpur University
Kolkata, 700032

…………………………………………

Prof. Nandini Mukhopadhyay
Head of the Department of
Computer Science & Engineering
Jadavpur University
Kolkata, 700032

Faculty of Engineering and Technology
Jadavpur University, Kolkata - 700032

Certificate of Approval

The forgoing thesis entitled "Designing Datasets for Smartphone Sensing Based Localization" is hereby approved as a creditable study of an Engineering subject carried out and presented in a manner that fulfils its acceptance as a prerequisite to the degree for which it is submitted. It is understood that by this approval, the undersigned does not necessarily endorse or approve any statement made, opinion expressed, or conclusion drawn therein but approves the thesis only for the purpose for which it is submitted.

The final examination for evaluation of the thesis

Signature of Examiners

……………………………

……………………………

……………………………

Declaration of Originality

I hereby declare that this thesis contains a literature survey and original research work done by me. All the information in this document has been obtained and presented according to academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name: SOURAV PRAMANIK

Examination Roll No. : M4CSE22037

Thesis Title: Designing Datasets for Smartphone Sensing Based Localization

Signature with Date:

ACKNOWLEDGEMENTS

I express my sincere gratitude to my supervisor**, Dr. Chandreyee Chowdhury** for her encouragement, suggestion and advices, without which it would not have been possible to complete my thesis successfully. It was a great honour for me to pursue my research under her supervision.

I would like to thank research scholar Manjarini Mallik for being a constant source of encouragement, inspiration and for her valuable suggestions coupled with her technical expertise throughout my research work.

I would also like to thank my co-workers, all the staffs of DST-FIST laboratory, and the other research scholars of our department for providing constant encouragement throughout my thesis work.

Last but not the least I extend my words of gratitude to my parents for personally motivating me to carry out the work smoothly.

ABSTRACT

Since the increasing use of wireless networks, there has been a surge in interest in making use of them for a number of reasons. One of them is localization of mobile devices in both indoor and outdoor settings. Wi-Fi-based localization has more potential than GPS does because of the many factors that reduce the accuracy of GPS positioning. These factors include using GPS in underground environments or in large cities with multi-story buildings, where GPS signals can be blocked or reflected.

A lot of data support is needed for Wi-Fi fingerprint-based localisation. The publicly available datasets lack the variety of data collected under different ambient conditions essential to generalize the performance of a localization system. The RSSI of Wi-Fi signals can change depending on the environment. So, The Wi-Fi data was collected and designed under different conditions at two different locations one of which is a ground-level metro station and another one is underground metro. In addition, an inertial sensor dataset based on smartphone inertial sensors (accelerometer, gyroscope, and gravity sensor) was designed to study the movement of smartphones in order to track the users. The details information of metro stations have described, where data was collected. Information on configuration of smartphones have provided which are used for data collection. Some Graphical plot has been shown here in order to assist the process of localization. The designed datasets of JU-MetroLoc have shown a maximum of 78.73% classification accuracy for Random Forest classifier.

List of Acronyms

Short Form	Full-Form
GPS	Global Positioning System
Wi-Fi	Wireless Fidelity
LAN	Local area network
RSS	Received Signal Strength
RSSI	Received Signal Strength Indicator
dBm	decibel-milliwatts
AP	Access Point
URUS	Ubiquitous Networking Robotics in Urban Settings
SLAM	Simultaneous Localization And Mapping
C-SLAM	Credibilist Simultaneous Localization And Mapping
GPS	Global positioning system
IMU	Inertial Measurement Unit
GSM	Global System for Mobile communication
RFID	Radio frequency identification
UHF	Ultra High Frequency
UWB	Ultra-Wideband
RF	Radio Frequency
ToA	Time of Arrival
TDoA	Time Difference of Arrival
WPANs	Wireless Personal Area Networks
WLAN	Wireless Local Area Network
MAC	Media Access Control Address
WSN	Wireless Sensor Network
GNSS	Global Navigation Satellite System
IoT	Internet of Things
FFT	Fast Fourier Transform
DFT	Discrete Fourier Transform
PCA	Principal Component Analysis
BSSID	Basic Service Set Identification

List of Figures

List of Tables

Chapter 1

Introduction

In the last couple of years, smartphones and other wireless devices have become more and more popular, which has led to a wide range of services, such as localization. Locating a device or user in an interior environment is known as indoor localization. It has been less than a decade since the widespread use of smartphones with wireless communication capabilities made it possible to find and track both the users and the devices. This made it possible for a wide range of related applications and services to be created [1]. Smartphone manufacturers have also included a vast combination of sensors in their products, which makes it possible for such applications to be developed. Wireless sensors are rapidly becoming one of the most essential types of sensors for localization. The localization of users and devices has several practical applications in a variety of fields, including the medical field [2], activity detection [3], smart architectures (such as smart cities) [4], and smart buildings [5], amongst many others. This success is made possible by several unique qualities of wireless sensors, such as its resistance to lighting conditions and occlusion, which overcomes the limitations of cameras, and its nonintrusive sensing, which does not need any further effort on the part of the user.

The purpose of indoor localization is to determine the precise position of a smart device inside the confines of a multi-story building or other indoor structure. The GPS system might not function correctly in an interior setting since the signal strength varies erratically and the user's device might not get position updates at regular intervals [6]. Indoor localization based on Wi-Fi (IEEE 802.11 WLAN standard) signal strength or RSSI has become a significant strategy since it does not require any extra hardware devices to be put in a building. This is because Wireless LAN now covers every location in the majority of buildings, including colleges, hospitals, workplaces, and shopping malls, among other types of establishments.

The popularity of received signal strength (RSS) fingerprint-based Wi-Fi localization methods has increased significantly in recent years because RSS can be easily collected by a Wi-Fi-integrated mobile device without the usage of any additional hardware. The RSS, or received signal strength, is the signal power strength that is measured at the receiver in decibel-milliwatts (dBm) unit [7-9]. The Wi-Fi fingerprint-based localization process can be broken down into two stages: the offline stage, and the online stage. During the offline part of the process, the database is built by monitoring and recording the RSS of Wi-Fi signals coming from a number of different access points (APs). During the online phase, the user's real-time RSS measurements are matched and compared with the fingerprints stored in the database using classification methods to estimate the user's location (shown in Fig. 1.1).

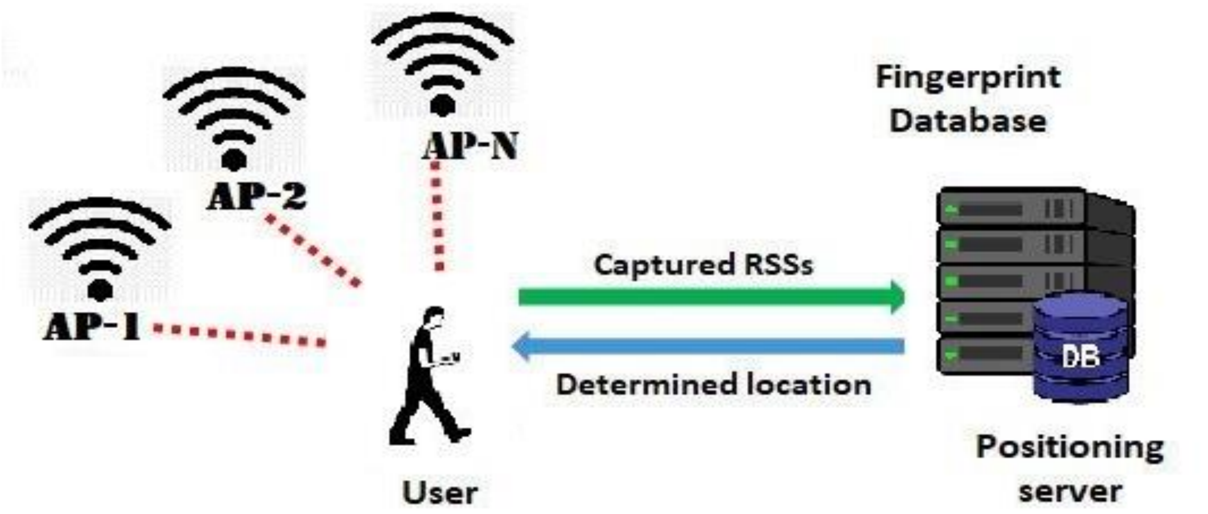


Fig. 1.1: RSSI-based localization

1.1 Applications of Indoor Localization

A. Health Services

Indoor localization has been used in healthcare services for various purposes. It can help employees, patients, and visitors by tracking their positions and suggesting possible ways of navigation. The Current technique entails broadcasting a message or paging a doctor or staff member who may not be near the patient. The staff's tardiness might kill the patient. Broadcasting the message will send unrelated messages to others. A location-based approach would track employees' locations. In an emergency, the localization system would identify a qualified staff member nearby.

Indoor localization can also help doctors keep track of their patients and see where they go to make sure they are safe. A localization system makes it easy for people who want to visit patients to find where they need to go [1].

B. Disaster Management and Recovery

Technology has the potential to make disaster management easier and aid in the recovery process following any type of natural (including tornadoes, earthquakes, storms, and floods, amongst others) or man-made (terrorist attacks, etc.) disasters. The recovery process may be sped up and made more efficient with the help of localization, which can also contribute to more effective disaster management. One of the hardest things about disasters is usually finding out where people are and if they are safe. Localization can help in these situations by giving the exact location of the missing people and, in the worst case, getting them medical help. In the same way, if there is a fire or other disaster in an indoor environment, the rescue team can find out where the users are using the localization system. They can then use this information to do targeted work in the affected area [3].

C. Security

Localization can make the world's security much better in many ways. Threats that could pose a security risk can be found by looking at how users move and interact. In the same way, the military can keep track of its assets and troops on the battlefield or in war zones with a localization system. This will improve the operation and make it more likely to succeed. A strong localization system can also help the soldier on the ground get around in places they don't know. This is a strategic advantage because it lets the soldiers focus on their mission and not worry about where to go next. With the help of localization, the central command can come up with better plans and strategies, which they can then give to the soldiers on the ground [1].

D. Robotics

One of the primary ways that indoor localization is used is in robotics. Applications for mobile multi-robot systems have been the subject of a lot of research and development. A very important topic is how robots move in real large indoor settings where they have to work together. For example, when robot teams work together, the results of tasks like surveillance, exploring unknown areas, guiding, and maintaining connectivity are better. The Ubiquitous Networking Robotics in Urban Settings (URUS) project is a great example of how localization can be used to help people get to safety in an emergency. In case of a fire, the robots lead people to safe areas along safe paths [14].

In the field of robotics, SLAM (simultaneous localization and mapping) is also a very interesting topic. A robot does tasks autonomously without knowing its surroundings or position. The construction of autonomous robots benefits from the usage of SLAM. Real-time SLAM requires many processors. Then, Credibilist-SLAM, or C-SLAM builds an environment map using multiple robots. In C-SLAM, a global map is created by integrating the individual maps that have been obtained by a number of different robots [16].

1.2 Localization Sensor

Sensors for localization are explored in detail in this subsection. The sensor on an Android device was chosen to collect data for localization because gives you the most options. Android gives users access to a number of unique sensors, although this varies greatly depending on the mobile phone brand and model. There are a variety of sensors available, the most common of which are:

- Motion sensors: Accelerometer, Gyroscope, Magnetometer, Gravity sensor
- Environmental sensors: Wi-Fi, Bluetooth, GPS, and GSM

1.2.1 Motion sensors

Motion sensors are important for localization because they give valuable information about the distance and direction travelled by a user. This information may be used to determine the phone's location. Additionally, it also can determine device orientation, user’s step detection and changes in the directions of users. Nearly all Android motion sensors measure their values with respect to the coordinate system of the device. A typical Inertial Measurement Unit (IMU) such as used in consists of six degrees of freedom unit composed of a three-axial accelerometer and a three-axial gyroscope. This set of motion sensors (accelerometer, gyroscope, gravity) stands out from the rest since they don't require any additional hardware.

1.2.1. A Accelerometer sensor

Three-axial accelerometers are now common on nearly all modern smartphones. For localization purposes, this sensor is an excellent choice because of its wide accessibility and affordability. Accelerometer values are measured with respect to the device (Fig. 1.2) and are measured in standard SI units of m/s^2.

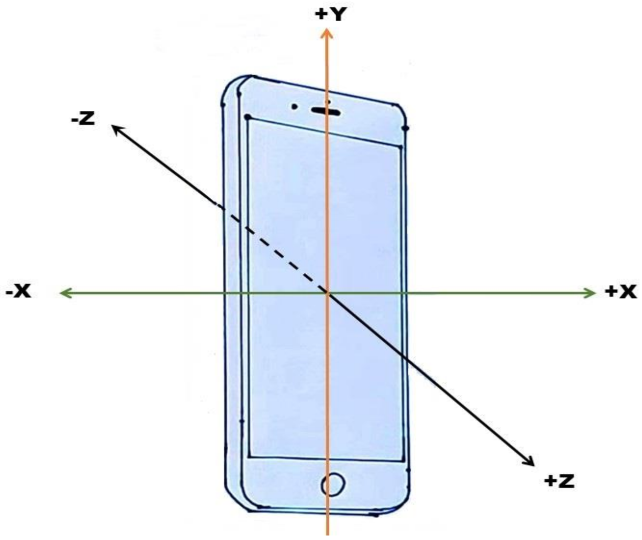


Fig. 1.2: The coordinate system used for a smartphone.

1.2.1. B Gyroscope sensor

A gyroscope measures angular velocity and can help an accelerometer sensor get more accurate motion information. The values of gyroscope sensors are measured in radians per second (rad/s). A three-axial gyroscope, which measures angular velocity along all three axes of a device, is a common feature on phones these days. If there is also a magnetic field sensor, a gyroscope might not be needed. Both a gyroscope and a magnetic field sensor can be used to measure changes in how the phone is turned. But a gyroscope can't take the place of a magnetic field. The gyroscope in a phone works much faster than the magnetic field sensor.

The output that is typically produced by a gyroscope is shown in Fig. 1.3. The data presented here were obtained by successively rotating the phone along the x-axis, the y-axis, and the z-axis of the device, while maintaining the orientation of the phone which was horizontal to the plane. The whole sequence of rotations takes about 8 seconds. If one looks carefully at the graph, one can deduce that the Y-axis is the most essential one. It can be changed if the phone’s orientation is changed.

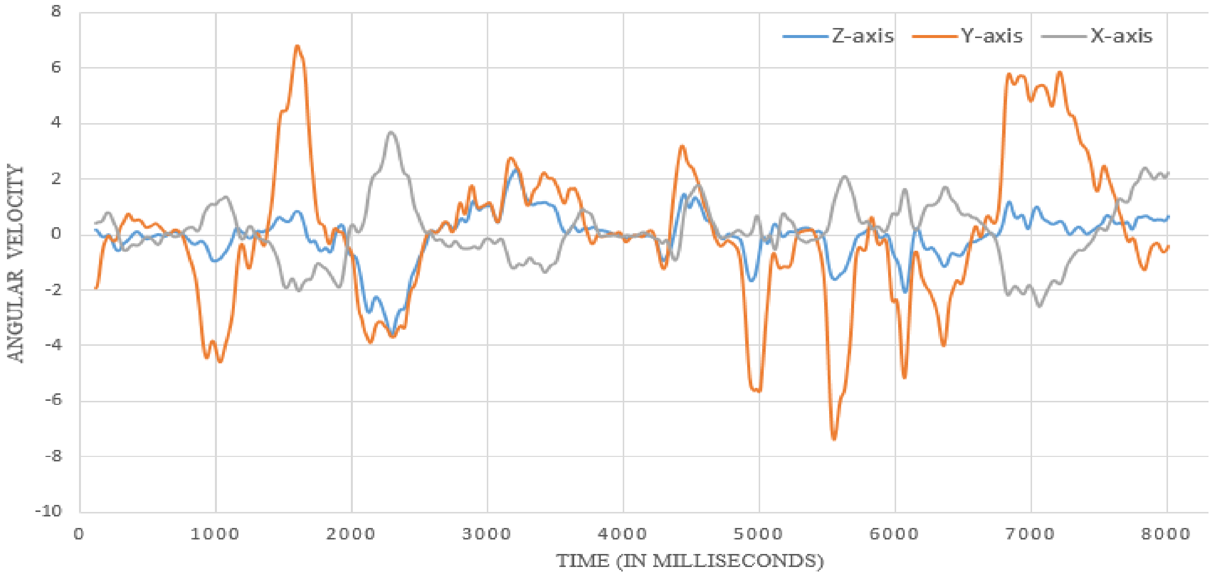


Fig. 1.3: Gyroscope sensors values are generated by rotating the phone consecutively along the x, y, and z-axis. These axes are represented by the grey, red and blue lines respectively

1.2.1. C Magnetometer

The magnetometer is a magnetic field sensor. Sensor values are measured in earth's magnetic field in each direction (x, y, and z), and can be utilized to detect the orientation of the phone with respect to the Earth's surface. Sensor values are measured in μT (micro-Tesla).

A cell phone's magnetic field sensor is extremely sensitive to nearby disturbances, making it easy to mislead it. Refrigerators, loudspeakers, and other electronic equipment can easily alter the measurements of this sensor. It is called noise when there are local disturbances in the magnetic field.

1.2.1. D Gravity sensor

The gravity sensor gives a three-dimensional vector indicating the direction and magnitude of gravity. Typically, this sensor is used to identify the device's relative orientation in the environment.

1.2.2 Environmental sensors

Environmental sensors are important because they can give extra information on the location of the device. Environmental sensors are used primarily for obtaining location of static users, rather than moving users. To track walking or moving users, inertial sensors are needed to be considered because walking with a device will result in high activity of the motion sensor. For instance, walking with a phone in your pocket will result in high activity of the motion sensor, and the signal strength of Wi-Fi, Bluetooth, and GSM are expected to change accordingly.

1.2.2. A Wi-Fi

The Received Signal Strength Indication (RSSI) and angle-of-arrival (AoA) are the two most common parameters used by localization techniques that are based on Wi-Fi signals. Signal strength is used by the vast majority of these systems. Because it needs more sensitive equipment, angle-of-arrival is used less. A key property to recognize when using RSSI for localization is that of signal attenuation. There is a relation between signal strength and distance between transmitter and receiver in all RSSI-based localization systems. As the distance between the transmitter and receiver rises, so does the signal intensity.

1.2.2. B Bluetooth

Like Wi-Fi, Bluetooth signal attenuation can be used to estimate distances. However, the Bluetooth protocol was made to get rid of wired communication systems so that people could move around more. But Bluetooth observations don't always give information about location. To make this situation even worse, using a mobile Bluetooth headset would lead motion sensors and environmental sensors to provide results that are inconsistent with one another. Since the headset moves with the user, the Bluetooth sensors would show that the user is not moving, while motion sensors might show that the user is very active. When there is no feedback and only information from sensors is available, there is no way to get rid of this contradictory information.

1.2.2. C GPS

GPS is a popular method for localization and navigation. The approach is based on satellite readings and is most commonly used outside. Due to poor precision, indoor applications are frequently ruled out. The accuracy of measuring devices in indoor applications is strongly dependent on the kind of building and the quality of the measuring equipment. However, GPS may be used as an extra feature for localization in order to overcome drift problems in areas of the building that are GPS-friendly.

1.2.2. D GSM

For the most accurate location based on GSM towers, the GSM network needs to have a high density. However, even in cities, the density is often not high enough to include GSM sensors as a feature. Compared to other techniques such as Wi-Fi and Bluetooth, GSM towers can only assist in coarse positioning. The primary advantage of using cell tower RSSI rather than Wi-Fi RSSI is that all mobile devices, by default, are already scanning for cell towers, whereas sensors such as Wi-Fi need additional power consumption from the device's battery. In addition, the quality of Wi-Fi signals can be rather low in many outdoor settings, and it's possible that they won't even be there at all. Scanning continuously for Wi-Fi networks would thus only result in unnecessary use of battery power if it was done continuously.

1.3 Localization Technology

1.3.1 Radio Frequency Identification (RFID)

RFID, or radio frequency identification, is a common kind of wireless data transfer. Radio waves are used to send digital data, which is usually an object's identification number. The most common application of this technology is in a large-scale system. When an RFID reader and an RFID tag are in communication, it instantly detects the distinct radio frequencies being exchanged. The readers will receive the emitted radio signal from the tags. The tags include a microchip that activates the tracking of objects. There are now two types of RFID tags. Active RFIDs and Passive RFIDs. Active RFIDs operate in the Ultra High Frequency (UHF) and microwave frequency range and Passive RFIDs are limited in communication range (1-2m) and can operate without a battery. They are smaller, lighter, and cost less than the active ones; they can work in the low, high, UHF, and microwave frequency range [1]. It has a transceiver, antenna, controllers, power supply, and a server connection interface. In order for RFID tags and readers to communicate, a radio frequency protocol must be followed.

1.3.2 Ultra-Wideband (UWB)

The term "Ultra-Wideband" refers to a Radio Frequency (RF) signal that has a large bandwidth communication. This signal is often more than 500 MHz and exceeds a range of 20-25 % of the broad-spectrum radio frequency. UWB has a large channel for communication that expands out over a wide range of radio frequencies. In addition, UWB transmitters use less transmission energy despite the fact that they can send an incredible number of bits of data. In the localization system, the distance from the reference point to the target point is calculated using Time of Arrival (ToA) or Time Difference of Arrival (TDoA) method, which is based on RF signals.

1.3.3 Ultrasonic

Ultrasonic waves are mechanical waves that transmit oscillation or vibration of pressure in a medium. Ultrasound waves have the advantages of being short-ranged and having a high resolution. Ultrasound waves, do not interfere with magnetic fields or electromagnetic waves. In indoor localization systems, ultrasonic waves move through the air and building materials. Time of Arrival (ToA) based approaches are also used to measure the travel distance between emitters and receivers of ultrasound pulses. There are often three or more emitters that are used to estimate the coordinates, but the receivers are always fixed in place.

1.3.4 Bluetooth

Bluetooth is a ubiquitous wireless technology that has a limited range and a common standard of Wireless Personal Area Networks (WPANs) is used for transferring data from one device to another. Bluetooth runs in the ISM band at a frequency of around 2.45 GHz. Because of its lower power consumption, lower gross bit rate, and shorter range as compared to WLAN, Bluetooth is a popular alternative communication method. Short-range Bluetooth connection is possible up to a distance of around 10-20 metres, which is known as Bluetooth 2.0. Since the range covered by Bluetooth signal is short in nature, and get affected by ambient factors, usually Bluetooth-based localization provides better accuracy for coarse-grained localization rather than fine-grained localization. The prior version of this is Bluetooth 3.0 also known as high speed is significantly faster than Bluetooth 2.0, the data transfer of this version is 25 Mbps.

1.3.5 ZigBee

ZigBee is built upon the IEEE 802.15.4 standard that is concerned with the physical and MAC layers. It is a wireless data transfer technology that has a low data transfer rate and a limited range. ZigBee is an ideal protocol for high-level communications, making it well-suited for the provision of personal area networks such as those used for medical operation application services, home automation systems, and security networks. The low cost and simple system requirements of this technology are among its many attractive features. ZigBee is a single-piece technology that uses a little amount of power and transfers data at a slow rate.

While ZigBee is favourable for localization of sensors in wireless sensor network (WSN), but it is not commonly available on the majority of users' devices, because of this, it is not a good option for the indoor localization of users.

1.4 Motivation

Localization-based services, such as human/object tracking and advertising, are becoming increasingly popular. Currently, infrastructure-based technology known as the global navigation satellite system (GNSS) is used to approach the majority of the localization services. The most well-known example of GNSS is the global positioning system or GPS. Though GPS is outstanding for outdoor localization, the signal suffers from fading or multipath propagation property in indoor areas, due to complex structures and obstacles. So, technologies for indoor positioning relies on the other sensors commonly available indoor areas, such as Wi-Fi, Bluetooth or IMU. However, the indoor sensor values may get affected by various ambient conditions. To evaluate a newly proposed localization method, at first, a dataset is needed that comprises of data collected under different ambient and/or physical conditions. Most of the publicly available datasets lack this variety and hence if a newly proposed localization approach is evaluated using these datasets, which will stick at a point to reach generalization. From this motivation, an effort has been made in this work to collect two different kinds of sensor data (Wi-Fi and IMU) that will reduce the time and laborious effort of on-site survey and enhance the evaluation process of researchers in the domain of localization.

1.5 Contribution

The contributions of this work are as follows:

1. It provides a Wi-Fi based dataset JU-MetroLoc for indoor localization collected using a commercial smartphone application implemented for this work. Data has been collected considering all kinds of diversity of user, device and environmental scenarios.
2. To our best knowledge, the JU-MetroLoc dataset is the first dataset in India collected from two metro stations, among which one is at underground level.

3. Localization performances on JU-MetroLoc dataset are evaluated using different supervised ML classifiers, with varying granularity levels.
4. It provides an inertial sensor dataset collected using commercial android smartphones that achieves device heterogeneity and user heterogeneity.
5. For inertial dataset, data from different sensors including accelerometer, gyroscope and gravitational sensor have been collected to analyze the effect of these sensors on step detection, direction change and device orientation.

1.6 Scope of the work

Using these two datasets, in future an inertial navigation system can be designed and/or implemented in which the static position of the user can be identified based on the Wi-Fi data, and the navigation path can be predicted using inertial sensor data, when the user is moving. As different kinds of sensor data are available to predict the steps, turns, and orientation, changes in routes can be predicted easily using these sensor values.

1.7 Thesis organization

The thesis is organized as follows:

- In chapter 1 localization and how RSSI-based localization works have discussed. Different possible applications of indoor localization have been discussed. Different sensors and technologies that can be used for localization are also discussed in this chapter. Motivation, contributions and scope of the work have been discussed in this chapter.
- In chapter 2 machine learning framework based on localization have discussed.
- In chapter 3 a proposed design for two different types of datasets to assist localization have shown.
- In chapter 4 the details information of these two datasets have described.
- In chapter 5 experiment result, the graphical plot and baseline Accuracy of these datasets have shown.
- Concludes the thesis.

Chapter 2

Machine Learning Framework

Large volumes of data can be gathered and utilised to assist in the process of localization thanks to the widespread availability of an increasing number of sensors in mobile devices. These sensors can be found in almost all mobile devices. The use of machine learning methods is a logical solution for the problem of filtering through these vast datasets and finding the key bits of information for localization. Machine learning algorithms could also be a fast and effective way for localization, which is often more useful for applications than static localization.

2.1 Data Collection

During the phase of data collection, users move about the building or any infrastructure and take measurements of the strengths of the Wi-Fi signals as well as gather data from the various inertial sensors (accelerometer, gravity, and gyroscope) contained in the device. The data is saved to a CSV file on the device, which can be retrieved from the device at a later time for analysis. Additionally, the data is strongly associated with a location that has been specified by the user.

The two different modes of dataset collection are described below.

2.1.1 Dedicated user-based data collection

Users who are interested to participate in the data collection work collect RSS of accessible APs from each location point or reference point using smart devices. Dedicated users collect inertial sensor data using the sensor measurement app. The location points are chosen according to the ground truth decided by the work. Ground truth is used to guide the selection of the location points.

2.1.2 Crowdsourcing-based data collection

The process of accomplishing a task by enlisting the help of several individuals is known as crowdsourcing. It's getting more common to use crowdsourcing approaches to produce the database of localization. Massive volumes of data can be captured because there are millions of users and devices participating in IoT-based Applications. As long as such data is used efficiently, the issue of creating localization databases may be handled.

According on whether the users actively participate, the crowdsourcing-based database generation technique can be separated into active and passive crowdsourcing approaches. Active methods of crowdsourcing mostly use an interactive interface to let users take part in building a database. With passive crowdsourcing, the data uploaded from consumer devices is analysed without bothering the users, and the database is set up automatically. Based on this idea, sensor-based localization makes it possible for mobile users to automatically collect Wi-Fi fingerprints while going about their daily lives.

2.2 Data pre-processing

The wireless nodes received raw data that included a variety of data, including a timestamp for when the signal was collected, a signal counter, and an RSSI value for each AP (access point). In order to make this data useful for localization, the raw data has to be processed. Data processing was used to analyse the original data by breaking it down into several steps. The first step involved filling up the blank (NULL) field with the previous (t–1) data point. Where t indicates the current timestamp and (t-1) represents the previous RSSI value. The removal of inconsistent AP is the next phase. During data collecting, APs from surrounding buildings or even hotspots might be heard. In addition to having limited signal strength, some APs may not always be accessible to use. So, it's possible that those APs won't be beneficial. Using these APs in analysis will result in inaccurate location predictions due to the noise. As a result, those APs have to be removed. And for inertial sensor data, there are many filtering tools like band-pass filter, mean filter, low-pass filter, Gaussian filter, wavelet filter and Kalman filter used to de-noise the raw sensor data.

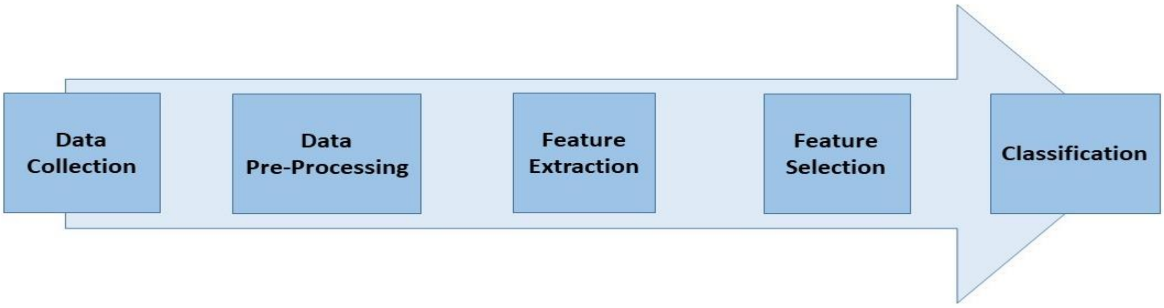


Fig. 2.1: Machine Learning Framework

2.3 Feature extraction

Feature extraction is a process of dimensionality reduction. It is able to transform data from a high-dimensional space into data from a lower-dimensional space without sacrificing any of the information. It is helpful for processing massive datasets that need a lot of computational resources in an effective and efficient manner. There are usually 3 types of features in the case of inertial sensor data, namely, time-domain feature, frequency-domain features and wavelet-domain features [26]. For Wi-Fi sensor data the APs are used as a feature, and the total number of APs determines the size of the feature (signal) space.

Time-domain features:

Mean: The direct component of the signal. The mean is calculated by adding all the data points and dividing them by the number of data points.

$$\text{Mean } (\bar{x}) = \frac{1}{m} \sum_{i=1}^{m} xi$$

Where:

xi = Value of the i th point in the data set.

m = the number of data points in the data set.

Standard deviation: Reduce the data spread. Data spread is measured by the standard deviation (σ), which tells us how far off our data is from the mean. The lower the standard deviation, the closer the data are to the mean, whereas the higher the standard deviation, the more dispersed the data are.

$$\text{Standard Deviation } (\sigma) = \sqrt{\frac{1}{m}\sum_{i=1}^{m} (xi - \bar{x})^2}$$

Where:

xi = Value of the i th point in the data set.

$\bar{x}$ = the mean value of the data set.

m = the number of data points in the data set.

Max and Min: Show the changing range of the signal.

Frequency-domain features:

In order to conduct a more accurate analysis of the data, we transform each Feature such that it exists in the frequency domain. A feature can be converted into its appropriate domain by Fourier transform. FFT (Fast Fourier transform) is the easiest and fastest way to use DFT (discrete Fourier transform)

When the orientation of the smartphone is changed, the results of the FFT can represent different activities, as demonstrated in Wang et al. [11]. It is quite helpful to have features like FFT energy and frequency domain entropy when you are trying to identify data collected from different positions.

Wavelet-domain features:

A continuous-time signal or a given function can be segmented into a variety of scale components by the use of a mathematical function known as a wavelet. The wavelet transform successfully mixes the time-domain and frequency-domain data because it offers benefits over conventional Fourier transforms for representing functions in various scale components.

2.2.1 Feature Extraction Technique

One of the main feature extraction techniques for Wi-Fi fingerprint data is PCA feature extraction. PCA reduces high-dimensional data to low-dimensional data without losing information. A PCA finds the first, second... and cth principle components, that are orthogonal vectors on which data are projected to achieve the biggest variance, second largest variance... and cth largest variance. Here's how PCA works. Given a set X = $\{x1,x2,...,xn\}$ of original data in d-dimensional feature space, PCA finds a d×c transformation matrix (or projection matrix) W to project data into c-dimensional feature space such that projected data have maximal variance, where $c \ll d$. Standardizing the data involves calculating the mean and standard deviation for each feature. Using standardised data, a d×d covariance matrix is created. After that, d eigenvectors of the covariance matrix are derived, and each eigenvector has an eigenvalue that corresponds to it. Eigenvectors are sorted by eigenvalues. The first c eigenvectors $v_1, v_2 \ldots, v_c$ with eigenvalues $e_1, e_2, \ldots, e_c$ are then combined to form the projection matrix W. The eigenvalue e_i, $1 \leq i \geq c$, of eigenvector vi represents the projected variance of v_i. According to Equation (4), ρ is the ratio of c eigenvalues to d eigenvalues. ρ is the cumulative proportion or explained variance ratio. The explained variance ratio ρ should exceed a threshold θ (e.g., 0.9 or 0.95) so the c principal components can explain (or reflect) the total data variance.

$$\rho = \frac{\sum_{i=1}^{c} e_\mathrm{i}}{\sum_{i=1}^{d} e_\mathrm{i}} \tag{1}$$

Each original d-dimensional data sample x_i can be projected onto the c principal components to be a new c-dimensional sample $x_\mathrm{i}^{'}$ according to Equation (2). The initial d-dimensional data sample x_i has d data features, while x_i' has c data features. This accomplishes dimensionality reduction and feature extraction [13].

$$(x_\mathrm{i}')_{1\times C} = (x_\mathrm{i})_{1\times d}(W)_{d\times c'}\ 1 \leq i \leq n \tag{2}$$

2.4 Feature selection

Because the efficiency of a classification algorithm hugely depends on the dimensions of the feature space, it is important to limit and eliminate the impacts of dimensionality. One method for reducing dimensionality is to use feature selection algorithms, which are used to identify and remove features that contribute nothing to the classifier's success. Sequential search methods, such as branch bound searches [12] and the Pupil algorithm [12], are commonly used in feature selection algorithms for inertial sensor data.

Feature selection is also a useful way to reduce biasing during training. In general. There are two ways of feature selection as follows.

(1) **Filter method:** In this method, features are scored and ordered based on a search algorithm. The ordered list of features shows which features contribute the most to the model. For the Wi-Fi dataset, there are some APs which have very less no of fingerprints. So, we can remove those APs and then we have to rewrite the database.

(2) **Wrapper method:** This method uses various combinations of the features and compare results for a given classifier, this information is then used to evaluate which feature combinations result in the most accurate model.

2.5 Classification

In general, there are three types of machine learning: supervised, unsupervised, and reinforcement learning. Unsupervised learning approaches use unlabelled data instances, while supervised learning requires labelled instances. Reinforcement learning lies between supervised and unsupervised learning because it doesn't label all training examples, only sparse, time-delayed feedbacks, called rewards. For Wi-Fi fingerprinting, the training dataset consists of both Wi-Fi RSS (the object) and a label. So it's supervised learning.

The classification of localization is an essential component in order to provide location-based services. It can classify two sequential queries. It can determine a target's location in relation to the site i.e., inside/outside of the region. If so, which area in the region the target is located. Once we established that the RSSI vector of the target is measured within the fingerprint region, we can localise the target's precise location.

Wi-Fi fingerprint-based localization could be performed in two phases, on offline phase where the Wi-Fi fingerprint and their corresponding area IDs are collected. After that, classification models are used for area categorization and in-out decisions and store the model parameters in the database. following that During the online phase, the user samples an RSSI vector and then directs his or her module to determine if the user is inside the fingerprint region or not. If so, the vector is then used to locate and return the user's current location by feeding it into the area categorization module [14].

2.6 Summary

Generalise a Machine learning framework which is commonly followed by smartphone-based localization applications are summarized in this chapter as it evident data collection, and hence dataset design from the very baseline for any localization system using smartphone. Hence our proposed design of dataset is described in detail in the next chapter.

Chapter 3

Proposed Dataset Design

3.1 Overview

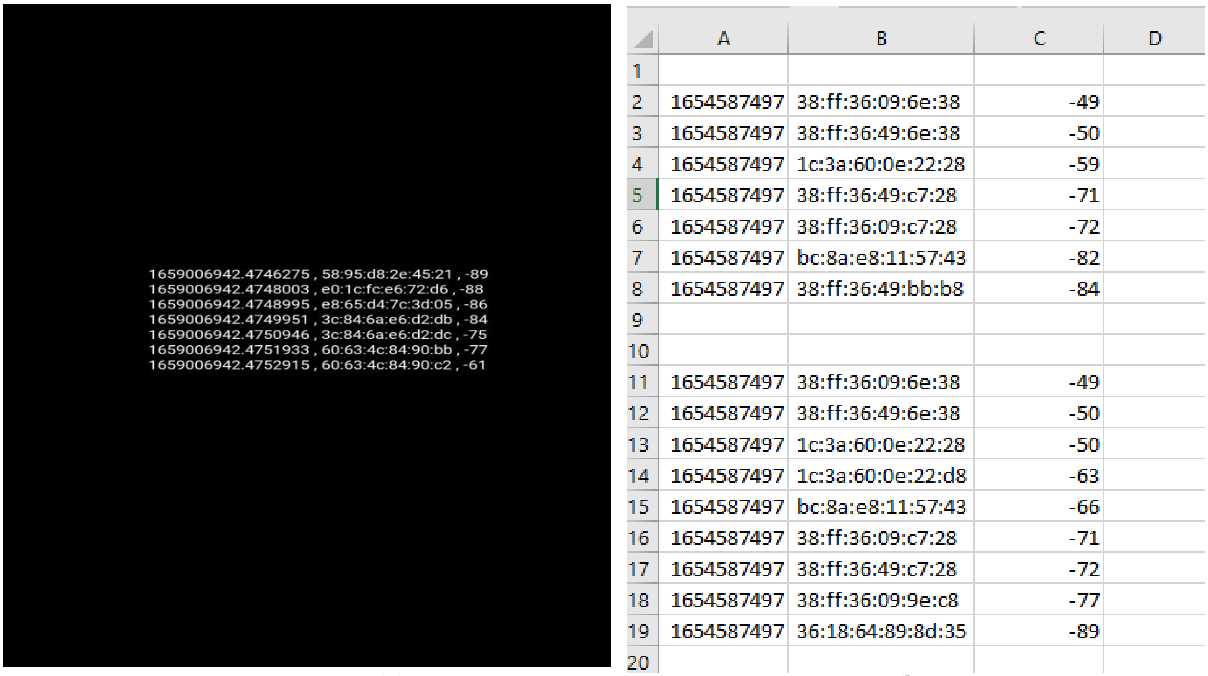
In order to assist localization, two different types of localization datasets were designed by us. One is based on smartphone sensors like accelerometer, gyroscope, and gravity sensor. The other one is infrastructure based where the data was collected from mobile Wi-Fi sensor, at two different types of metro stations one is underground metro and the other one is a ground-level metro station.

3.2 Data Collection

The data collection aim was to test the feasibility and accuracy of localization using realistic data. An android application named JU-Wi-Fi Scanner was implemented and used to collect Wi-Fi data for localization (fig. 3.1(a)), To collect inertial sensor data, another app called Sensor Logger was used, which is available on the Google Play Store for free of cost(fig. 3.2(a)). At the same time, these applications can be used to collect data and display the value in real time.

In order to gather WI-FI data, users walked from one end of a metro platform to the other while the app is running on his phone. The application scans and captures RSSI values from all the available APs around it, along with specific timestamps. The collected RSSI data is stored as a CSV format (shown in fig. 3.1(b)). RSSI sampling can be affected by different environmental factors and person’s movement. Even when using the same device in the same area at different times, RSSI value could be different, which could effect on the accuracy of the localization. So, to achieve device heterogeneity, four different smartphones were used to collect RSSI samples. Since the RSSI fingerprint values are not robust to environmental and time changes, enough number of samples were collected at each location point with a considerable timespan.

During the inertial sensor data collecting-phase, the user walked a predefined number of steps while holding or placing their phone in various positions. That time the application collect measurement of accelerometer, gyroscope, and gravity sensor. At the same time, it saved the data as CSV format (shown in fig. 3.2(b)).

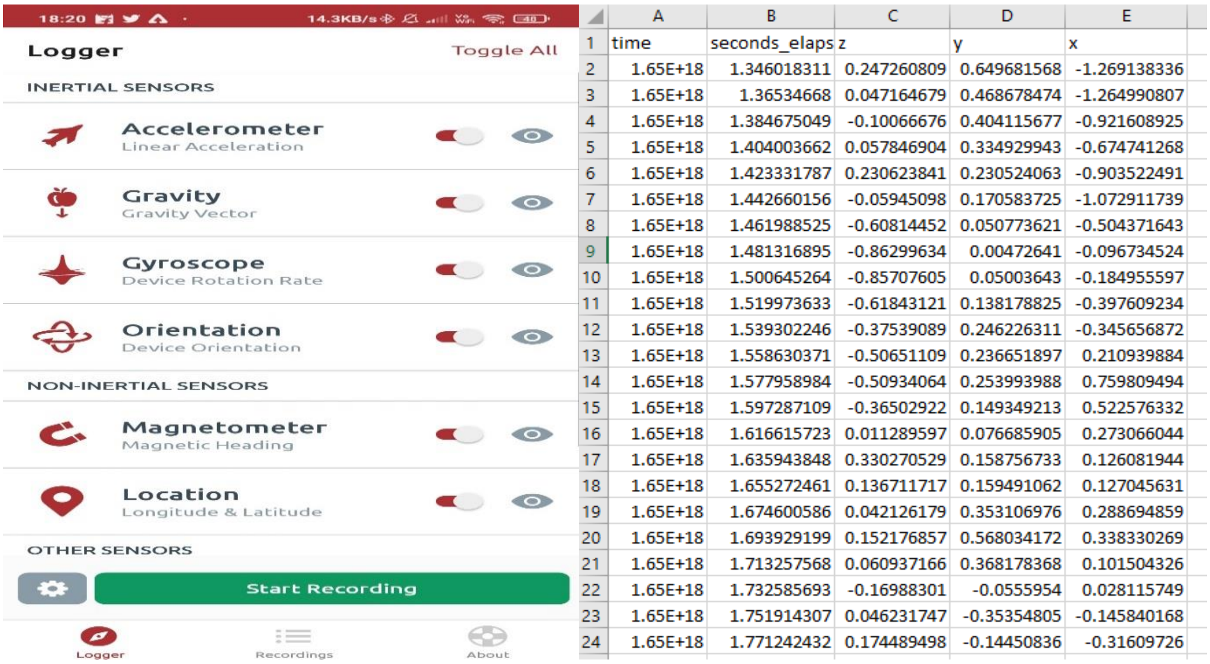


```
1659006942.4746275 , 58:95:d8:2e:45:21 , -89
1659006942.4748003 , e0:1c:fc:e6:72:d6 , -88
1659006942.4748995 , e8:65:d4:7c:3d:05 , -86
1659006942.4749951 , 3c:84:6a:e6:d2:db , -84
1659006942.4750946 , 3c:84:6a:e6:d2:dc , -75
1659006942.4751933 , 60:63:4c:84:90:bb , -77
1659006942.4752915 , 60:63:4c:84:90:c2 , -61
```

	A	B	C	D
1				
2	1654587497	38:ff:36:09:6e:38	-49	
3	1654587497	38:ff:36:49:6e:38	-50	
4	1654587497	1c:3a:60:0e:22:28	-59	
5	1654587497	38:ff:36:49:c7:28	-71	
6	1654587497	38:ff:36:09:c7:28	-72	
7	1654587497	bc:8a:e8:11:57:43	-82	
8	1654587497	38:ff:36:49:bb:b8	-84	
9				
10				
11	1654587497	38:ff:36:09:6e:38	-49	
12	1654587497	38:ff:36:49:6e:38	-50	
13	1654587497	1c:3a:60:0e:22:28	-50	
14	1654587497	1c:3a:60:0e:22:d8	-63	
15	1654587497	bc:8a:e8:11:57:43	-66	
16	1654587497	38:ff:36:09:c7:28	-71	
17	1654587497	38:ff:36:49:c7:28	-72	
18	1654587497	38:ff:36:09:9e:c8	-77	
19	1654587497	36:18:64:89:8d:35	-89	
20				

(a) (b)

Fig. 3.1: (a) JU-Wi-Fi Scanner application main page (b) A snapshot of raw data structure



	A	B	C	D	E
1	time	seconds_elaps	z	y	x
2	1.65E+18	1.346018311	0.247260809	0.649681568	-1.269138336
3	1.65E+18	1.36534668	0.047164679	0.468678474	-1.264990807
4	1.65E+18	1.384675049	-0.10066676	0.404115677	-0.921608925
5	1.65E+18	1.404003662	0.057846904	0.334929943	-0.674741268
6	1.65E+18	1.423331787	0.230623841	0.230524063	-0.903522491
7	1.65E+18	1.442660156	-0.05945098	0.170583725	-1.072911739
8	1.65E+18	1.461988525	-0.60814452	0.050773621	-0.504371643
9	1.65E+18	1.481316895	-0.86299634	0.00472641	-0.096734524
10	1.65E+18	1.500645264	-0.85707605	0.05003643	-0.184955597
11	1.65E+18	1.519973633	-0.61843121	0.138178825	-0.397609234
12	1.65E+18	1.539302246	-0.37539089	0.246226311	-0.345656872
13	1.65E+18	1.558630371	-0.50651109	0.236651897	0.210939884
14	1.65E+18	1.577958984	-0.50934064	0.253993988	0.759809494
15	1.65E+18	1.597287109	-0.36502922	0.149349213	0.522576332
16	1.65E+18	1.616615723	0.011289597	0.076685905	0.273066044
17	1.65E+18	1.635943848	0.330270529	0.158756733	0.126081944
18	1.65E+18	1.655272461	0.136711717	0.159491062	0.127045631
19	1.65E+18	1.674600586	0.042126179	0.353106976	0.288694859
20	1.65E+18	1.693929199	0.152176857	0.568034172	0.338330269
21	1.65E+18	1.713257568	0.060937166	0.368178368	0.101504326
22	1.65E+18	1.732585693	-0.16988301	-0.0555954	0.028115749
23	1.65E+18	1.751914307	0.046231747	-0.35354805	-0.145840168
24	1.65E+18	1.771242432	0.174489498	-0.14450836	-0.31609726

(a) (b)

Fig. 3.2: (a) Sensor logger app main page. (b) A snapshot of accelerometer sensor’s raw data.

3.3 Details Information of JU-MetroLoc Dataset

During the design phase of a dataset, four types of information was captured by us.

3.3.1 AP’s fingerprint

Fingerprint is the data vector containing RSSI values collected at specific location points from different APs fixed at different positions. Considering different ambient conditions, multiple number of data samples have been collected for each location point. Based on the bunch of RSSI vectors for some location point, that particular location point is uniquely identified, and the term “fingerprint” arises from this background concept. Increase in the number of data samples for a particular location point increases the localization accuracy.

3.3.2 Infrastructure properties

The data was collected by us at two different types of infrastructure (metro station), one of which is a ground-level metro and the other one is underground metro. We formed a grid with a size of 27 × 4 across the station area. The RSSI information was collected inside the perimeter of the grids.

The platform of the ground-level metro station has a total of 27 pillars (T1-T27) on one side and a railway track on the other. The data was gathered between those pillars and track (shown in Fig. 3.3). Total of four users walk on the platform in four different positions (P1-P4) from one end to another end in a straight line to collect the data. In comparison the underground station platform has a total of 27 pillars (R1-R27) parallel to each other and the train tracks are on both sides of the platform (shown in Fig. 3.4).

In Fig. 3.3 & 3.4 the overview map of metro stations is shown.P1, P2, P3, and P4 indicate the position of four users, and the Dotted line represent the user’s walking path on the platform. The cylinder shapes T1, T1…, and T27 represent the pillar of ground-level metro station and R1, R2…, and R27 represent the pillar of underground metro station.

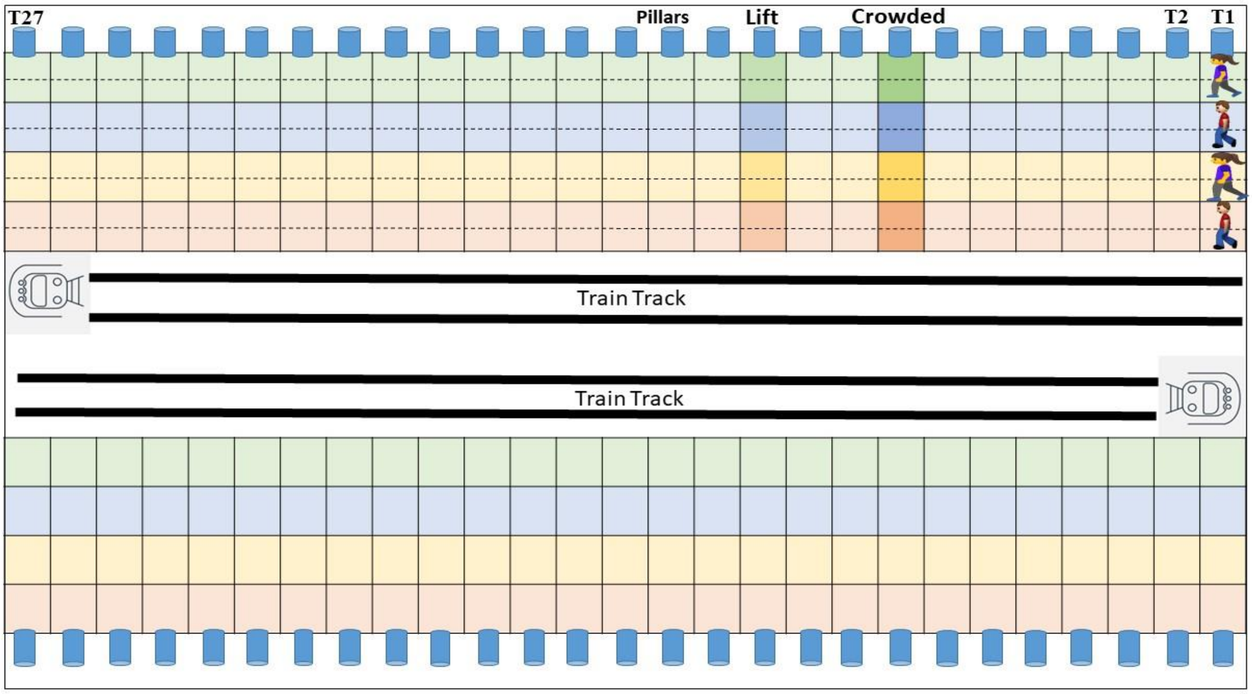


Fig. 3.3: Overview map of ground-level metro station

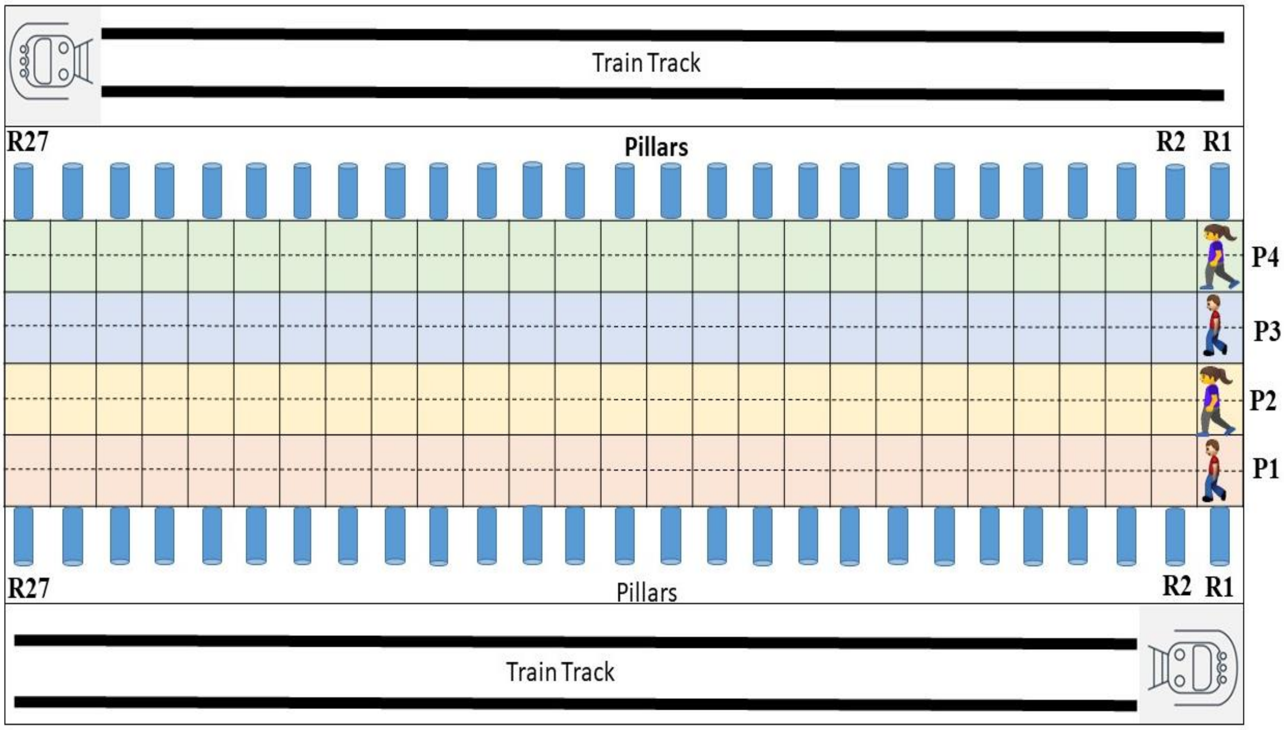


Fig. 3.4: Overview map of underground metro station.

3.3.3 Ambient conditions

The data was collected across a wide range of environmental conditions, such as when there was a significant number of people present in a metro station. In addition to this, we take note of the presence and absence of the train and presence of heavy electrical appliances like near lift (Accelerator) area (shown in Fig. 3.3).

3.3.4 User and Device Information

Four users with Smartphones with unique configurations were used for data collection to understand the variation of signal strength with respect to hardware change (Table. 3.1). The signal strength of the Wi-Fi signal can also be affected by individual users. Because water makes up approximately 70% of the human body, it has the ability to absorb some of the radio frequency signal. Due to the fact that people spend the majority of their time indoors, the human body has a significant impact on radio frequency signal by causing different degrees of signal decline and uncertainty [16].

Device name	Technology	Configuration
Poco x3 pro	GSM/HSPA/LTE, HSPA 42.2/5.76 Mbps	Android 11, upgradable to android 12,6 GB RAM Qualcomm Snapdragon 860, Octa-core (1x2.96 GHz & 3x2.42 GHz & 4x1.78)
Redmi note 10 pro	GSM/HSPA/LTE, HSPA 42.2/5.76 Mbps	Android 11, upgradable to android 12, 6 GB RAM Qualcomm SM7150 Snapdragon 732G, Octa-core (2x2.3 GHz & 6x1.8 GHz)
Samsung Galaxy A03 Core	GSM/HSPA/LTE, HSPA 21.1/5.76 Mbps	Android 11,2 GB RAM, Unison SC9863A, Octa-core (4x1.6 GHz & 4x1.6 GHz)

Samsung galaxy a21s	GSM/HSPA/LTE, HSPA 42.2/5.76 Mbps	Android 10, 6 GB RAM, upgradable to android 12, Exynox 850, Octa-core (4x2 GHz & 4x2 GHz)
Samsung galaxy M12	GSM/HSPA/LTE, HSPA 42.2/5.76 Mbps	Android 11, 6 GB RAM, Exynox 850, Octa-core (4x2 GHz & 4x2 GHz)

Table. 3.1: Details specification of smartphone used in Wi-Fi data collection.

3.4 Details Information of JU-Accelerometer Dataset

3.4.1 Walking pattern

The designed dataset has a mix of different types of walking patterns. The pattern includes the number of turns and the direction of turn (Right turn, Left turn, U-turn). We also take note that after how many steps user took the turn.

3.4.2 User and Device usage details

One user with two different Smartphones were used for data collection to understand the variation of sensor measurement value with respect to different chipsets (Table. 3.2). Smartphone’s inertial sensors measurement value changes for different motions like shaking, tilting, swinging, and rotating. The measurement value also changes according to the change of phone orientation. Because of this, the data was collected by placing the smartphone at six different positions of the body (shown in Fig. 3.5**)**. We also record the smartphone's orientation at the time of data collection.

Device name	Sensors	Configuration
Poco x3 pro	accelerometer, gyro, proximity, compass	Android 11, upgradable to android 12, Qualcomm Snapdragon 860, Octa-core (1x2.96 GHz & 3x2.42 GHz & 4x1.78)
Realme 5 pro	accelerometer, gyro, proximity, compass	Android 9.0 (Pie), upgradable to Android 11, Qualcomm SDM712 Snapdragon 712, Octa-core (2x2.3 GHz & 6x1.7 GHz)

Table. 3.2: Details specification of smartphone used in inertial data collection.

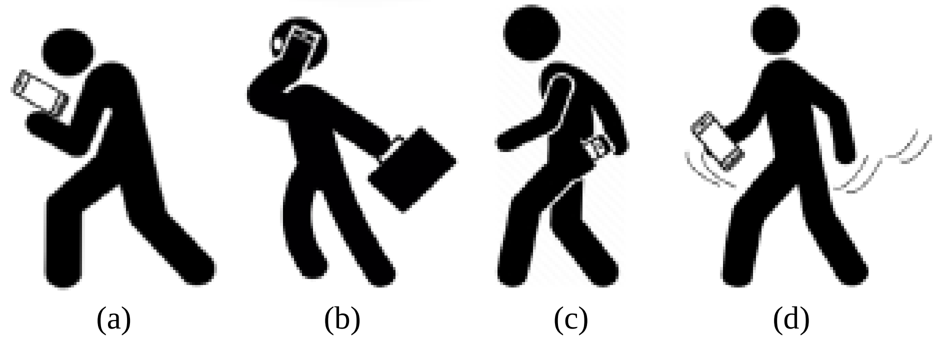


Fig. 3.5: Different Smartphone positioning (a) Normal using. (b) Talking on phone. (c)Trousers’ pocket. (d) Swing.

3.4.3 Ambient conditions

Data was collected using a number of various environmental conditions, such as placing the phone in a variety of different positions on the body, because each position on the body generates a unique motion pattern. The data from an accelerometer can be affected by a variety of environmental factors, such as noise, humidity, temperature, and electromagnetic interference.

Chapter 4

Dataset Description

4.1 JU-MetroLoc Dataset

Datasets were generated at two different locations: (1) Tollygunge Metro [A ground-level metro) (2) Rabindra Sorobor Metro [An underground metro]. The JU-MetroLoc Dataset were gathered using an application named JU_WIFI_Scanner.

The JU-MetroLoc dataset contains signal strength measurements at 27 pillars of each station. At each of the 27 pillars, 4 users collected the data at 4 different positions by 4 different devices. This resulted in a dataset of 3031 data points at metro station 1 and 1600 data points at Metro station 2, where each data point consists of (1) Timestamp, (2) MAC address, (3) RSS (received signal strength) measured in dBm, (4) User, (5) Pillar position, (6) User position (Table 4.1).

(1)	(2)	-----	(78)	(79)	(80)	(81)	(82)
Timestamp	AP1	-------	AP77	User	Pillar	Position	Piller_Position
1654584837	-74		-90	U1	T1	P1	T1_P1

Table 4.1: JU-MetroLoc dataset structure

Timestamp (Column 1) indicates the data collection time in milliseconds. A Service Set Identifier (SSID) and a Basic Service Set Identification (BSSID) are used to identify each AP. But it has been shown that many APs may have the same SSID. Thus, BSSIDs or MAC addresses are used to identify APs [6]. These BSSIDs have been changed to AP1-AP77 for privacy concerns. The most crucial information for localization is supplied by the RSSI values of the various APs in AP1 to AP77 (Column 2 — 78). User (Column 79) represents the unique user who collected the data. Pillar (Column 80) represents the data collection point at the station where data was collected. Where T represent the Tollygunge station. It varies between T1 to T27 because every station has 27 pillars (Table 4.2). Position (Column 81) represents user's location at the pillar. Piller_Position (Column 81) is the pinpoint position of the user, it is used to represent the location point where the RSSI values are recorded. In order to train various machine learning algorithms, RSSI values are used. Other attributes are utilised for classification.

Parameters	Values	
	Metro station 1	Metro station 2
No. of Pillar	27	27
Data collection Area	701.22 m^2	784.81 m^2
Crowd density	3 people per m^2	4 people per m^2

Table 4.2: Details information of Metro station

Parameters	Values	
	Metro station 1	Metro station 2
No. of Instance	3031	1600
No. of APs	77	60
No. of User	4	4
No of measurement point between 2 opposite side pillar	4	4

Table 4.3: Description of JU-MetroLoc dataset

The data collection space of metro station 1 is 128.44 m long and 5.46 m wide. There were several APs directly inside the metro station to provide strong RSS also, there are many other weaker APs from nearby buildings and person’s Wi-Fi which brought the total number of APs in the dataset to 77. Respectably the metro station 2 is 152.1 m long and 5.16 m wide and number of APs in the dataset is 60 (Table 4.3).

4.2 JU- Accelerometer Dataset

The JU- Accelerometer datasets were gathered using the dataset gatherer application named sensor Logger which is available on play store.

The datasets are stored in CSV format, and each data point in the dataset contains: (1) Total Elapsed time of data collection, (2) Tri-axial values of accelerometer, (3) Gyroscope values (device rotation rate), (4) Gravity vector values, (5) Phone Orientation, (6) Direction of Turn (7) Phone model (Table 4.4).

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Elapsed time	AX	AY	AZ	GX	GY	GZ	gx	gy	gz	Ph. Orientation	Turn	Ph. Model
1.3460	-0.921	……	….	….	…..	….	….	….	2.030	C3	(85L)	P1

Table 4.4: JU- Accelerometer dataset structure

Elapsed time (column 1) indicates the data collection time duration in seconds. Accelerometer values of three-axis are shown in AX, AY, AZ (column 2 — 4).GX, GY, GZ indicates the values of gyroscope (column 5 — 7).) Gravity vector values are shown in gx, gy, gz (column 8 — 10). Column 11 represents the holding position of phone. The direction of turns is depicted in Turn (Column 12), it also shows that after how many steps the user takes a turn. At last, the Ph. model (Column 13) represent the unique phone which collected the data.

The details description of JU- Accelerometer dataset has been provided in Table 4.6. JU-Accelerometer data obtained by one participant. Two different smartphones were used to collect the data. Participant collected the data while walking and holding or putting the phone at different positions. A total of 120 steps of walking data were gathered. While walking participant hold or put the phone at 6 different positions (i.e. Holds the phone in right hand and watches the screen, Holds the phone in right hand and swings, Puts the phone in left pocket of trouser). The combined tri-axial values of the smartphone accelerometer, gyroscope and gravity were used to record sensor data, while holding the phone by the participants. At a steady rate of 50 Hz tri-axial values of linear acceleration, angular velocity and intensity of gravity data were obtained. A detailed description of the dataset collected by single smartphone is provided in Table 4.5.

Phone Position	Abbreviates	No. of Samples
Hold phone in right hand and use the phone	C1	3713
Hold phone in right hand and use the phone (with turn)	C1	3580
Holds the phone in right hand and swings	C2	3866
Holds the phone in left hand and swings	C3	3746
Holds the phone in left hand and swings (with turn)	C3	3809
Puts the phone in right pocket of trouser	C5	3754
Puts the phone in left pocket of trouser	C6	3705
Holds the phone in right hand and talks	C7	3625

Table 4.5: Description of JU- Accelerometer dataset obtained from a POCO X3 PRO phone at 50Hz sample rate while walking

Parameters	Values
No. of Steps	120 steps
No. of Instance	54765
No. of User	1
No. of Device	2
No. of Phone holding position	6
No. of Phone orientation	2
No. of Maximum turn	2

Table 4.6: Description of JU- Accelerometer dataset.

Chapter 5

Experimental Result

5.1 JU-MetroLoc Dataset

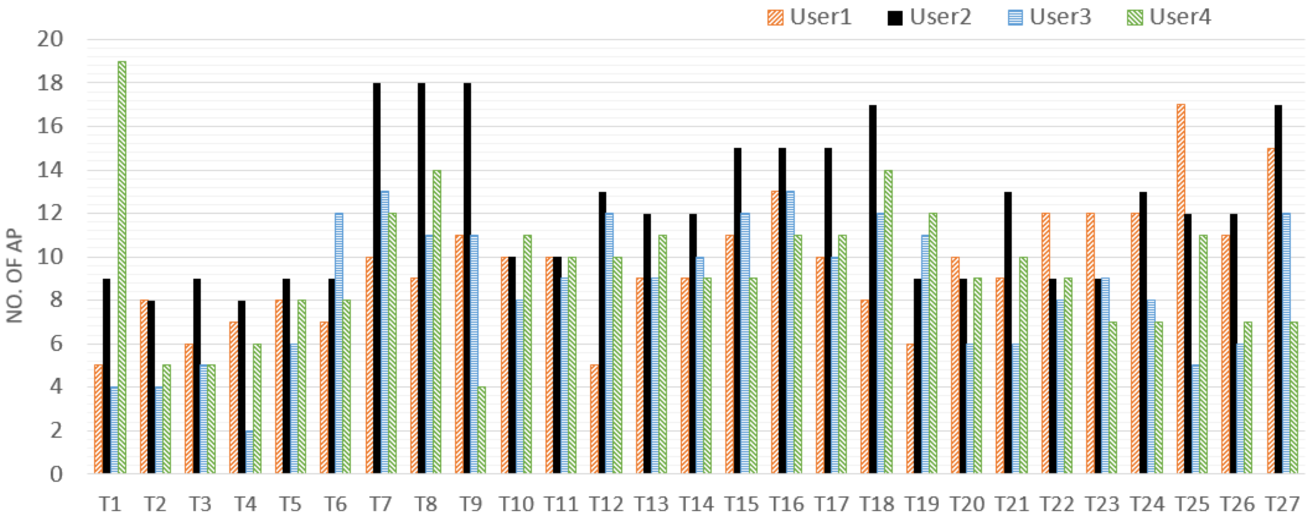


Fig. 5.1: Total number of APs covered by each pillar (T) at metro station 1.

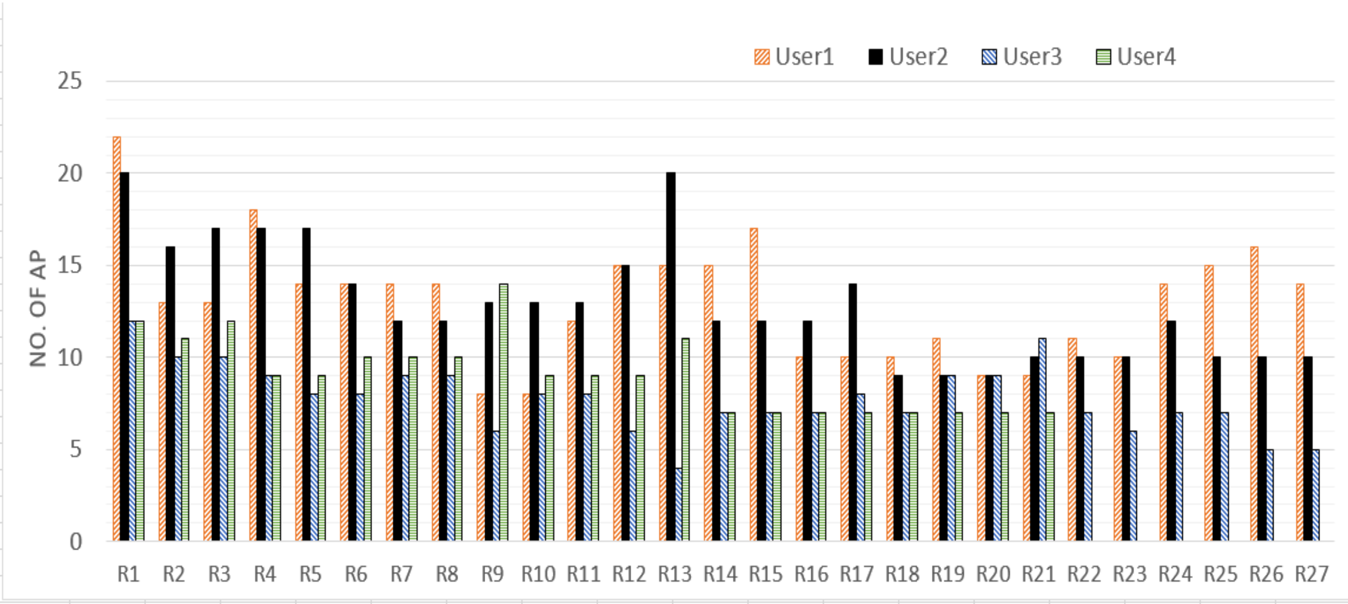


Fig. 5.2 Total number of APs covered by each pillar (R) at metro station 2.

Fig. 5.1 & 5.2 present a general information about the dataset. It shows the total number of APs covered by each pillars. Each pillar cover some APs, a max of 19 APs out of total of 60 APs are covered by a single pillar at metro station 1 and a max of 22 APs out of total of 77 APs are covered by a single pillar at metro station 2. It is important to observe where the maximum APs are covered so the best result for localization can be provided.

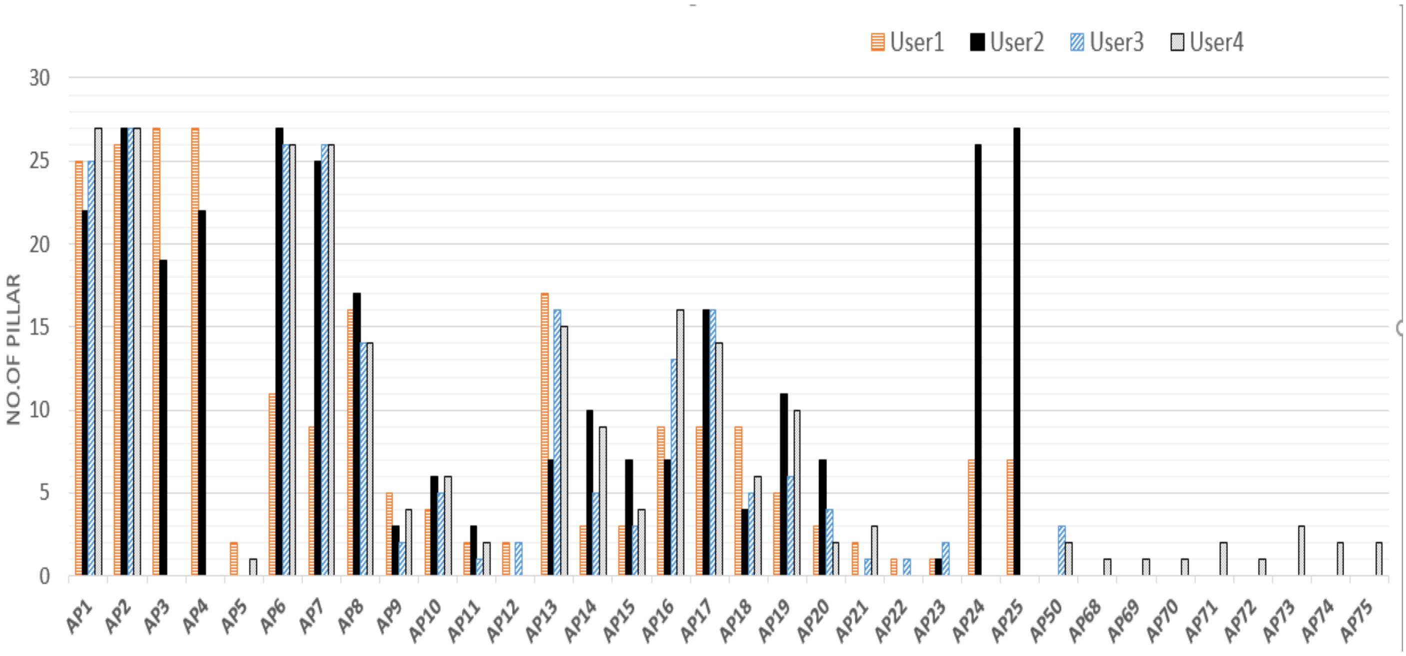


Fig. 5.3: Total number of pillars covered by each AP at metro station 1.

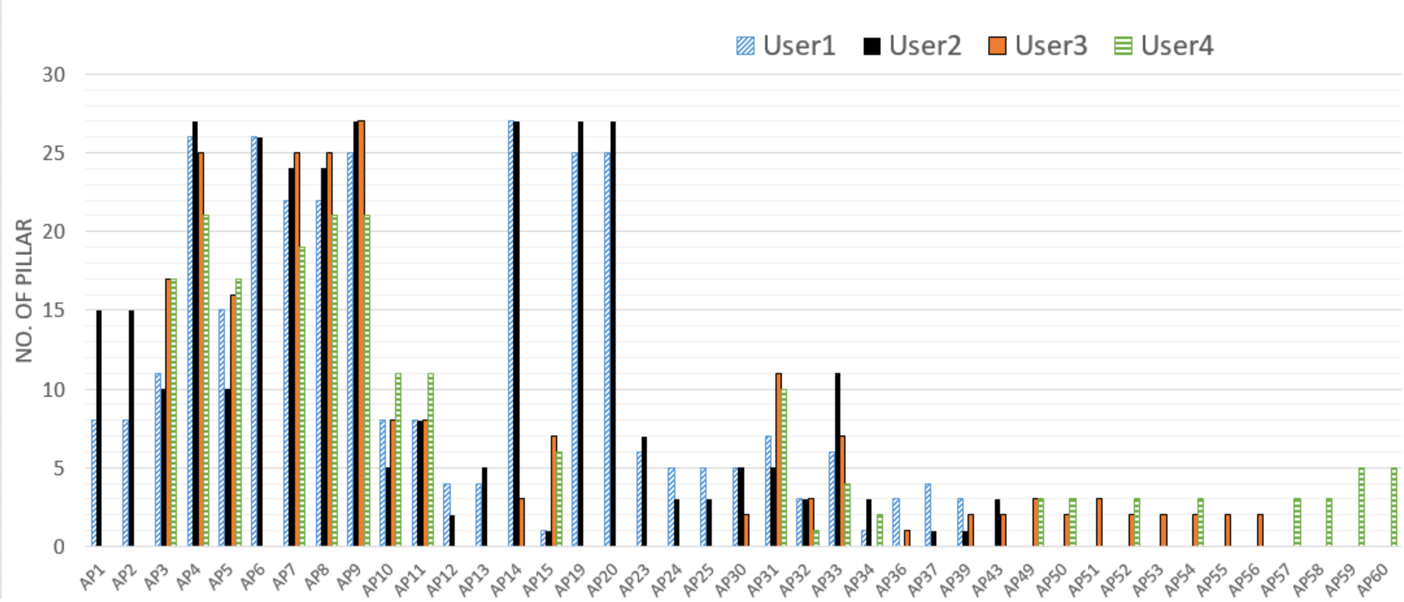
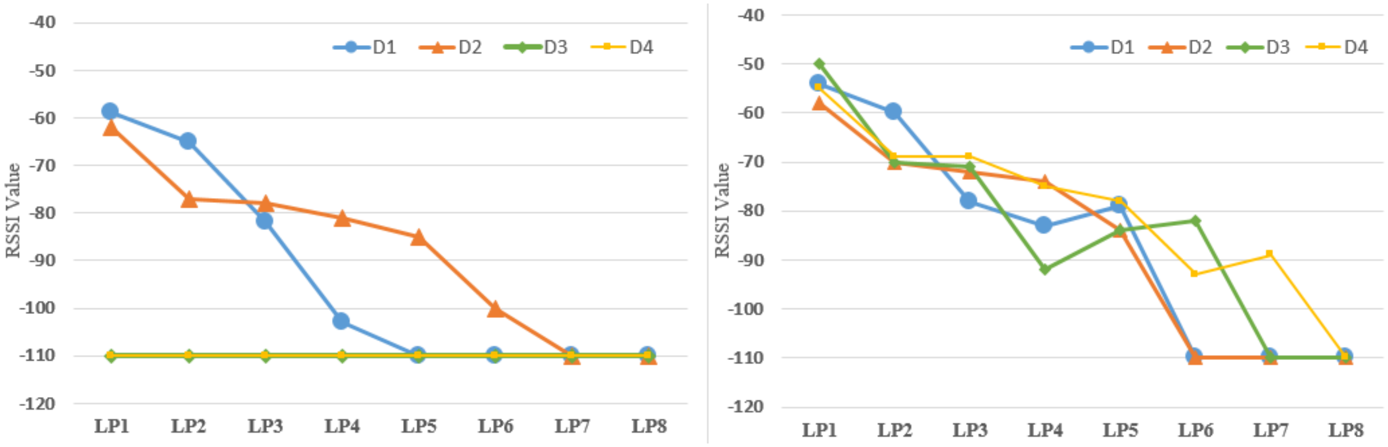
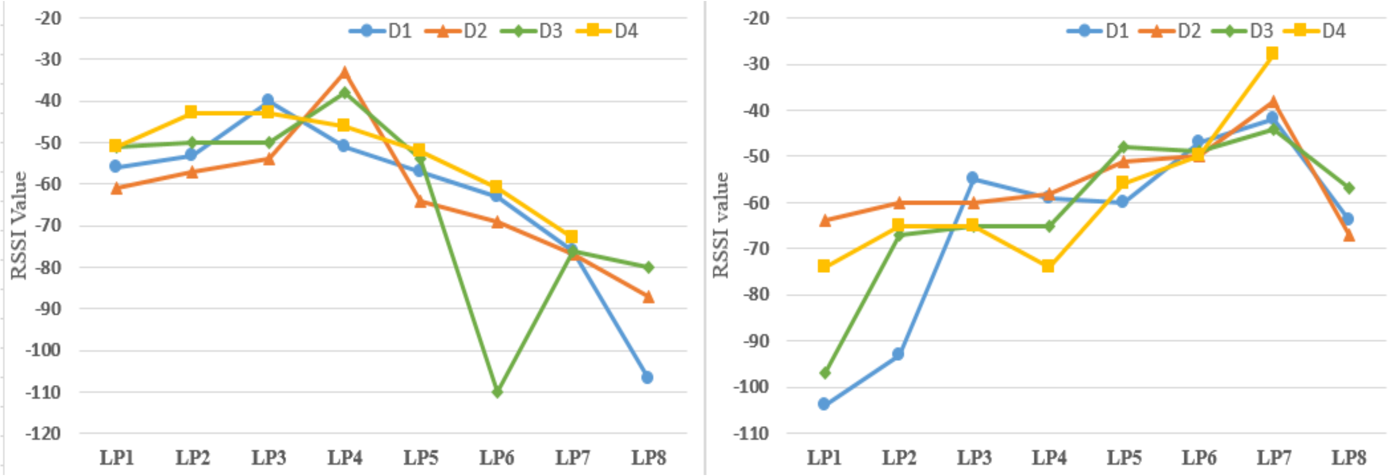


Fig. 5.4: Total number of pillars covered by each AP at metro station 2.

Fig. 5.3 & 5.4 shows the total number of pillars covered by each AP. From these plot graphs, the APs which cover all the pillars of stations can be identified. Here, AP4 through AP9 covered the majority of the pillars, and all four users recorded those APs. The APs which cover only 2-5 pillars can be identified also. Those APs’ signal can come from surrounding buildings or even someone’s hotspots.

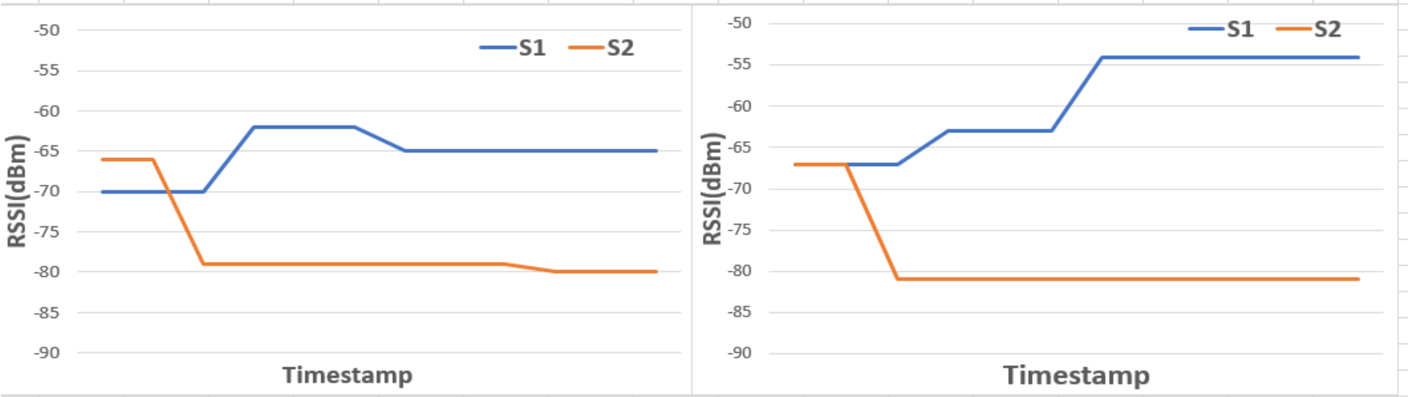


(a) AP1 (b) AP3

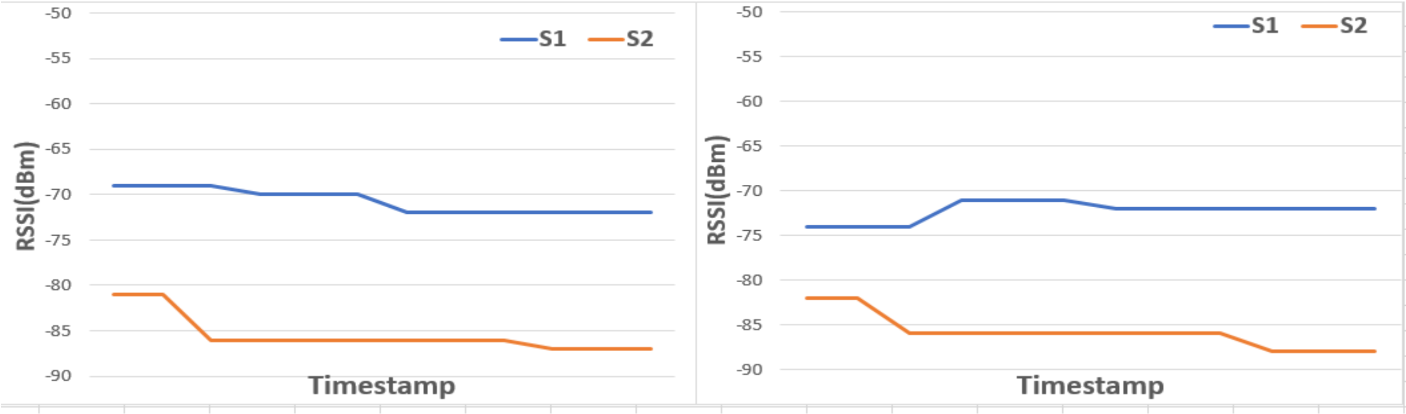


(c) AP4 (d) AP9

Fig. 5.5: Variation of RSSI values in different devices at different location points; D*i* indicates *i* number of devices used in data collection.

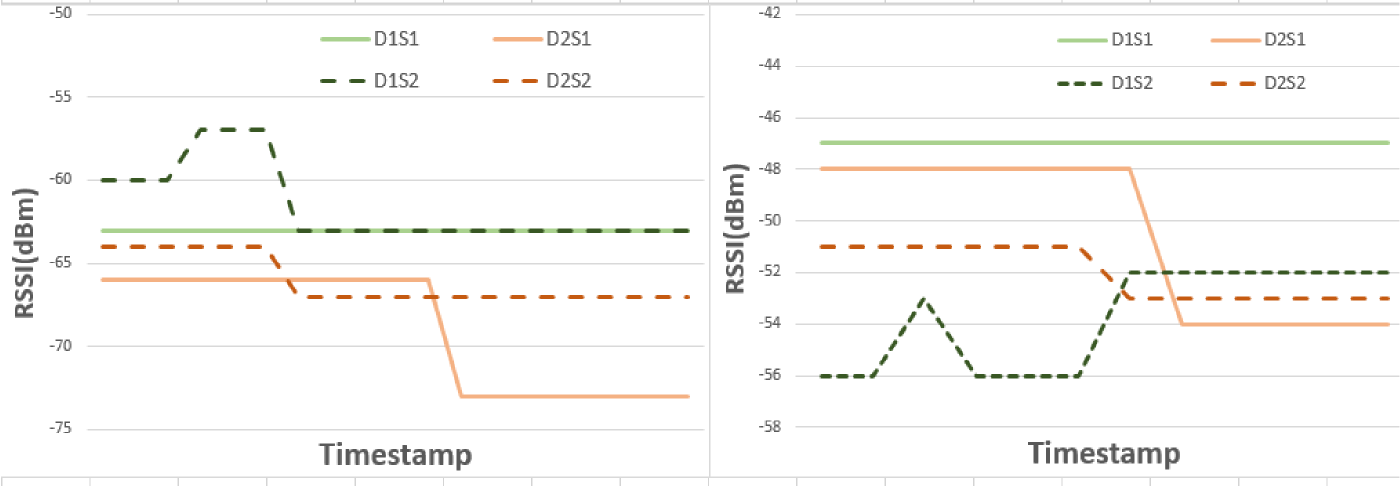


(a) AP1 (b) AP2



(c) AP3 (d) AP4

Fig. 5.6: Distortion of RSSI values due to the presence of train for different APs at metro station 1; S1 indicates normal RSSI value and S2 indicates RSSI values in presence of a train.



(a) AP4 (b) AP9

Fig. 5.7: Changes in RSSI values due to the presence of train at metro station 2; S1 indicates normal RSSI value and S2 indicates RSSI values in presence of train; Di indicates i number of devices.

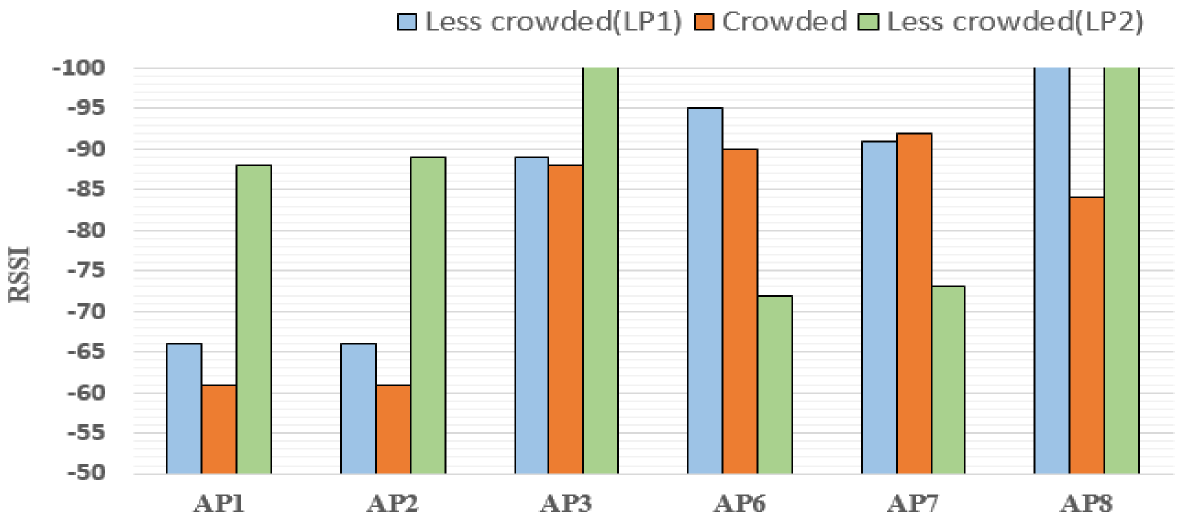


Fig. 5.8: average RSSI values in crowded areas and less crowded areas; here consider two ends of the platform have been considered as less crowded.

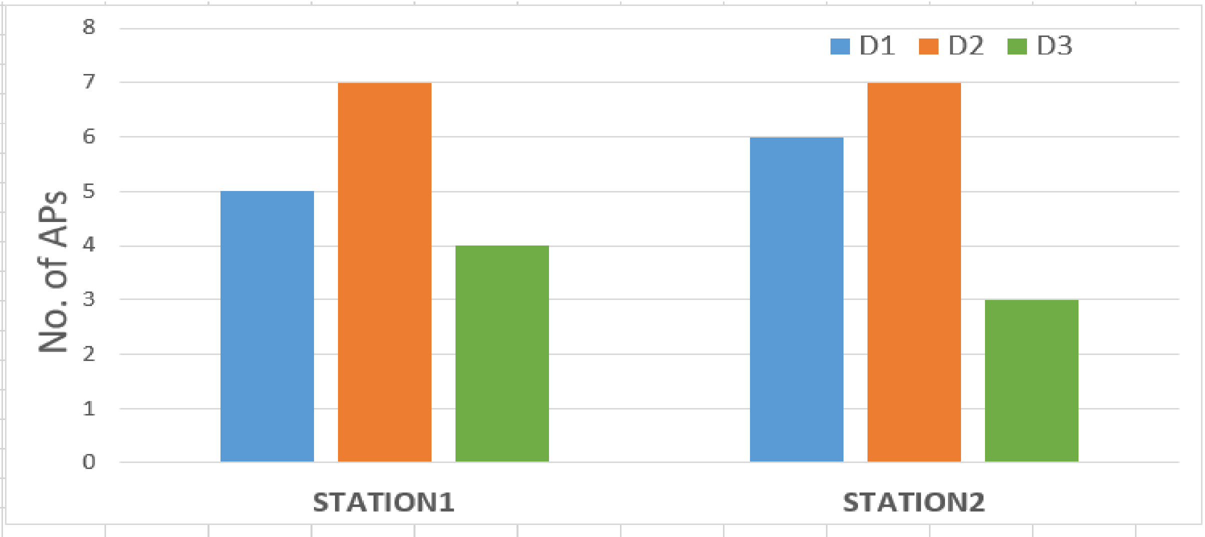


Fig. 5.9: Number of common APs when user travelled from metro station 1 to station 2; Di indicates i number of devices used in data collection.

RSSI values are affected by time, location points (LPs), and devices. Variation of RSSI values in different devices at different location points for different APs are shown in fig.5.5.

To have a better understanding of the problems and obstacles, the JU-MetroLoc data have been analysed for a variety of environmental factors. The RSSI values are impacted by a varying number of environmental factors, including crowded or less crowded locations and the presence or absence of trains. In figures 5.6, 5.7, and 5.8, how the RSSI values change as a result of various external factors have been shown. Fig.5.6 and 5.7 display RSSI value distortion caused by the presence of a train for some selected of APs and Figure 5.8 display the average RSSI values in crowded and less crowded areas.

The APs range depends on the strength of its device transmitter, receiving device, the nature of physical obstructions and radio interference in the surrounding area. Fig. 5.9 shows the number of common APs when the user travelled from metro station 1 to metro station 2.

5.2 JU- Accelerometer Dataset

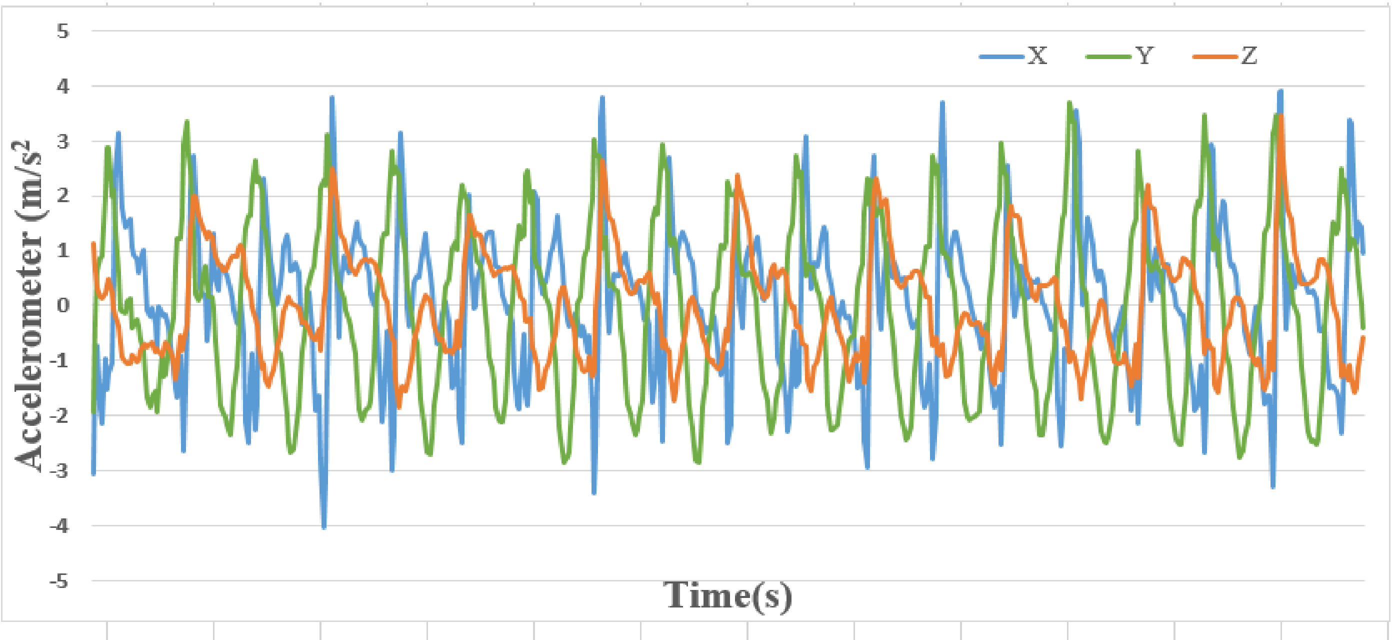


Fig. 5.10: Graphical plot for the smartphone tri-axial accelerometer data when user was talking on smartphone while walking.

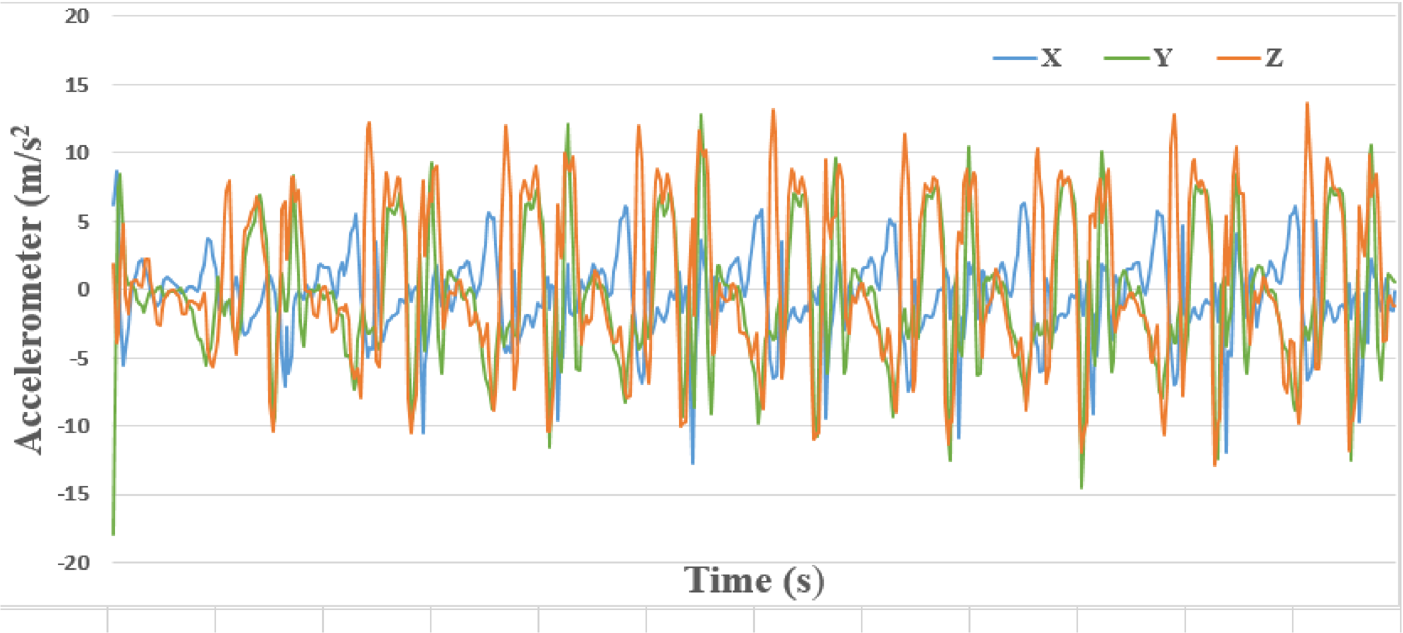


Fig. 5.11: Graphical plot for the smartphone tri-axial accelerometer data when a smartphone placed in user's trouser pocket while walking.

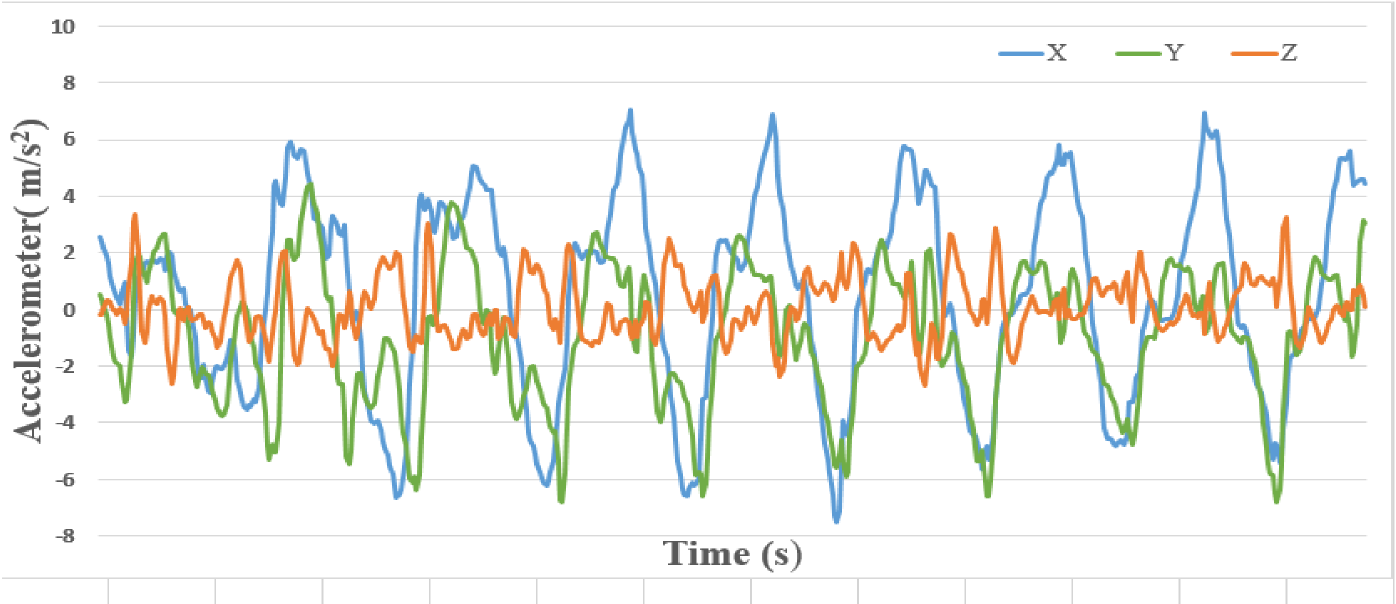
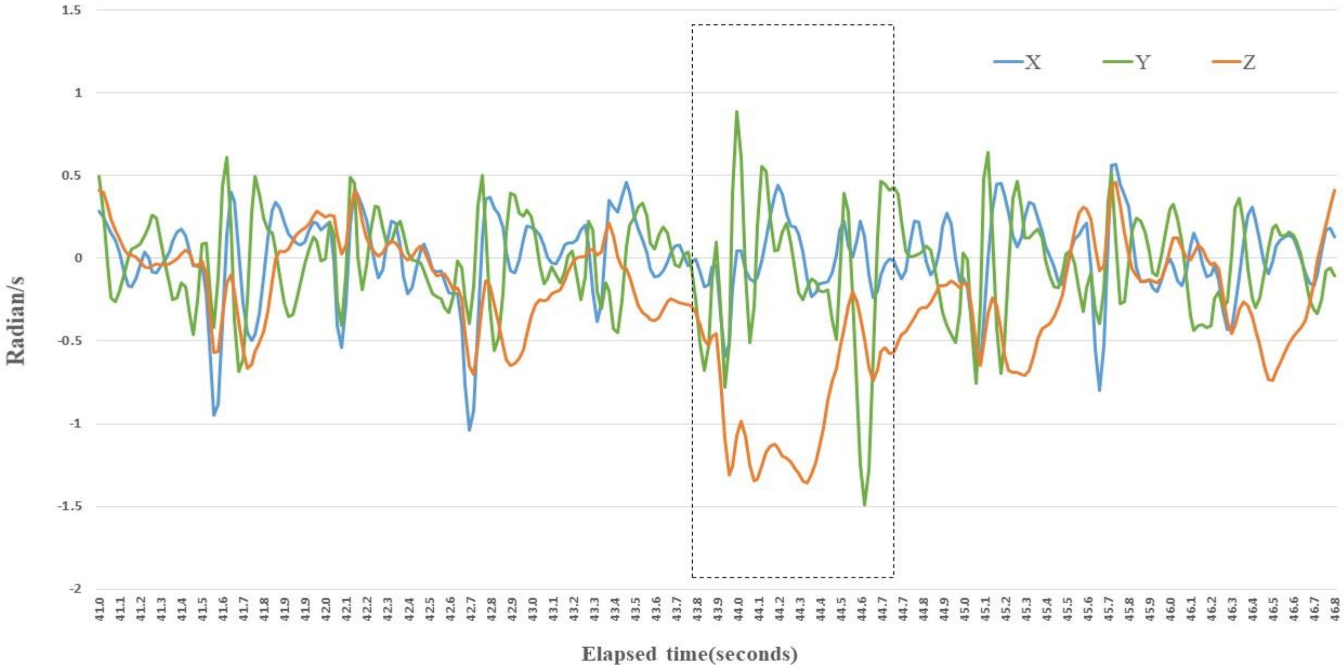


Fig. 5.12: Graphical plot for the smartphone tri-axial accelerometer data when smartphone was placed in user's hand and swinging his arms while walking.

From Fig. 5.10, 5.11, and 5.12 the dominant axis can be easily identified, like when smartphone was in hand and the arms are swinging, the dominant axis is X-axis. And when the smartphone was placed in user's pocket, Z-axis is the dominant axis. From these plots, one can determine in which position of the body the smartphone was placed. These details could be useful in some situations.

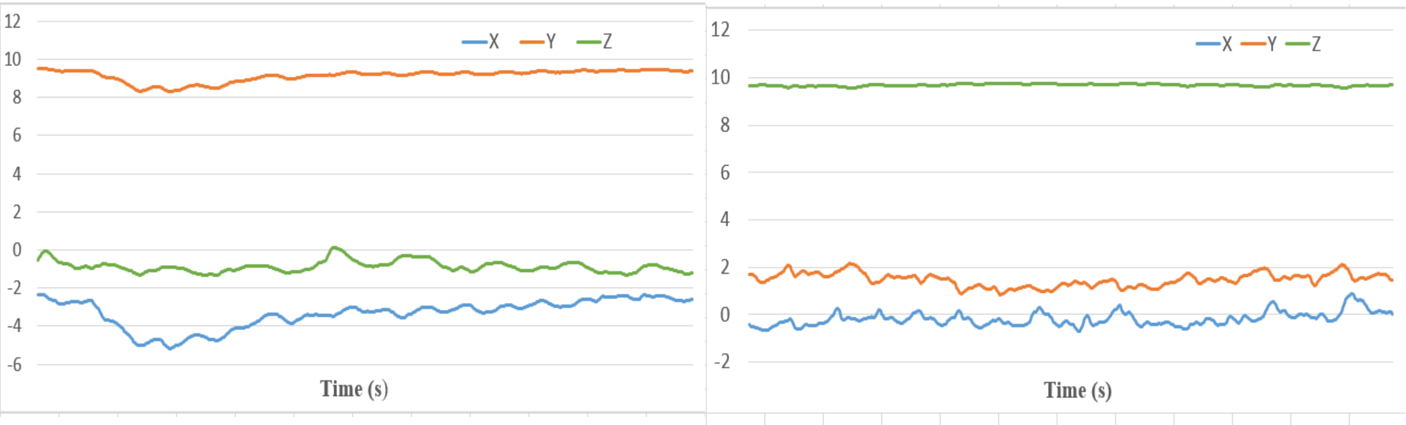


(a)

(b)

Fig. 5.13: Graphical plot for the smartphone tri-axial gyroscope data when (a) a right turn occurred (b) a U-turn occurred

The gyroscope data plot for the user, while the user was walking and normally using the phone is shown in Figure 5.13. From these plots the spike of Z axis can be easily identified. The dotted line in the (a) and (b) graph shows the Z-axis spike where the turn occurred. The first graph spike indicates there was a right turn occurred and second graph indicates a U-turn.



(a) (b)

Fig. 5.14: Gravity sensor data plot (a) Smartphone orientation was vertical (b) Smartphone orientation was horizontal while walking.

With the use of the data from the Gravity sensor, it is simple to establish the orientation of the phone (Fig. 5.14).

5.3 Baseline Accuracy

The results of the localization accuracy for the JU-MetroLoc dataset in different classifiers have been shown in this subsection. The maximum accuracy obtained from this dataset was 78.23% (Table 5.1).

Level of data granularity	Grid size(m)	Classifiers Accuracy (%)					
		GaussianNB	Decision Tree	kNN	SVM	Random Forest	Neural Network
Fine Granularity	4.94 × 1.36	62.57	76.99	78.23	75.92	78.73	75.76
Coarse Granularity	4.94 × 5.46	48.24	77.03	74.06	76.37	77.36	75.16

Table 5.1: Accuracy of Classifiers based on different granularity levels of metro station 1 dataset

Level of data granularity	Grid size(m)	Classifiers Accuracy (%)				
		Decision Tree	kNN	SVM	Random Forest	Neural Network
Median Granularity	5.86 × 2.58	46.45	48.54	45.83	47.9	48.12
Coarse Granularity	11.72 × 2.58	71.25	67.5	70	71.87	70.62

Table 5.2: Accuracy of Classifiers based on different Granularity levels of metro station 2 dataset

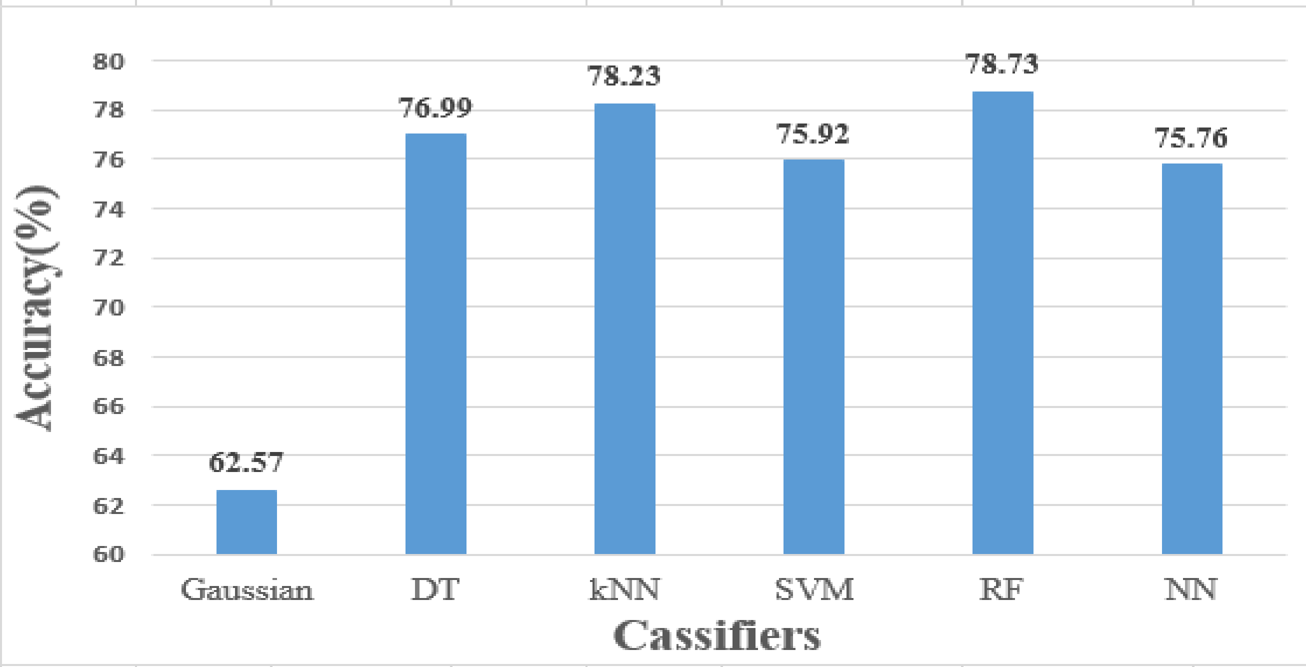


Fig. 5.15: Baseline accuracy results of different Classifiers for the JU-MetroLoc dataset of Metro station 1

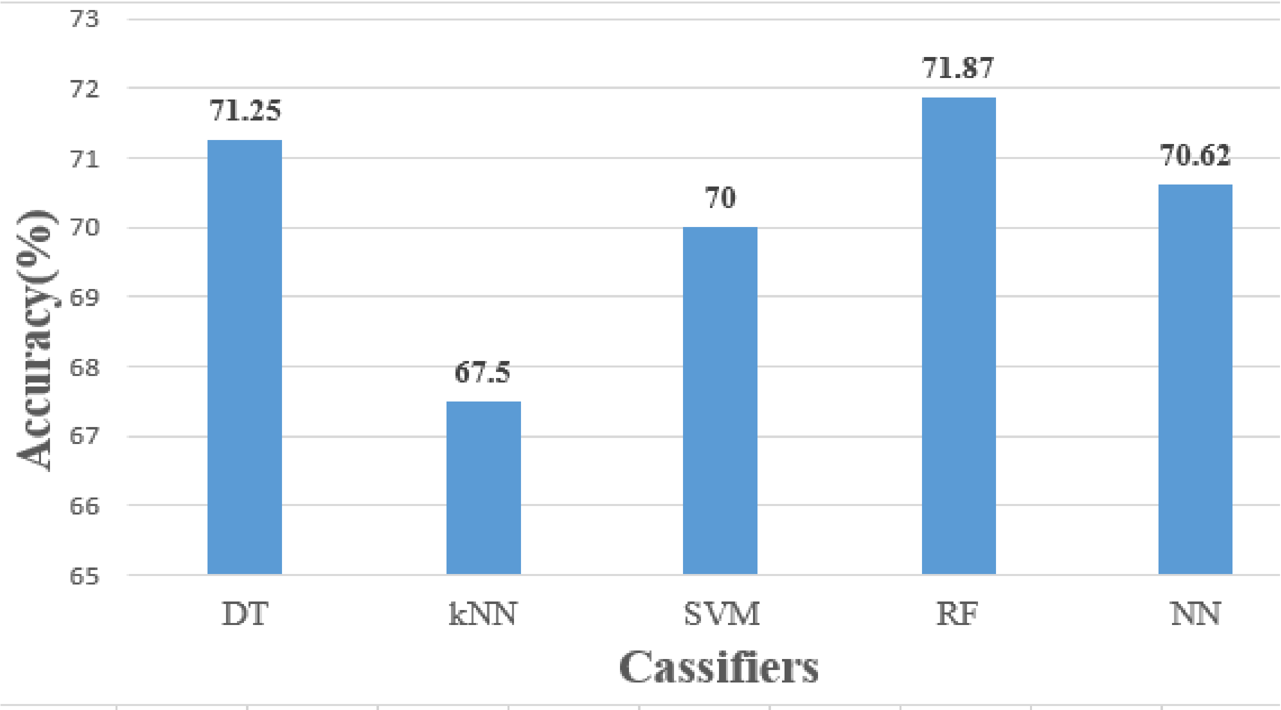


Fig. 5.16: Baseline accuracy results for the JU-MetroLoc dataset of Metro station 2.

The JU-MetroLoc dataset of metro station 1 has shown a maximum of 78.73% classification accuracy for Random Forest classifier (Fig. 5.15**)**. This result was obtained when dataset was split into 60 % for training and 40% for testing. And the JU-MetroLoc dataset of metro station 2 has shown a maximum of 71.87% classification accuracy for Random Forest classifier (Fig. 5.16**)** which is based on coarse-level Granularity.

Table 5.1 & 5.2 have shown the accuracy of datasets on different granularity levels for different classifiers. Here, the grid area where data was collected sets the level of granularity. The RF (random forest) classifier achieves better results compared to other classifiers for both datasets. However, accuracies of GaussianNB (Gaussian Naive Bayes) and SVM (Support-vector machine), kNN (k-nearest neighbors) and NN (Neural Network) vary between 62.57% to 78.73% and 67.5% to 71.87% for metro station 1 and metro station 2, respectively.

Chapter 6

Conclusion

In this thesis, two datasets for user localization by using Wi-Fi RSS sampling from smartphones and sensor data of smartphones have been designed. The JU-MetroLoc dataset based on Wi-Fi fingerprint was collected under several environmental factors at two different infrastructures in order to contribute in the process of user localization. Furthermore, our JU-Accelerometer dataset was designed to track the movements of the user. Mobile phone inertial sensors and Wi-Fi sensors provided the crucial data for predicting the user's location in each of these processes. Inertial sensors in mobile phones were used to track and guess where the next locations would be. Additionally, it can determine step detection and device orientation. And, based on the locations of Wi-Fi access points, mobile phone Wi-Fi sensors provided localization information of users. Our designed dataset can predict 46.45% to 78.73% location accuracy based on different levels of granularity.

Wi-Fi fingerprint-based localization also has several drawbacks of its own. It requires a lot of data support, and the preliminary work takes a lot of time. Since the method is modelled by the strength of each AP signal at a certain location, environmental factors also can affect it. So, data re-acquisition and re-designed are required whenever there is a change in either the location of the AP or the environment.

References

[1] F. Zafari, A. Gkelias and K. K. Leung, "A Survey of Indoor Localization Systems and Technologies," in IEEE Communications Surveys & Tutorials, vol. 21, no. 3, pp. 2568-2599, thirdquarter 2019, doi: 10.1109/COMST.2019.2911558.

[2] I. Qudah, P. Leijdekkers, and V. Gay, “Using mobile phones to improve medication compliance and awareness for cardiac patients,” in Proc. Int. Conf. Pervasive Technol. Related Assisted Environ., 2010, pp. 1–7.

[3] Lee, Dong-Gyu & Suk, Heung-Il & Park, Sung-Kee & Lee, Seong-Whan. (2015). Motion Influence Map for Unusual Human Activity Detection and Localization in Crowded Scenes. IEEE Transactions on Circuits and Systems for Video Technology. 25. 1-1. 10.1109/TCSVT.2015.2395752.

[4] K. Su, J. Li, and H. Fu, “Smart city and the applications,” in 2011 International Conference on Electronics, Communications and Control (ICECC), pp. 1028–1031, IEEE, 2011.

[5] D. Snoonian, “Smart buildings,” IEEE spectrum, vol. 40, no. 8, pp. 18–23, 2003], and surveillance [E. Asimakopoulou and N. Bessis, “Buildings and crowds: Forming smart cities for more effective disaster management,” in Fifth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), pp. 229–234, IEEE, 2011.

[6] Roy, Priya & Chowdhury, Chandreyee & Ghosh, Dip & Bandyopadhyay, Sanghamitra. (2019). JUIndoorLoc: A Ubiquitous Framework for Smartphone-Based Indoor Localization Subject to Context and Device Heterogeneity. Wireless Personal Communications. 106. 10.1007/s11277-019-06188-2.

[7]Castro P, Chiu P, Kremenek T, et al. A probabilistic room location service for wireless networked environments [J]. 2001.

[8] C. Wu, Z. Yang, Y. Liu, et al., WILL: wireless indoor localization without site survey [J]. IEEE Transactions on Parallel & Distributed Systems **24**(4), 839–848 (2013)

[9] M. Bshara, U. Orguner, F. Gustafsson, et al., Fingerprinting localization in wireless networks based on received-signal-strength measurements: a case study on WiMAX networks [J]. IEEE Transactions on Vehicular Technology **59**(1), 283–294 (2010)

[10] Haider Malik, Amir & Yiqiao, Wei & Liu, Shuzhi & Hwang, Seung-Hoon. (2019). Pre- and Post-Processing Algorithms with Deep Learning Classifier for Wi-Fi Fingerprint-Based Indoor Positioning. Electronics. 8. 195. 10.3390/electronics8020195.

[11] N.Wang, E. Ambikairajah, N. H. Novel, and B. G. Celler, ``Accelerometry based classification of walking patterns using time-frequency analysis,'' in *Proc. 29th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Lyon, France,Aug. 2007, pp. 4899_4902.

[12]J. Wannenburg and R. Malekian, "Physical Activity Recognition From Smartphone Accelerometer Data for User Context Awareness Sensing," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 47, no. 12, pp. 3142-3149, Dec. 2017, doi: 10.1109/TSMC.2016.2562509.

[13]Jiang, Jehn-Ruey & Subakti, Hanas & Liang, Hui-Sung. (2021). Fingerprint Feature Extraction for Indoor Localization. Sensors. 21. 5434. 10.3390/s21165434.

[14] K. -H. Chow, S. He, J. Tan and S. . -H. G. Chan, "Efficient Locality Classification for Indoor Fingerprint-Based Systems," in IEEE Transactions on Mobile Computing, vol. 18, no. 2, pp. 290-304, 1 Feb. 2019, doi: 10.1109/TMC.2018.2839112.

[15] Yassin, Ali & Nasser, Youssef & Awad, Mariette & Al-Dubai, Ahmed & Liu, Ran & Yuen, Chau & Raulefs, Ronald. (2016). Recent Advances in Indoor Localization: A Survey on Theoretical Approaches and Applications. IEEE Communications Surveys & Tutorials. PP. 1-1. 10.1109/COMST.2016.2632427.

[16] Xia, Shixiong & Liu, Yi & Yuan, Guan & Zhu, Mingjun & Wang, Zhaohui. (2017). Indoor Fingerprint Positioning Based on Wi-Fi: An Overview. ISPRS International Journal of Geo-Information. 6. 135. 10.3390/ijgi6050135.

[17] Wang, Jingjing & Park, Joongoo. (2021). An Enhanced Indoor Positioning Algorithm Based on Fingerprint Using Fine-Grained CSI and RSSI Measurements of IEEE 802.11n WLAN. Sensors. 21. 2769. 10.3390/s21082769.

[18] P. Zhang et al., "A Localization Database Establishment Method Based on Crowdsourcing Inertial Sensor Data and Quality Assessment Criteria," in IEEE Internet of Things Journal, vol. 5, no. 6, pp. 4764-4777, Dec. 2018, doi: 10.1109/JIOT.2018.2817599.

[19] J. Xue, J. Liu, M. Sheng, Y. Shi and J. Li, "A WiFi fingerprint based high-adaptability indoor localization via machine learning," in China Communications, vol. 17, no. 7, pp. 247-259, July 2020, doi: 10.23919/J.CC.2020.07.018.

[20] A. Yassin et al., "Recent Advances in Indoor Localization: A Survey on Theoretical Approaches and Applications," in IEEE Communications Surveys & Tutorials, vol. 19, no. 2, pp. 1327-1346, Secondquarter 2017, doi: 10.1109/COMST.2016.2632427.

[21] Wu, Chenshu & Yang, Zheng & Zhou, Zimu & Liu, Yunhao & Liu, Mingyan. (2013). DorFin: WiFi Fingerprint-based Localization Revisited.

[22] Yu, Xiao-min & Wang, Hui-qiang & Wu, Jin-qiu. (2020). A method of fingerprint indoor localization based on received signal strength difference by using compressive sensing. EURASIP Journal on Wireless Communications and Networking. 2020. 10.1186/s13638-020-01683-8.

[23] Apostolo, Guilherme & Sampaio, Igor & Viterbo, Jose. (2019). Feature selection on database optimization for Wi-Fi fingerprint indoor positioning. Procedia Computer Science. 159. 251-260. 10.1016/j.procs.2019.09.180.

[24] Ying, Julang & Pahlavan, Kaveh & Xu, Liyuan. (2019). Using Smartphone Sensors for Localization in BAN. 10.5772/intechopen.80472.

[25] F. Wang, J. Feng, Y. Zhao, X. Zhang, S. Zhang and J. Han, "Joint Activity Recognition and Indoor Localization With WiFi Fingerprints," in IEEE Access, vol. 7, pp. 80058-80068, 2019, doi: 10.1109/ACCESS.2019.2923743.

[26] Y. Chen and C. Shen, "Performance Analysis of Smartphone-Sensor Behavior for Human Activity Recognition," in IEEE Access, vol. 5, pp. 3095-3110, 2017, doi: 10.1109/ACCESS.2017.2676168.