Analysis of social interaction, crowd density and mobility pattern using smartphone sensing

by

Mehta Dhara Vipinkumar
202111054

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of

MASTER OF TECHNOLOGY
in
INFORMATION AND COMMUNICATION TECHNOLOGY
to

DHIRUBHAI AMBANI INSTITUTE OF INFORMATION AND COMMUNICATION TECHNOLOGY

July, 2023

Declaration

I hereby declare that

i) the thesis comprises of my original work towards the degree of Master of Technology in Information and Communication Technology at Dhirubhai Ambani Institute of Information and Communication Technology and has not been submitted elsewhere for a degree,

ii) due acknowledgment has been made in the text to all the reference material used.

D.V. Mehta

Mehta Dhara Vipinkumar

Certificate

This is to certify that the thesis work entitled "Analysis of social interaction, crowd density and mobility pattern using smartphone sensing" has been carried out by MEHTA DHARA VIPINKUMAR for the degree of Master of Technology in Information and Communication Technology at *Dhirubhai Ambani Institute of Information and Communication Technology* under my supervision.

Prof. P. S. Kalyan Sasidhar
Thesis Supervisor

Acknowledgments

I would start by showing my gratitude to my research supervisor and my mentor Prof. P. S. Kalyan Sasidhar, for his invaluable guidance and dedicated involvement in every part of my work. The way he conveyed the detailed information required for this research paperwork left a strong impression on me. Also, his humbleness and focused vision steered me in the right direction. He taught me a proper methodology for conducting research. I am grateful to him for providing me with a workspace and making me a part of lab-003 'UbiSense Lab', and also thanks a lot for providing me with a workstation to carry on my research work.

I would also like to thank Ph.D. scholar Ms. Maitri Vaghela for her thorough guidance. I would be glad to mention the contribution of all the volunteers who took part in the data collection and spared their time for this research.

Getting through my MTech curriculum and research required more than just academic support, and I have many friends and the entire batch of Mtech-2021 to thank for teaching and helping throughout the course of two years. I would appreciate friendship and patience in sharing their valuable expertise in the required data analysis. I also would like to thank my friends Kavan Hajare, Vyom Shah, Jaineel Trivedi, Khushali Ratanghayara, Jui Telvane, Janvi Thakkar and many others for sticking with me through thick and thin and making the entire journey a memorable one.

Above all, I would like to thank my parents and my whole family for constantly believing in me and pushing me throughout the process.

Contents

Abstract

The smartphone has become an important part of human life and with their integrated sensor ecosystem, they offer a practical platform for analyzing human behavioral patterns. Traditionally, studies done till now in this domain have employed dedicated hardware or regular questionnaires to study human behaviors. In order to better understand the psychology and behavior of college students, researchers have used the smartphone as a measurement tool. As college life can be a transition period for any student with respect to psychology and behavior. Using embedded sensors included in standard smartphones, we can leverage these capabilities to track social interactions and examine students' social context and network strength.

A significant rise has been observed in research interest in crowd behavior analysis and crowd density estimation, due to its crucial significance in ensuring the seamless management of events. Indoor localization helps people in navigating in indoor spaces. But the wealth of information regarding user's location inference is not being used in an efficient manner that is present with the servers of indoor localization architecture. So here we propose a method of extending indoor localization through Wi-Fi Connectivity information with the combination of smartphones. We implemented an in-house app, Usage Tracker, that will automatically collect data from its built-in sensors to analyze continuous sensing data from 35 students of our institute for 5 weeks. After that preprocessing and algorithms were applied to infer crowd estimation-related information and individual user-level social behavior analysis were carried out.

List of Figures

Chapter 1
Introduction

1.1 Era of smartphone's evolution

Smartphones have become an integral part of people's lives, they are now more accessible and equipped with a great variety of applications, making them an indispensable part of people's daily lives. Instead of only being speech and internet access devices, mobile phones are evolving into devices that can recognize a variety of user-related activities. Smartphones are in people's hands and on their minds everywhere they look. People's reliance on cell phones has grown exponentially in recent years due to the increased usage of these devices[25]. These phones have a multitude of applications that make life easier. In the last decade, smartphones have supplanted other platforms as the go-to devices for socializing, communication, entertainment, and many other activities. Many studies have already been conducted on how individuals use their smartphones, and a significant amount of data indicates that new digital devices are harmful to people's mental health[14]. In the last decade, the use of smartphones is rising as more and more teenagers discover their devices for relaxation. Excessive usage of mobile devices and smartphones is raising serious concerns due to the negative effects of such use on kids and teenagers. The effects of this technology appear to lead to adjustments in human behavior, such as rising levels of phone addiction, reduced physical activity such as walking, running, etc., more engagement online and less in-person, and shifting sleep patterns or less sleep hours[25][23]. Mental health includes emotional, psychological, and social well-being but all of these are not prioritized which affects their mood and may lead to stress and anxiety.

1.2 Smartphone Sensing Paradigm

Smartphones have become an essential element of the Internet of Things (IoT) paradigm because their integrated sensor ecosystem offers a practical platform for analyzing behavioral patterns in people. These devices have proven crucial in a number of IoT-related fields, including smart cities and smart health, by making it easier to monitor activities, identify different modes of transportation, and analyze mobility patterns, among other uses. Recent developments in sensor technology have enabled pervasive sensing, which has transformed many areas of our life. Notably, smartphones are equipped with a wide range of sensors, such as microphones, gyroscopes, barometers, light sensors, and proximity sensors. Using machine learning algorithms to infer multiple aspects of human behavior, such as physical activity, mobility patterns, location data, and social interactions, has been considerably easier due to this comprehensive sensor suite. These developments have accelerated the mobile sensing paradigm, enabling a thorough understanding of behavioral patterns in people. Smartphone sensors can be used to infer a variety of information about human behavior, including physical behavior, which includes movement patterns and activity levels, and social behavior, which includes the social dynamics of user and interpersonal interactions. Smartphone sensing capabilities provide valuable means to gain deep insights into social interactions, encompassing communication patterns and social networks.

1.3 Application of smartphone sensing

Smartphone sensing helps in keeping an eye on several health indicators of an individual, which can be either physical health or mental health. They help in tracking physical activity, sleeping patterns, excessive smartphone usage, and social behavioral patterns of an individual user on a daily basis [25][15][24]. Mental health parameters such as stress, depression, and anxiety scale of individual users can be inferred more accurately by correlating all mentioned behavioral measures with wellbeing scores [3][28]. While the heart rate monitor and GPS track heart rate, the accelerometer and gyroscope calculate steps, mobile unlocking patterns, and burned calories. All the existing wellness mobile application's core work includes results and analysis from the smartphone's sensed data[7][27][5].
Sensor present inside smartphone such as magnetometer provides three-dimensional magnetic field values and helps in developing navigation application. And makes it easier for users to find the directions inside the indoor spaces by detecting

the movement of the device and providing precise positioning data. The indoor localization-based system uses Wi-Fi and Bluetooth sensors present inside phones to estimate the crowd inside an indoor space by correlating proximity and location for the purpose of event management and urban planning [31][29].

1.4 General process of smartphone sensing

The method by which smartphones gather information from numerous sensors that are built into the device is referred to as "smartphone sensing." These sensors enable a variety of applications and capabilities by enabling smartphones to observe and engage with the real environment. Here is a general procedure for smartphone sensing:

1. **Sensor detection:** The operating system of a smartphone recognises and initialises the sensors that are installed in the device when it is turned on. Usually, a sensors which are found in smartphones include accelerometers, gyroscopes, magnetometers, proximity sensors, ambient light sensors, GPS receivers, barometers, and cameras.

2. **Sensor data acquisition:** When the sensors are discovered, the smartphone starts gathering data from them continuously. Each sort of sensor offers particular knowledge about an environment of the smartphone. For example, the accelerometer measures acceleration forces, the gyroscope measures angular velocity, the magnetometer measures magnetic fields, and the camera captures visual information.

3. **Data processing:** The operating system and other software elements of the smartphone process the sensor data that has been extracted. To obtain useful information, raw sensor data is often processed and filtered. This may involve applying algorithms for noise reduction, calibration, and data fusion, combining information from multiple sensors to increase accuracy.

4. **Contextual interpretation:** The smartphone's environment and user interactions are taken into account as the processed sensor data is interpreted. For instance, combining data from the accelerometer and gyroscope can help determine the phone's orientation like portrait or landscape, while GPS data can provide information related to location.

5. **Application integration:** The smartphone's applications get the access to the interpreted sensor data. Through the proper application programming

interfaces (APIs) that the operating system provides, app developers can access certain sensor data. This allows applications to utilize the sensor data to offer various functionalities like step counting, augmented reality, fitness tracking, navigation, gesture recognition, and more.

6. **User interaction:** Smartphone sensing frequently involves user engagement. Users can interact with the device through touchscreens, buttons, voice commands, and gestures, which are also captured by the sensors. Applications can leverage this interaction data to deliver intuitive and individualized user experiences.

1.5 Motivation and Problem Statement

1.5.1 Motivation

In the present era, college students have increasingly become immersed in their smartphones, devoting a substantial amount of time, approximately 6-7 hours per day, to these devices[22]. The predominant factor behind this extensive mobile usage is the prevalence of social media platforms, particularly Instagram and WhatsApp and frequent mobile unlocks were also observed[16]. Consequently, this excessive reliance on smartphones has had a profound impact on social engagement, reducing opportunities for face-to-face interactions and diminishing sociability. Regrettably, the prioritization of mental health, encompassing emotional, psychological, and social well-being, often takes a backseat, leading to detrimental effects on mood, heightened stress levels, and anxiety [1] [19][14].

1.5.2 Problem Statement

The problem addressed in this thesis is inferring social interaction and crowd density without relying on wearable equipment or computer vision systems. The objective is to develop a smartphone sensing application that utilizes the built-in sensors of a smartphone. By leveraging smartphones' capabilities, this solution provides a convenient and accessible method for measuring and understanding social interaction and crowd density, eliminating the need for additional devices or complex setups.

1.6 Contribution in thesis

In this study, we are doing an experimental study using our in-house application usage tracker, which will collect data from 35 students for 5 weeks from in-built smartphone sensors such as a microphone, Wi-Fi, and Bluetooth sensors. We observed the social interaction patterns, examining students' social circles and their discussion dynamics. Following is the list of Contributions:

- **Duration of social interactions:** The duration of social interactions of students on a daily basis will give us insights into their social behavior.
- **Frequent locations of social gatherings:** By knowing the frequent location where most of the students are socially engaged gives us an insight into the places of the institute that are mostly preferred by students.
- **Mobile usage during social interaction:** During social involvement, the amount of mobile usage gives us insights into their partial involvement in the interaction.
- **Small or Large group discussions:** This will help us to get insights about student's social network and their dynamics, based on involvement they prefer while interacting with others in form of either smaller or larger group.
- **Crowd Density/Crowd behavioral analysis:** In this, we find population density, identify congested regions, and crowd flow patterns in sub-areas, and detect peak usage patterns.

1.7 Organization of thesis

In this section, we will discuss the flow of the thesis organization.

Chapter 1(Introduction): This chapter briefly describes the evolution of smartphones in recent years, their rise as a smartphone-sensing paradigm, and the effects of it's excessive mobile usage. Then I added motivation and a problem statement for my thesis.

Chapter 2(Literature Survey): This chapter discusses the existing work where social behaviors were inferred using regular questionnaires, and sensor-based approaches as well. It also discusses the different problems with RFID, extra hardware infrastructure installation-based, and video surveillance-based crowd density estimation work carried out to find crowd density. Also, it includes the various indoor localization concept which helped us to understand the approach we could use for implementing our idea.

Chapter 3(System Design): This chapter gives a brief about the entire methodology which includes, system design, usage tracker app, data collection, and data prepossessing. It will briefly describe the entire data collection procedure and the followed system architecture.

Chapter 4(Algorithm): This chapter includes three different algorithms that were implemented on combinations of three different Wi-Fi, Bluetooth, and sensor datasets.

Chapter 5(Results): This chapter gives insights into the student's social interaction behavioral trends, frequent locations of student gatherings, student's social networks, and mobile usage while their social interaction.

Chapter 6(Crowd Density): Brief description of the need for crowd control and management in public spaces, indoor localization-based Wi-Fi infrastructure approach and results of crowd density, and stay duration at each Wi-Fi APs of our institute.

Chapter 7(Conclusion and Future Work): This chapter concludes the work done in my thesis and also discusses the future work that could be carried out to improve the work.

CHAPTER 2

Literature Survey

In this Section, we have explored Different methods: Survey based methods, Sensor-based methods, Bluetooth and RFID,Indoor Localization Systems and Camera based Detection for collecting data to identify human behaviour and understand the application of different sensors.

2.1 Survey Based methods

Griffioen, Nastasia et al.[14] used correlations and a comparison of the feeling scores on the data, which was collected using the questionnaires phase. 110 out of the 114 emerging adults in their sample (96.5%, n = 114) utilized their smartphones at some point. For instance, work in [2] utilizes mobile call logs, text messages, Wi-Fi and Bluetooth data, microphone and sensor data to related disorders, such as loneliness, depression, stress, etc. The [2]study examines the relationship between collaborative learning and students' interaction skills.A total of 100 secondary school students received the survey form at random. The findings indicate that students like group projects over solo ones.

2.2 Sensor based approaches

In [27], the Student life continuous sensing app assesses the day-to-day and week-by-week impact of workload on stress, social dynamics, sociability, mental well-being, and academic performance of a single class of 48 students across a 10-week term at Dartmouth College using Android phones. BeWell adopted a practical approach and develops a basic well-being model that may be expanded. A score between 0 to 100 was associated with their social well-being, calculated by taking the high and low amount of social interaction or face-to-face duration using linear regression but overestimates by 14% [7]. Work in [5] infers social behaviour through Bluetooth IDs, microphone activity detection, SMS, call logs, and step counts. Work in [5] infers social behaviour through Bluetooth IDs, microphone activity detection, SMS, call logs, and step counts. The application lets user to set their goals for their well-being. In this work, [6] the authors presents a study with 24 participants where they were asked to socialize with each other for 45 minutes. Develops a system for detecting stationary(Not moving) social interactions inside crowds using mobile sensing data such as Bluetooth Smart (BLE) and in total, 99 one-to-one interactions were observed with a mean duration of 254.9sec (±161.7) and 22 group interactions (i.e., interactions that include more than 1 two participants) with a mean duration of 117.2sec (±139.4).

Matic et al. [11] aimed to detect face-to-face social interactions on a small spatiotemporal scale relying on widely available sensing technologies. The approach recognizes both speech activity and spatial settings among subjects. Zhang et al. [33] proposes a solution based on mechanism not relying only on simply smartphone based sensors and uses bluetooth wearable tags, minimally invasive and low-cost. This solution is based on the analysis of the RSSI emitted by BLE beacon messages and received by the user personal device through a dedicated mobile app.

Minshu et al. [8] investigates aspects and activity patterns of human social interaction using location information and communication records obtained from mobile phones. Movement patterns are included in activity patterns, which pertain to social interaction aspects, which are the temporal and spatial interactive information.

2.3 Bluetooth and RFID based crowd density

Eight noninvasive bluetooth sensors were deployed on different floors throughout the museum to analyze visitors' behaviour and for estimating the crowd density [30]. And identified the length of stay at each node varies from 15 sec to 3 min, according to monuments.To collect crowd positioning information, each visitor was asked to wear a bracelet equipped with a proximity sensor that is available at the museum to operate at conditions of high crowd density [10]. Weppner j. et al. [29] Bluetooth scans to analyze social context and extends it with more advanced features, leveraging collaboration between close by devices and the use of relative features that do not directly depend on the absolute number of devices in the environment. and its shows over 75% recognition accuracy on seven discrete classes.

The work presents in [32] an extensive review of state-of-the-art advances in detecting abnormal behaviour in dense crowds. The techniques are based on range-free localization for detecting the direction and speed of the crowd movement by surveying Radio Frequency Identification (RFID) and Wireless Sensor Networks (WSN).

Versichele et al. [26] provides a case study of using Bluetooth sensing to track visitors at a large-scale event. Bluetooth scanners at 22 locations throughout the event area. Over the course of 10 days, the scanners detected over 100,000 devices, corresponding to approximately 80,000 visitors. It also highlights the advantages of Bluetooth sensing for event management and urban planning.Solmaz et al. [20] present a study on human mobility models are key components of various research fields, including transportation, mobile networks, disaster management, urban planning, and epidemic modelling.

Yuan et al. [32] proposed a novel method for estimating crowd density using wireless sensor networks (WSNs). The method uses the received signal strength (RSSI) of radio signals to estimate the number of people in a given area. The paper also discusses the potential applications of the method, such as crowd management, traffic control, and disaster response.

Hussein et al. [18] presents an extensive review of state-of-the-art advances in detecting abnormal behavior in dense crowds. The techniques are based on range-free localization for detecting the direction and speed of the crowd movement. Radio Frequency Identification (RFID) and Wireless Sensor Networks (WSN) are surveyed.

Mohandes et al. [12] discusses the use of RFID technology to improve the efficiency and safety of the Hajj pilgrimage. A prototype system was developed and tested with a group of 1000 pilgrims from the Ivory Coast. The results of the pilot study showed that the system was effective in identifying and tracking pilgrims, and could be used to improve crowd control and safety during the Hajj pilgrimage.

2.4 Indoor Localization System based crowd density

Tang et al. [21] proposed method for estimating crowd density by anonymous, non-participatory, indoor Wi-Fi localization of smart phones.The approach is predicated on the idea that estimating the population density in a specific area may be done by counting the number of Wi-Fi probes produced by a smartphone. The study's findings demonstrate that the suggested method can precisely estimate crowd density in a range of interior settings, including lecture halls, meeting spaces, and retail centers. Yuan Y, et al. [31] implemented two phase iterative process,consists of crowd detection phase and calibration step works k-mean algorithm and noise elimination, to perform experiments using 16 sensor nodes and large scale simulations.It divides the level according to RSSI values received from WSN.

Georgievska S. et al. [4] proposed a method to estimate crowd by anonymous,non-participatory and indoor Wi-Fi localization of smartphones and uses big data analytics and probabilistic models to overcome the ambiguity of Wi-Fi localization. Pie et al. [17] approached crowd sensing is a promising approach for indoor localization using Wi-Fi access points (AP) and fingerprinting methods. With that, they also used the signal strength of multiple access points to identify the 2-meter range accuracy.

Nakatsuka et al.[13] proposed RSSI and LQI-based systems to develop a popular wireless network. They create a prototype system that integrates visual marker detection to pinpoint the location, Web services to gather and share sensor data, and augmented reality visualization to improve the estimated crowdedness.

2.5 Camera based Detection

Onkar S. et al. [9] provided a technique for wide-area surveillance applications that uses video to estimate crowd density. The device can effectively estimate the precise number of people and crowd velocity from crowded regions. The technique first extracts moving objects from the movie using a foreground-based technique. After that, the moving objects are categorized as either cars or people using a feature-based approach.

Zhang et al. [34] suggested a method for estimating and mapping crowd density. The technique combines crowd semantic segmentation, crowd denoising, and BP neural network to accomplish high-precision crowd extraction and density estimate in large scenes. The findings demonstrate that the approach can reliably predict crowd density in large scenarios.

Ligun et al. [33] presented a video-based crowd density analysis and prediction system for wide-area surveillance. The Accumulated Mosaic picture Difference (AMID) approach was used to get estimate of a crowd's density several minutes in advance by using a multi-camera network. Numerous experiments carried out in actual settings (such as a station, park, or mall) show the usefulness and resilience of the suggested strategy. The system has been employed in practical applications.

The literature has given us useful insights into current concepts and methods for estimating crowd density and social interaction. Various studies have explored the use of smartphone sensors and also referred to daily questionnaires for analyzing the social behavior of users. Additionally, approaches utilizing RFID technology, camera-based detection, and indoor localization systems have been explored. But we can not rely on survey-based and complex hardware-related methods. However, a comprehensive and infrastructure-free solution that leverages the capabilities of smartphones for accurately measuring social interaction and crowd density is still needed. Building upon the knowledge gained from the literature, this thesis proposes a novel system design that utilizes a custom smartphone application called "Usage Tracker."

CHAPTER 3
Methodology

This chapter will briefly discuss the entire methodology which includes, system design, data collection, and data processing. This chapter will also give insights into the system architecture that we have followed for the thesis.

3.1 System Design

Due to the lack of an all-in-one app that could sense sensor data and mobile usage, we created our own mobile application, called "Usage Tracker," that works on smartphones with the Android operating system. Through built-in sensors inside the smartphone, our application gathers data on the phone unlock, accelerometer, microphone, light sensor, Bluetooth, and Wi-Fi. Our application is displayed in Fig 3.1. The "Usage Tracker" app's user interface has been kept basic and straightforward. The application will automatically function in the background. It is not required to track the process or the scanning of the data. The user only needs to enable "Record activity" on the toggle for the app to be running in the background.

The "Usage Tracker" app keeps the user experience smooth by reducing unnecessary notifications and avoiding unwanted interruptions. The "Usage Tracker" application carefully captures sensor data in real time and effectively maintains the value or information for each sensor in a CSV file, ensuring precise and structured entries. Users have the option to manually export data at any time via the interface, either as a CSV file or a raw database (DB) file. Since the data was produced by sensors, it was guaranteed to be accurate and free from any potential manipulation by users.

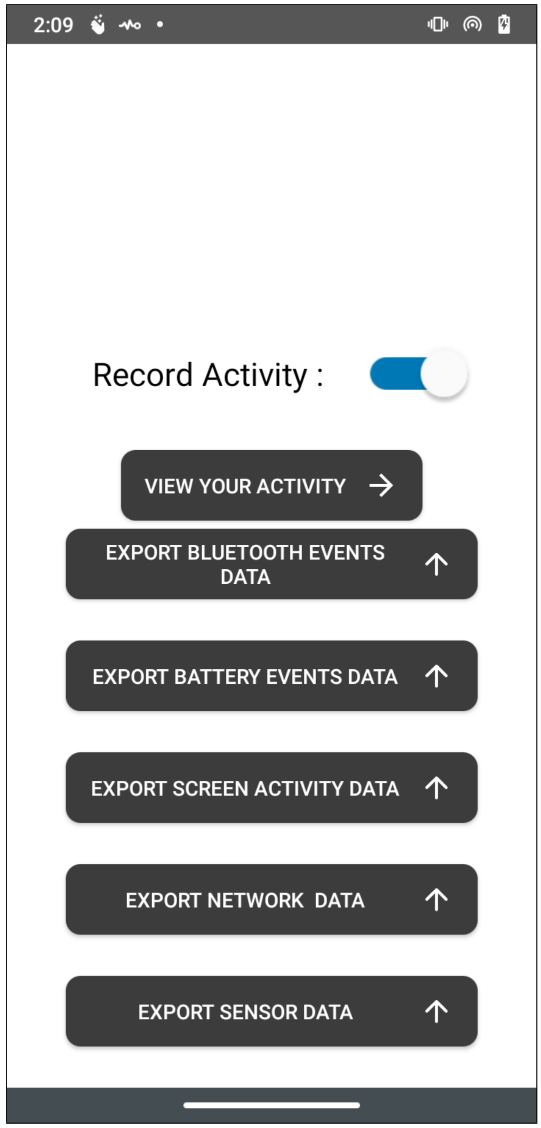


Figure 3.1: Usage Tracker App

Since there wasn't a third-party application that offered all the necessary information, it had to be tried, tested, and secured. The 0.5 Hz rate of sampling was chosen. We installed our software as an .APK file, which is the file format for applications used on the Android operating system, on each participant's phone after explaining the data-gathering method and study process.

We have beta tested the app on many Android devices to ensure that it runs successfully on various versions from Android 6.0 to the latest Android 12.0. It also shows compatibility with all of the specified versions of Android.

3.2 System Architecture

In the system architecture shown in Fig 3.2, the usage tracker app plays a central role in collecting continuous sensing data from various inbuilt sensors available on the users' devices. These sensors include the accelerometer, gyroscope, Bluetooth, microphone, and Wi-Fi. The app continuously gathers data from these sensors, capturing information about users, and collected sensor data is used to capture physical activities, device interactions, Bluetooth interactions, ambient sound, and Wi-Fi connectivity.

All the collected data is stored in a CSV (Comma-Separated Values) format. Storing the data in this format allows for easy organization, and analysis using various data processing tools and techniques.

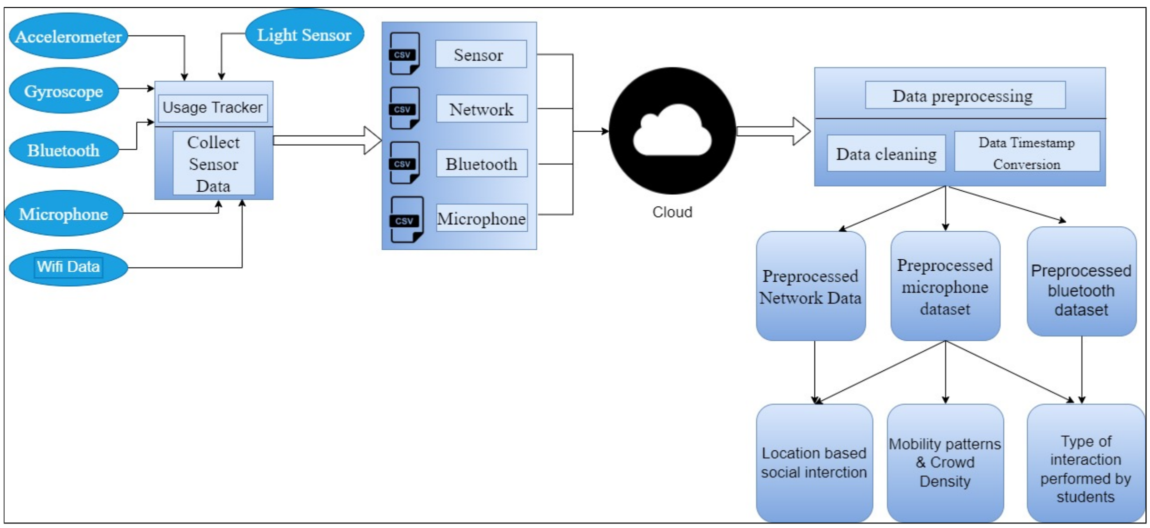


Figure 3.2: System Architecture

After applying the preprocessing techniques to the entire datasets, we proceeded with the application of various algorithms and scripts to extract meaningful insights and identify behavioral patterns among the students, which will be covered in upcoming chapters. By combining the three datasets, we were able to derive valuable insights that shed light on different aspects of student behavior. Such as the sociability level of students, the frequent location where most of the students gather, sociability patterns, the type of interaction performed by students be it single or group discussions, and their crowd density and mobility-based patterns.

3.3 Data Collection

The data collection exercise in our study targeted students who were pursuing undergraduate and postgraduate programs. The total population consisted of 35 students, with 20 men and 15 women. Out of these participants, 22 were enrolled in undergraduate programs, while 13 were pursuing postgraduate studies. This approach enabled us to include students from a wide range of academic backgrounds, ensuring diversity in our study population. To ensure the voluntary participation of students, we advertised the opportunity to take part in the study. Interested individuals were then required to provide their consent to share app usage and sensor data from their mobile phones. By obtaining informed consent, we respected the privacy of the participants, as they had the freedom to decide whether they wanted to contribute their data to the study. Once the participants were recruited, we provided them with a detailed explanation of the app's functionality and purpose. This included information on the installation process, data storage, and exporting format. It was essential for the students to have a clear understanding of how the app would collect and handle their data to ensure transparency and build trust in the research process. The average age of the participants was determined to be 23 years, with a standard deviation of 3 years (±3). This age range suggests that the majority of the participants fell within a relatively young and similar age group, which could potentially minimize the impact of age-related variations on the study findings.

3.4 Dataset

The data we collected through the usage tracker app, was stored in google drive and the dataset contains the following different sensed data during the experimental period.

- **Wi-Fi DataSet:** The Wi-Fi data-set gives an idea regarding at one given timestamp students are present at which Wi-Fi APs of the institute along with its IP address throughout the day using Wi-Fi sensor.

- **Sensor DataSet:**The Sensor data set comprises three types of sensor data collected throughout each day that is an accelerometer, microphone, and mobile unlock count.

- **Bluetooth Dataset:**The Bluetooth dataset contains information about discoverable devices that are detected using the Bluetooth. It records the presence of other Bluetooth devices within range of the students' devices throughout at the frequency of 5 sec.

The data was collected at one sample every five mins. We received an average of 20 hr/day of data from each of the 40 participants. We collected a total of 3,36,357 data samples which contain each student's WiFi connectivity throughout the day. The below shows the sample dataset record in Fig 3.3. The table shows the date and time at which a particular student id's mobile phone is connected to WiFi APs across the campus.

Date	HH:MM:SS	ID	Ip address	Wifi Id
27-09-2018 00:00	8:33:06	2	10.200.5.231	"Hostel"
27-09-2018 00:00	8:38:07	2	10.200.5.231	"Hostel"
27-09-2018 00:00	8:43:25	2	10.200.5.231	"Hostel"
27-09-2018 00:00	8:48:32	2	10.200.5.231	"Hostel_5GHZ"
27-09-2018 00:00	8:53:55	2	10.200.38.65	"Hostel_5GHZ"
27-09-2018 00:00	8:58:19	2	10.200.38.65	"Canteen"
27-09-2018 00:00	9:03:19	2	10.200.38.65	"Canteen"
27-09-2018 00:00	9:08:20	2	10.200.38.65	"Canteen"
27-09-2018 00:00	9:13:19	2	10.200.38.66	"Event"
27-09-2018 00:00	9:18:21	2	10.200.38.67	"Event"
27-09-2018 00:00	9:23:47	2	10.200.38.67	"Event"
27-09-2018 00:00	9:28:21	2	10.200.38.67	"Event"
27-09-2018 00:00	9:33:30	2	10.200.38.65	"Event"
27-09-2018 00:00	9:38:30	2	10.200.38.65	"Event"
27-09-2018 00:00	9:43:30	2	10.200.38.65	"Event"

Figure 3.3: Sample WiFi Data Records

The data shows the date, timestamp, i/p address, and the ID of the WiFi access points at the institute.

3.5 Data preprocessing

Data preprocessing plays a crucial role in data analysis as it involves transforming raw data into a format suitable for further analysis. It is an essential step in the data analysis pipeline that helps ensure the quality, consistency, and reliability of the data, leading to more accurate and meaningful results. Data aggregation was carried out to combine each 40 students' separate 35 days' data files into one combined CSV different for each three datasets sensor, WiFi and Bluetooth to have a combined dataset. Data preprocessing was conducted to filter out invalid data and transform the data into desired forms. We collected a total of 3,36,357 WiFi interaction data samples which contain each student's WiFi connectivity throughout the day each at a 5-minute time interval. Those WiFi ids could be another person's mobile or laptop hotspot or the WiFi Access Points of the institute. The data is filtered to consider only the major Wifi APs such that all the different locations are covered. We get a total of 3,21,457 records after the filteration as shown in Fig 3.3

3.5.1 Data Timestamp Conversion

As we are collecting continuous sensor data from our app, each dataset has a timestamp column. In the timestamp column, data was present in HH:MM: SS 12-hour format without specifying AM/PM. So converting the timestamp into a 24-hour format to make it analysis compatible was really important to make data more readable and understandable the way data changes with time.

Applying timestamp conversion to each different data file such as Wi-Fi, Bluetooth data, or sensors from the dataset was really important as by using a standardized timestamp format, it becomes easier to perform sorting, grouping, and comparison operations based on time. Converting the timestamp from a 12-hour format without AM/PM to a 24-hour format improves the compatibility, readability, and consistency of the data, enabling effective time-based analysis and interpretation. It ensures that all the data points have a standardized timestamp format, facilitating seamless integration and analysis of multiple datasets.

Python's pandas and numpy libraries function helped to carry out data preprocessing. Some of those functions are pivot_table, replace, sort_values, merge, unique, random, describe, reshape, concatenate, etc.

CHAPTER 4
Algorithm

This chapter includes three different algorithms that were implemented on a combination of different datasets that were available with us to derive important insights.

4.1 Social Interactions

To estimate the duration of social interaction, we utilize microphone sensor data collected from the microphone that captures ambient sounds, including both vocal and non-vocal components. By analyzing the recorded data, which includes both noisy vocal data and quiet non-vocal data, we can determine the duration of the student's social engagement. To determine the total hours of social interaction, interaction during lecture timings needs to be excluded. To find the places where students are socially engaged frequently, we require both vocal microphone sensor data as well as Wi-Fi data. Fig 4.1 shows the overall flow and the algorithm will give a detailed idea about it. This work focuses on the parallel analysis of row-by-row data from Wi-Fi and microphone sources. By synchronizing the data using a shared timestamp, we aim to identify the location where a student is engaged in vocal interactions at a specific time.

Input to algorithm 1, microphone sensor data csv and network data csv for each student for same day. Here we want to find the frequent preferred locations and the duration of social interaction performed in the different areas of institute by each student on single day.

Algorithm 1 will iterate through each row of sensor data csv and it will check whether sensor data's microphone label is vocal or environmental noise.If row data found some interactions then at that moment we have to check each Wi-Fi data entry till the time of detected social interaction is present in between the start_time and end_time of particular location's duration in network data. Here we find the respective location from the network data where vocal interaction has

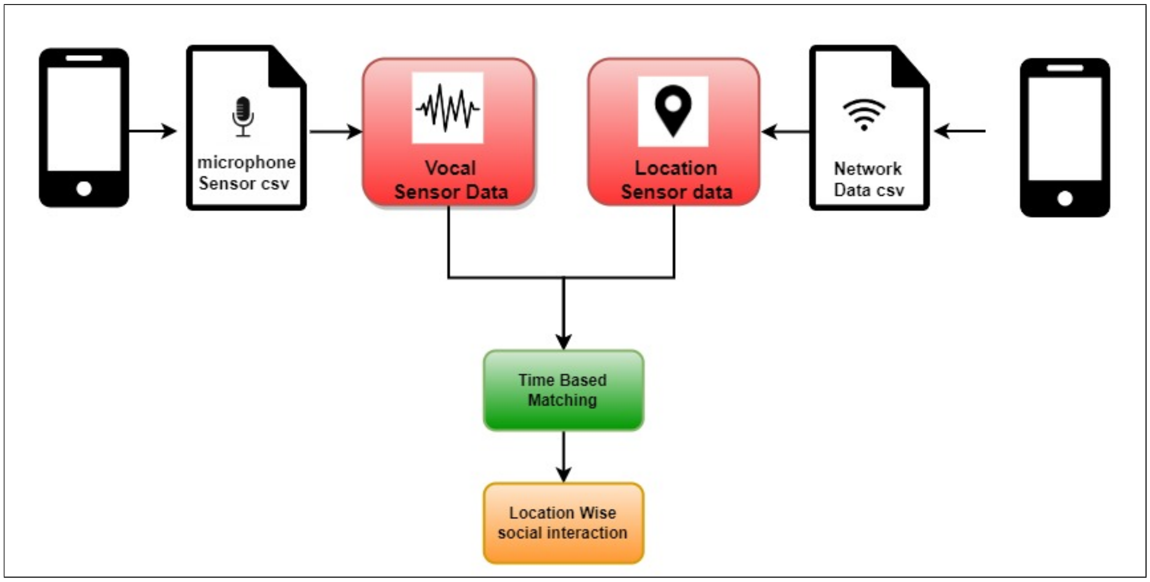


Figure 4.1: Location wise Social Interaction

been found. Here row_loc['Wi-Fi AP'] gives us the name of particular Wi-Fi AP and row_sensor['time'] shows the timestamp of sensor data-entry. We store the row_loc['Wi-Fi AP'] and row_sensor['time'] in respective two attributes location and duration. They will be stored in the result_df as two new attributes,and in each iteration they will get appended.

Algorithm 1: Vocal Detection

```
input  : Two data files Sensor and Location as CSV
output: location wise social interaction
for row_sensor ∈ sensor_data_csv do
    for row_location ∈ n/w_data_csv do
        if Microphone_Label == "VOCAL" then
            if row_loc['start_time'] ≤ row_sensor['time'] ≤ row['end_time'] then
                time = row_sensor['time']
                location = row_loc['Wi-Fi AP']
                result_df['duration'] = time
                result_df['location'] = location
        location_data = result_df.groupby['location'](['duration'].sum())
```

After all the iterations,result_df will be group the data by location and will calculate the total interactions performed by students at each location. This data helps us into getting insights for all students for entire data experiment duration and get insights into preferred location where students involved in social engagement.

4.2 Small group or large group interaction involvement

Increasing levels of phone addiction brings more social engagement online or through social media applications, and less interpersonal or face-to-face conversations. As we examined the duration of students' social involvement in the previous section, another important aspect to better understand students' social behavior is exploring campus-level social networks. A student's social circle and level of interpersonal involvement can be determined by counting the number of other students or friends they interact with on a regular basis.

Another social behavior of a student is the type of conversation, the student is more interested in smaller group interaction or larger group interaction. While evaluating a student's social behavior, it is crucial to understand the kind of discussion they prefer, whether in a large group or small group. Small group discussions refers to 1-4 students, and large group discussions involves 5 or more students.

To achieve this, we can analyze both the sensor data and the Bluetooth scans data. The sensor data contains information about whether a detached conversation is vocal or non-vocal. Bluetooth scans data provides timestamp-wise entries for all detected devices within range. The flow diagram of this procedure is shown here in Fig 4.2 By determining timestamp-based matching we can find the number of unique students/peers with a particular student communicates. Social interaction detection is presented in algorithm 2.

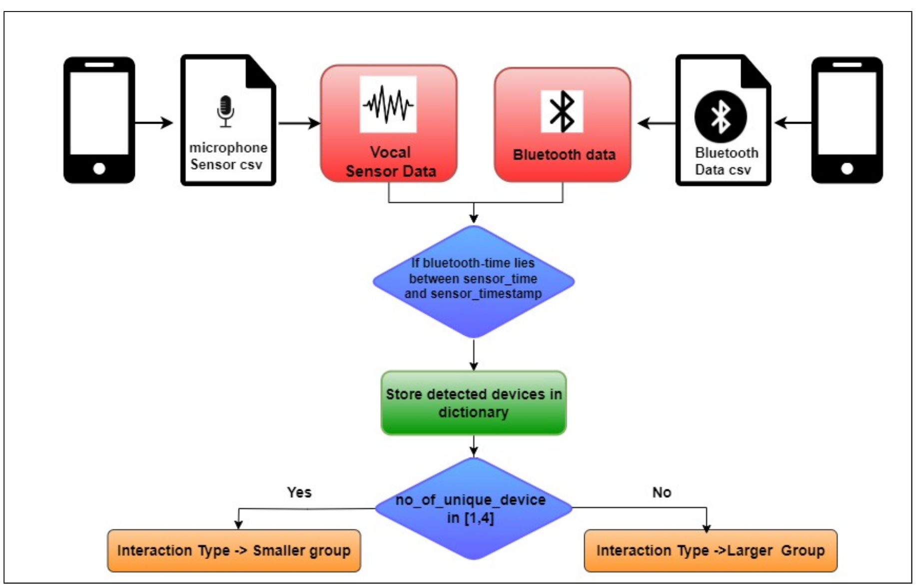


Figure 4.2: Interaction Detection Diagram

Algorithm 2: Interaction Detection

```
input  : bluetooth and sensor data csv files
output: interaction: small group or large group
Define variable called detected_devices.
for row_sensor ∈ sensor_data do
    if Microphone_Label == "VOCAL" then
        for row_bt ∈ bluetooth_data do
            if row_bt['time'] in row_sensor['timestamp','timestamp_interval']
             then
                detected_devices = row_bt(['time'],row_bt['device'])
        no_of_devices=len(detected_devices[1])
        if no_of_devices ∈ [1,4] then
            interaction ← small group
        else
            interaction ← large group
```

Input to algorithm 2 will be two csv files, microphone sensor data csv and bluetooth data csv for each student for same day. Here we want to find number of people involved in each detected vocal interaction. Algorithm will iterate through each row of sensor data csv and it will check if that sensor data's Microphone_Label has value vocal or quiet,if Microphone_Label is Vocal then it will iterate through row by row bluetooth data csv,until time of bluetooth scan fall in duration of that vocal interaction.Then it will store time and all the unique devices of bluetooth scan into a dictionary called detected_devices. To find the number of communicating devices length of unique detected device was measured and then based on the length if it fell within the range of 1-4, the interaction was termed to be representative of a small group. Conversely, if the number of detected devices exceeded this range, the interaction was termed as large group.

This way we found the number of people and type of discussion involvement preferred by all the students on daily basis during data experiment period which provide insights into their social dynamics.

4.3 Mobile Usage With Social Interaction

The frequency of mobile device usage during social interactions is a significant aspect to investigate, as it gives insights into the impact of mobile technology on interpersonal communications and social dynamics. It is essential to assess the extent to which students utilize their mobile devices while engaging with others.

Gaining insights into this behavior can provide a valuable understanding of the way mobile technology influences interpersonal communication and social dynamics.

We can determine mobile usage during social engagement by calculating general mobile usage and duration of unlocks that are happening during vocal interactions. We have the csv file containing sensor data comprised of mobile unlocks events and vocal annotations. The algorithm 3 will give a detailed idea about our approach to measure the level of mobile device usage specifically during social engagement.

We have sensor data csv consisting of mobile unlock events and vocal data, and want to determine amount of mobile usage during the social interaction to identify number of time a student unlock his/her smartphone and duration of mobile usage during that time.

Algorithm 3: Mobile usage while interaction

```
input : sensor_data csv contains unlock counts and vocal data
output: mobile usage data during vocal interaction
for row_sensor ∈ sensor_data do
    if row_sensor['microphone_label'] == "VOCAL" then
        if row_sensor['phone_unlock_count'] !=
          next_row_sensor['phone_unlock_count'] then
            phone_unlock_count=phone_unlock_count+1
        Mobile_Usage ← Calculate time duration of phone_unlock_count
```

In algorithm 3, while iterating through csv file the first thing we check is that microphone_label is vocal or not,if vocal interaction has been captured then in the same file check if the phone_unlocks is changing or not with respect to the next row.If phone_unlocks is not same then phone_unlocks must have changed which shows the unlock of phone and then based on that we will be able to get the count of unlock_events and calculate the time duration of the interaction to find mobile usage.The resulting output includes the count of vocal events and the count of unlock changes specifically related to the vocal label.

CHAPTER 5

Results and Discussion

5.1 Social behavioural trend

An average interaction hour of all students was calculated on each day and presented as a graph, the plot of the graph is fluctuating which represents distinct behaviors of students on individual days. To observe their social behavioral trend, the average interpersonal communication of all students on each day is plotted in the following Fig 5.1. The average social interaction hour was around 4.8 ± 1.9.

On the range of days (out of 30) 5-8, 12-15,19-20, and 26-29 there was moderate interaction duration observed ranging from 3.2 - 3.9 hours approx due to students having lectures, assignments, deadlines, etc on weekdays. In our experimental period, students were having examinations from days 6-8 around so 1.9 to 2.1 hours of average interaction was observed which was considerably less than any other day due to the busy study schedule highlighted using green labels. During the weekend, students were found having approx. 4.1 to 5 hours of interaction/discussion which shows students were more socially active with their friends, partying, or going out. Exams affected the weekend due to which comparatively less interaction was observed from days 2-4. From the graph, red labels were observed on days 16-18 for most social interactions, students gathering were found around 5.8 to 6.1 hours due to the technical fest organized at the institute.

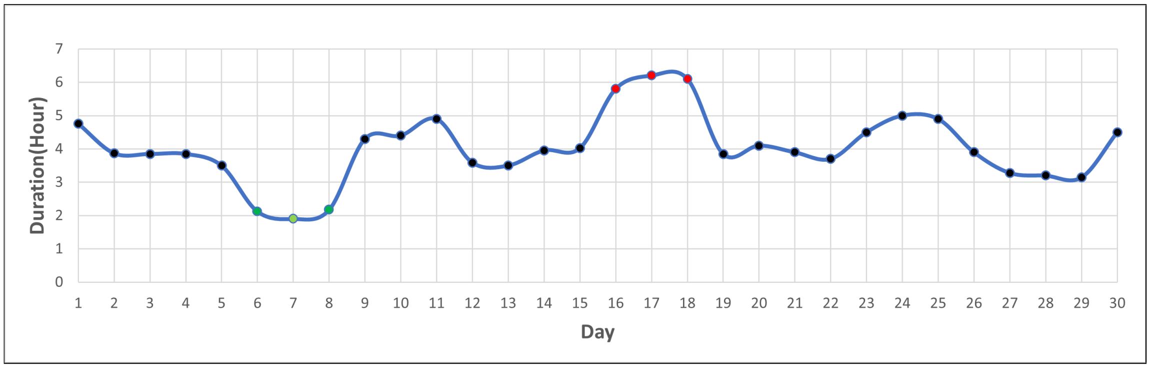


Figure 5.1: Social Behavioural Trend

5.2 Vocal Interaction at various Wi-Fi APs

This result shows the range of social interactions that happened at different locations of our institute. By visualizing the box plot, we observed the social interaction duration distribution over various locations. Students socially engaged in the canteen for around 22 minutes to 85 minutes 50% of the total interaction at the canteen happened around 25-30 mins and approx. 75% of the interaction that happened in the canteen lasted for 45 mins. And in a hostel, 25% of the social interaction happens for around 60 mins, and 50% of the total interaction that happened at the hostel lasts for 1.9 hours. The social interactions among students that happens at the canteen last sometimes approx. 3 hours - 4.5 hours which was observed on the tech-fest in our institute. At the library, students are comparatively less socialized, and 75% of social interaction duration noticed was for approx. 27-28 minutes. At the lecture hall, many time social interaction were observed that were due to classroom interaction, and sometimes outliers for around 2-3 hours was observed due to organized workshop. And 50% of interaction was observed at the hostel, which was approximately 1.5 times more than the total social engagement observed among students at a different location. This shows us that student most of the time gathers at a hostel, then the canteen, followed by lecture halls and lab buildings.

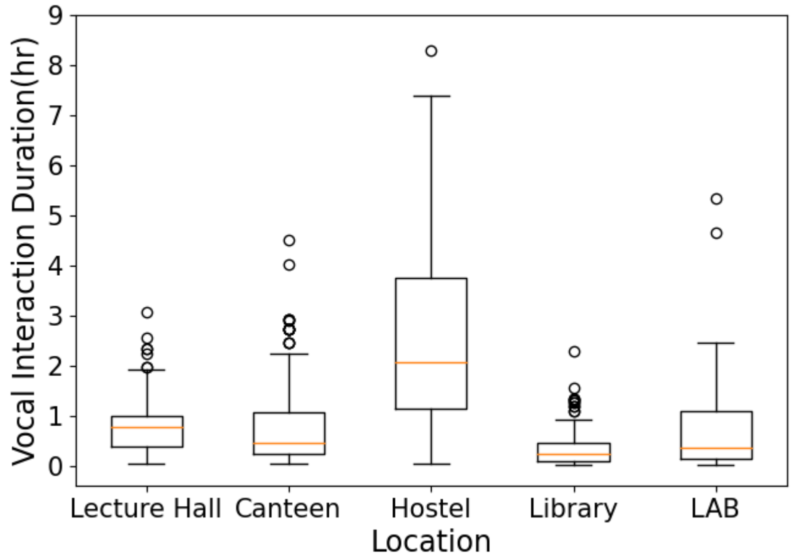


Figure 5.2: Location-wise social interaction

5.3 Small or large group discussions

Identifying the kind of interaction/discussions preferred by students will give insights into their level of sociability. Fig 5.3, shows the avg amount of small group and large group interaction(%) performed by 35 students during their data experiment period. From the graph, we can infer that almost about 70% of students were more involved in small group discussions in comparison with large group discussions during the experimental period which shows that they prefer interaction with less no. of people rather big groups. There are 20% of the students also interested more in large group interaction rather than small group which gives us insights that students are more interested in large group discussions and activities. And remaining 10% would prefer both smaller and larger group interaction equally. This shows the diversity of students' behavior in their social circle. The preference for smaller or larger group discussions among students holds significant implications for their social development and overall well-being.

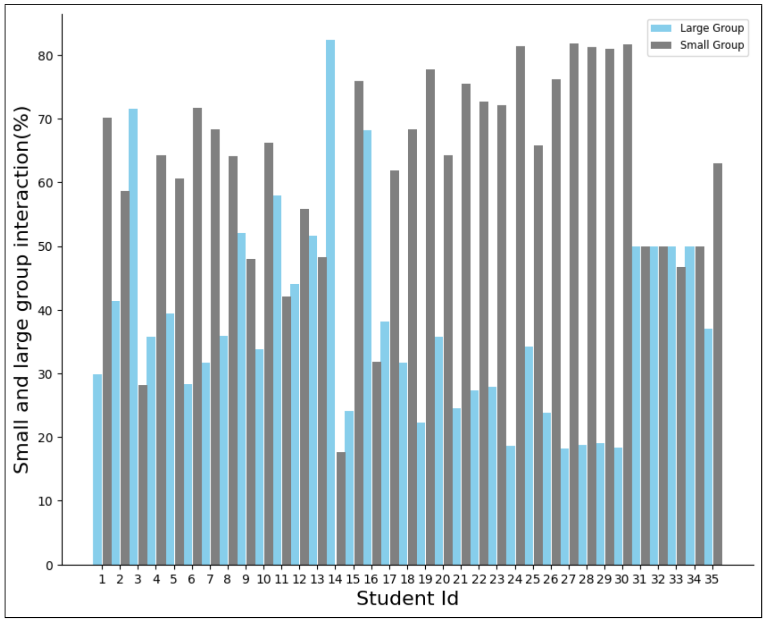


Figure 5.3: Large and Small Group Interaction

5.4 Mobile usage while social interaction

This graph shows a combined result of mobile usage and social interaction. It shows the average smartphone usage of 30 students during social interaction. Out of the total time of social interaction, approximately on average 45% of the respective total time, 80% of students are using mobile while interacting which shows a lack of social engagement or distraction in talking. While there were also 10% of students, who are more interested in social engagement by having approx,30-35% of their total time on mobile.

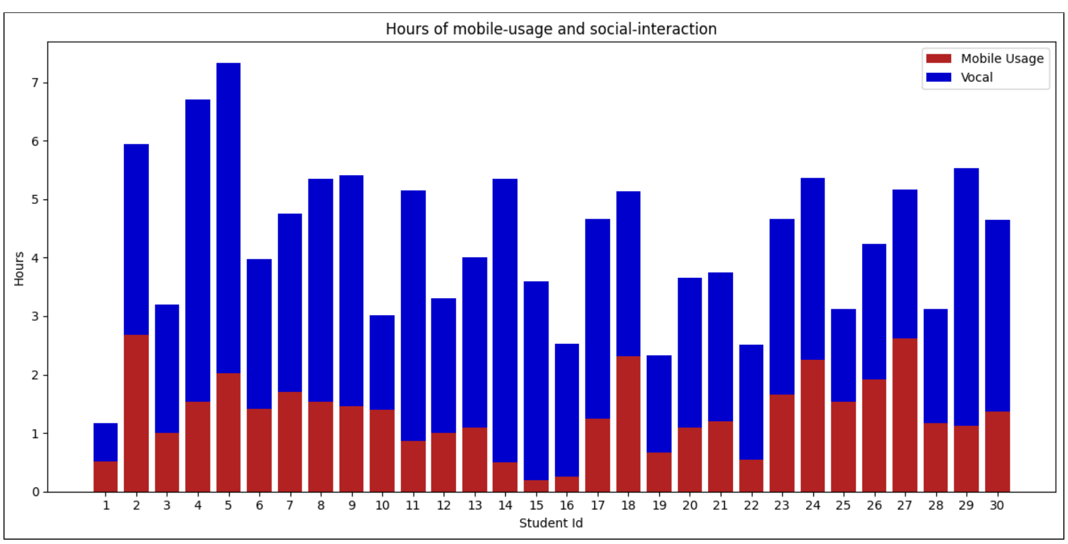


Figure 5.4: Mobile Usage while interaction

CHAPTER 6

Crowd Density

In this chapter, we will explore crowd density variations in different sublocations, examine crowd movement patterns, analyze crowd mobility, and visualize crowd flow movements.

6.1 Crowd Density

Crowd density refers to the number of people that are present in a given indoor space. Recent years have seen a significant rise in research interest in crowd behavior analysis and crowd density estimation, especially because of their critical importance in promising the smooth organization of events. Effective crowd management and monitoring strategies have become essential for modeling and constructing public spaces such as shopping malls, temples, stadiums, and airports, aiming to prevent hazardous conditions.

Understanding occupancy levels, improving space utilization, and increasing both the administrative and visitor experiences depend on the analysis of the most frequently used locations and the identification of peak usage hours. Existing methods for estimating crowd density, however, frequently call for additional infrastructure or hardware configuration, complicating installations and restricting their ability to be widely used.

Existing work in these areas demands extra hardware-based infrastructure and video surveillance-based techniques such as Bluetooth sensors and CCTV installment have been carried out to estimate crowd density but are not much efficient due to light constraints at night time [31][13] To collect crowd positioning information, each visitor was asked to wear a bracelet equipped with a proximity sensor that is available at the museum to operate in conditions of high crowd density[10].

The main goal of this study is to create a system for measuring and estimating crowd density using Wi-Fi connectivity data using a smartphone sensor. Wi-Fi access is becoming widespread in public areas, and making use of this already-existing infrastructure presents a promising method for crowd research and monitoring without the expense and inconvenience of installing invasive technology. For the advantage of both administrators and visitors, it is feasible to evaluate population density, identify congested regions, and crowd flow patterns in sub-areas, and detect peak usage patterns by utilizing the data provided by Wi-Fi networks.

6.2 Indoor Localization

Indoor localization looks at estimating the location of users in indoor spaces and further extending the system to navigate them to their destination within indoor environments where GPS signals are limited or unreliable. Most indoor localization systems have a client-server architecture and concentrate on locating individual users, primarily smartphones. The server determines and communicates the user's location while ignoring the vast amount of collective data that includes the spatiotemporal distribution of activities, mobility patterns, and facility utilization within buildings. Such a system typically utilizes a combination of hardware and software components to determine the location of objects or people within a building or enclosed space. Indoor localization systems can be implemented using many techniques, such as Wi-Fi, Bluetooth, Computer vision, and inertial sensors [21].

In indoor localization systems, the minimal attention paid to user location inference often ends in a neglect of the vast amount of available collective information. By transferring the focus to the server side, a thorough understanding of crowd behavior can be gained, enabling the deployment of crucial services like improving facility planning, maintaining crowd flow, and maximizing space utilization. Exploring a server's perspective can reveal valuable data, improve system performance, as well as improve user experience.

Smartphones have become an integral part of human life. With the embedded sensor ecosystem, these devices offer a practical platform for analyzing human behavioral patterns. A significant rise has been observed in research interest in crowd behavior analysis and crowd density estimation, due to its crucial significance in ensuring the seamless management of events. The majority of the work, however, aims at locating an individual in an indoor space leading to a multitude of indoor localization solutions. But the wealth of information regarding user's location inference is not being used in an efficient manner that is present with the servers of indoor localization architecture. So here we propose a method of extending indoor localization through Wi-Fi connectivity information obtained from users' smartphones. The proposed approach is briefly explained in section 6.3.

6.3 Proposed Method

The proposed method here is an extension to an indoor localization system primarily using smartphone and Wi-Fi APs n/w.In order to simplify the process of monitoring crowd density and analyzing mobility patterns in indoor spaces, we propose utilizing existing Wi-Fi AP and smartphones. By collectively examining server-based information retrieval and access point data, we can develop a crowd density-based estimation system.

In this system, we consider a scenario where N Wi-Fi AP is present in a public area. Each access point may have multiple connected gadgets at any given time, and these connected devices can move between various Wi-Fi AP. We can count the number of these devices present at a specific moment per AP, define prime AP usage times, and examine mobility trends by tracking devices across several Wi-Fi AP by tracking their movements. We want to make it simpler via already-installed Wi-Fi AP and a smartphone-based solution, which will gather information from each roaming user. Since users moving through an indoor area will undoubtedly connect to Wi-Fi infrastructure it will be easy to gather information. Consider below Fig 6.1 to understand the given work. We suggest an Android software named "Usage Tracker" that combines in-built sensors and a Wi-Fi sensor to simplify the implementation. This app gives data about which Wi-Fi AP a user is currently connected to in real-time. The Usage Tracker app is lightweight and doesn't occupy much storage. As users move from one AP to another during the day while using the Usage Tracker software installed on their mobile devices, location inference data is gathered into a CSV file.

We can extract crowd density, identify mobility trends, and monitor user movements at multiple Wi-Fi APs by merging indoor location-based Wi-Fi connectivity information with smartphone sensor data gathered by the Usage Tracker app.

As described in section 3.1, we conducted a study where the Usage Tracker app was installed on the mobile phones of 40 students all over a data collection period of 5 weeks in order to gain insights into crowd behavior within our institute. And we applied data preprocessing, algorithms on collected data to find significant results.

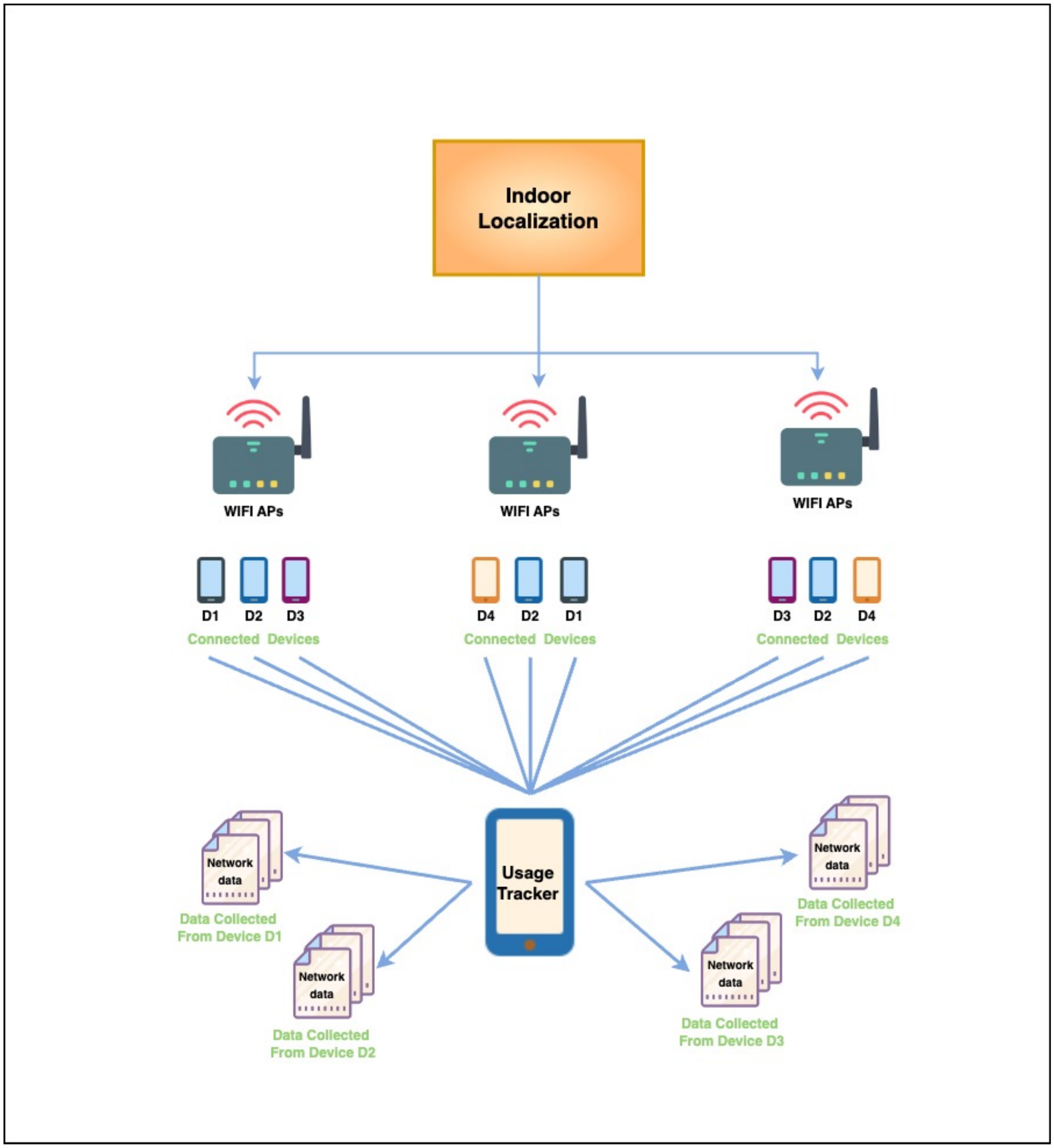


Figure 6.1: Architecture of proposed approach

6.4 Algorithm and Data Preprocessing

6.4.1 Login and Logout time identification

The algorithm 4 is designed to analyze the mobility patterns of students by tracking connections to various Wi-Fi AP and capturing their movements over time. It also helps in estimating the duration or amount of time a student connected to particular Wi-Fi AP by using the login time and logout time of each connection. It also helps in finding crowd density at different Wi-Fi AP with changing times.

Algorithm 4: Login-time and Logout-time at Wi-Fi AP

```
input  : network data csv
output: time_loc_dict: login and logout time for each Wi-Fi APs
time_loc_dict= { }
previous_Wi-Fi=None
for row in Wi-Fi_data do
    current_Wi-Fi=getattr(row['Wi-Fi ID'])
    if previous_Wi-Fi is None then
        time_list.append(getattr(row['Hour'])
    else
        if current_Wi-Fi!=previous_Wi-Fi then
            time_list.append(getattr(prev_row['Hour']))
            time_list.append(getattr(row['Hour']))
    previous_Wi-Fi=current_Wi-Fi
    previous_row=row
convert time_list into time_loc_dict by making sublist of time_list into 2
  part as login and logout time
for time in time_list do
    location=getattr(row['Wi-Fi ID'])
    time_loc_dict.append([time[0],location])
```

The algorithm takes a CSV file containing network data, including the date, time instance, connected Wi-Fi APs, and their IP addresses. It analyzes the data to determine the start time and end time for each Wi-Fi AP throughout the day. To find the changing location, we need to access track the previous row and the current row's Wi-Fi APs. It will iterate through each row of network data and it will store the current Wi-Fi AP's value and will also update the previous Wi-Fi AP's value. It will check the conditions of the current Wi-Fi and previous Wi-Fi and if both of them are not the same means the student has moved from one to another location. Then those time instances at the previous row and current row both will be stored in a dictionary as time_list and with the Wi-Fi APs. After the loop completion, time_list will be divided into two parts and given name start_time and end_time with Wi-Fi AP's.

This way we will be able to find the correct login time and logout time at a particular Wi-Fi APs for all students throughout the entire data experiment period which will eventually help us to identify crowd behavior analysis.

6.4.2 Data Polishing

In order to ensure an accurate analysis of student behavioral patterns, a data polishing technique was applied to refine the timestamp data. The goal of this technique was to standardize the time values and account for the presence of specific IDs during the first_connect and last_connect instances at different locations.

The data polishing process involved adjusting the timestamp values to align them with a consistent interval, thus enabling more meaningful analysis. Specifically, the original timestamp data were rounded to the nearest multiple of 10 minutes. This rounding approach allowed for the creation of fixed time intervals and facilitated the identification of crowd density patterns.

To achieve this, the minutes component of each timestamp was evaluated to determine its proximity to the nearest multiple of 10. If the difference between the minutes and the nearest multiple of 10 was equal to or greater than 5, the minutes were adjusted forward to the next multiple of 10. Conversely, if the difference was less than 5, the minutes were adjusted back to the previous multiple of 10. This adjustment ensured that the timestamps were aligned with the appropriate time intervals for subsequent analysis. By applying this data polishing technique, the timestamp data was effectively prepared for further analysis, such as data binning. The rounded timestamps ensured that the data could be grouped into fixed time intervals, allowing for the identification of meaningful crowd density patterns.

6.4.3 Data Binning

To analyze data more effectively and to derive meaningful insights in finding smaller duration crowd density, it is necessary to divide the login time and logout time into smaller time intervals. Data binning helps us to group the data based on specific time intervals, in this case, 10-minute intervals. By dividing the login and logout time into 10-minute intervals, we were able to find more insights into crowd density at each Wi-Fi APs in more detail.

For example, consider a scenario where the login_time is 10:10 and the logout_time is 12:00, indicating a time period of almost two hours. Dividing this time period into 10-minute intervals allows us to break down the data into multiple smaller segments, such as 10:10-10:20, 10:20-10:30, and so on. This facilitates the analysis of user activity and behavior within each specific interval, providing valuable insights into the patterns and trends that may emerge.

6.4.4 Time-based movements between different wifi AP

Finding time-based movements of the number of students moving from one location to another location will give us insights into how much amount of time a crowd moves. After applying data polishing and data binning on login time and logout time data, it will give us divided time intervals. For example, if a student stays at the hostel for 30 mins the input to this algorithm will be 6 5 mins time interval-based entries where the student is connected to the hostel.

To implement algorithm 5, we need 3 hashmap variables namely hm1,hm2, and hm3. hm1 represents the input CSV file given to it as is by mapping student_id to a particular location with respective current_log_in time which shows the initial step in time for that student at the given location. hm2 is constructed from hm1 which gives the mapping of current_log_in time to particular strudent_id along with the respective location. To get the crowd movement from one location to another location, we travsere hm2 and collect the current_log_in time for particular location and store it in prev_log_in_time variable. Then we fetch the details of previous_location and current_location for particular student from hm2 based on prev_log_in_time and current_log_in_time respectively. Moreover, take one list to store the count of students moving from previous_location to current_location in 5 minute of time intervals. To store these details in the list hm3 used, It represents the mapping of current_log_in time to list of crowd movement count happening between respective previous_location and current_location. Whenever hm3 contains current_log_in_time, movement count is incremented for particular pair of previous_location-current_location for respective current_log_in time present in hm3 else movement count is set to 1 for the same.

Algorithm 5: Crowd Movement Analysis

```
input  : CSV file with first and last connect details at a particular location
         for every student
output: Crowd movement count from every source to destination
Take hm1 = HashMap<s_id, <curr_login_time,location>>
Generate hm2 = HashMap<curr_login_time,<s_id,location>>
list=[]
movementCount=0
flag=1
for ∀curr_login_time ∈ hm2 do
    if flag==1 then
        prev_login_time=curr_login_time flag=0
    else
        for ∀value ∈ hm2 do
            prev_location ← location wrt prev_login_time
            curr_location ← location wrt curr_login_time
            list ← student_count(prev_location,curr_location)
            Take hm3 = HashMap<curr_login_time ,
              list[movementCount]>
            if prev_location!=curr_location then
                if curr_login_time ∈ hm3 then
                    movementCount=count(list(prev_location,curr_location))+
                      1
                else
                    movementCount=count(list(prev_location,curr_location))=1
                prev_log_in_time=current_log_in_time
```

6.5 Results

In this, we have analysed the Crowd Density at Different Locations,Crowd movements and the frequency of crowd movements.

6.5.1 Crowd Density at Different Locations

6.5.1.1 Crowd density at hostel location

Fig 6.2 shows the crowd density at the hostel location during different time slots throughout the day, which will help us estimate the crowd at the hostel location. In the morning, 8–9, approx 45% to 82% of students stay at the hostel. Some crowd changes according to different circumstances, as some students might have a lecture at morning 8 AM, or some might go to the Cafeteria. Midnight 10–12 are the prime time slots where most of the approx 65 % to 90% of students stay in the hostel. In the afternoon time slot, that is, 2-5 PM, around 50% to 70% of the total students were found to be present at the hostel. At midnight 12 PM- 7 AM morning almost all of the students stay at the hostel.

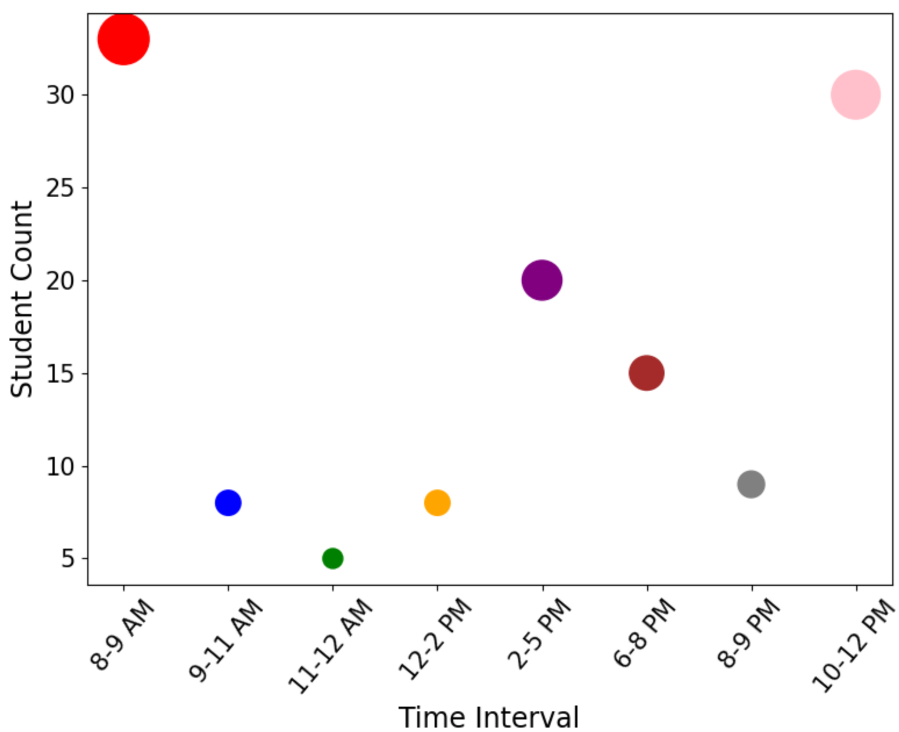


Figure 6.2: Crowd density at Hostel

6.5.1.2 Crowd density at lecture hall location

Fig 6.3 shows the crowd density at the Cep/lecture hall. In the morning, 8-9 AM, on average, around 25% to 50% students stay at cep because of lecture timings. From 9 AM-12 PM morning, around 40% - 80% of the students stay at CEP. During the two-day afternoon, the workshop was held at the Cep building around 2:30 PM - 4:30 PM in our data experiment period, so 30% of the relevant crowd was present. Otherwise, lectures and seminars/workshops are not much utilized throughout the day.

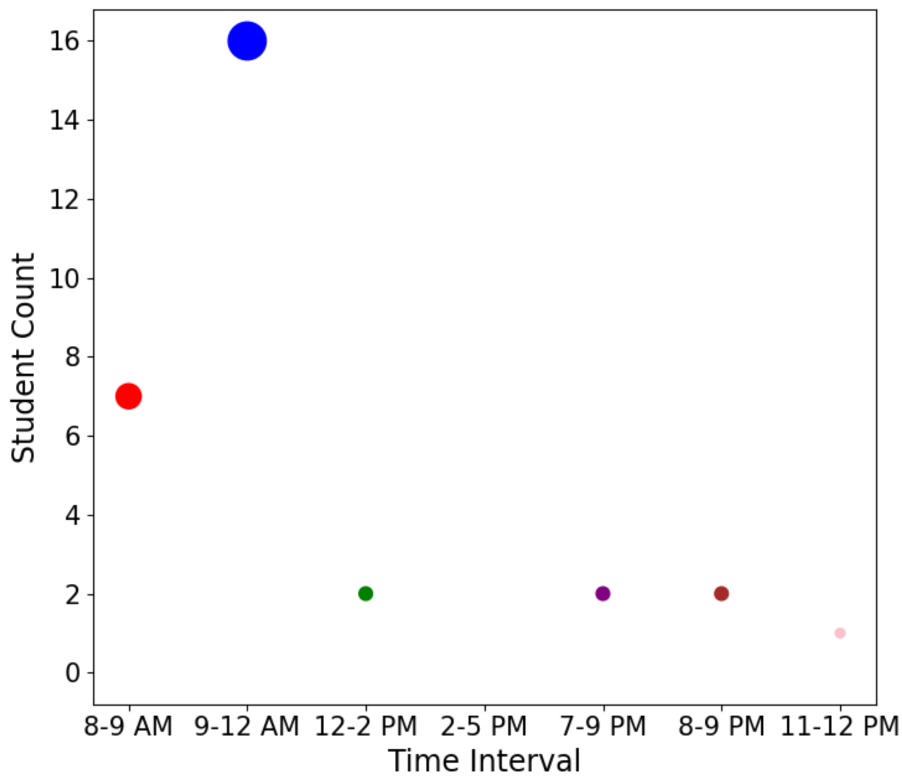


Figure 6.3: Crowd density at Lecture Hall

6.5.1.3 Crowd density at cafeteria location

Fig 6.4 shows crowd density at cafeteria locations at different time slots of the day. The crowd will be present at the cafeteria location mainly at breakfast, lunch, and dinner timings. However around 10% - 25% of the crowd will always be present at the cafeteria throughout the day. During the morning 8 AM - 9 AM around 12% to 20% of students will be present at the canteen location, but during those times some students also may present at the hostel location and may be attending lectures. Due to that more crowd is visible at around 9 AM - 11 AM. During lunch timings from 12:00 AM- 2 PM on an average 65% to 75%, the crowd is the present canteen. And at dinner from 7 PM - 9 PM around 50% - 65% of the students from the total crowd present at the canteen. Students also tend to spend time at the canteen from midnight 10 PM-12 PM.

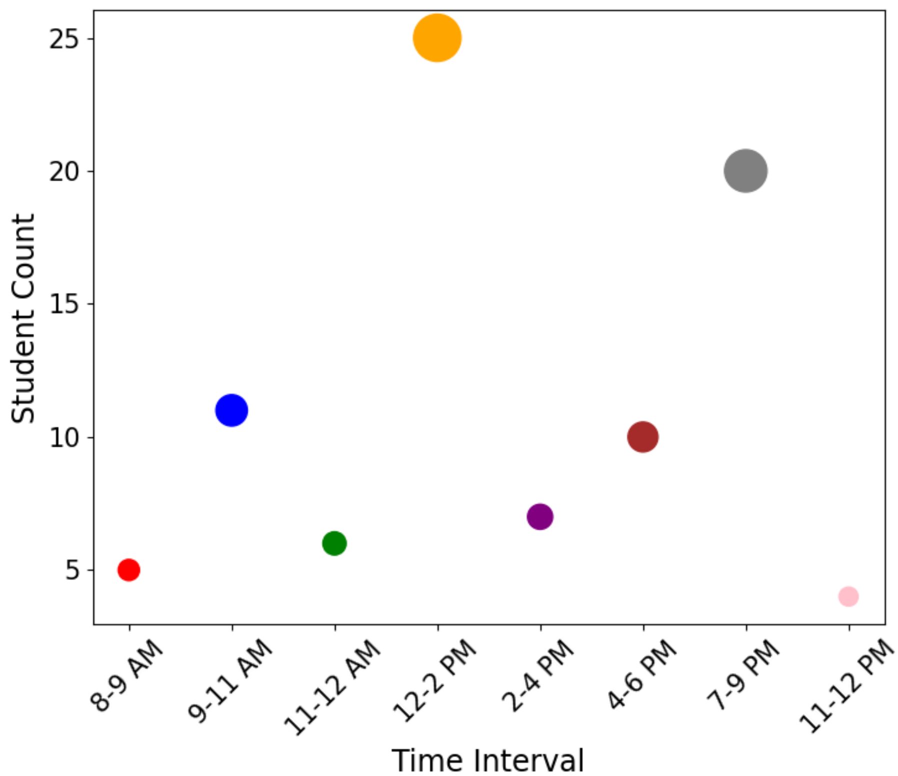


Figure 6.4: Crowd density at Cafeteria

6.5.2 Access Duration per Wi-Fi AP

This graph shows the amount of time spent by the crowd during the entire experimental duration at each Wi-Fi APs. This helps us in understanding where the crowd resides throughout a day on average. At the Lecture Hall, the crowd spends time within the range of 15 mins to 3.3 hours. At the cafeteria, the crowd spends around min 20 minutes to 3 hours. In the library, the crowd spends around 7-8 mins to 1.7 hours with an average of around 50 min. At the Hostel, the location crowd spends around 7 hr - 16 hr with an average of 12.5 hr. At the lab building, students spend around 15-20 mins to 1.6 hr with an average of 45 mins. From this box plot, we can see access duration at various Wi-Fi - APs of the institute. This will help us in identifying the maximum, minimum, and average amount spent by crowds at different Wi-Fi APs. One important insight which can be inferred is that there are some outliers as well other than average crowd access behavior as described here. We can also infer that one of the most used Wi-Fi AP is the hostel, which will be used even 55% - 60% more than all combined locations.

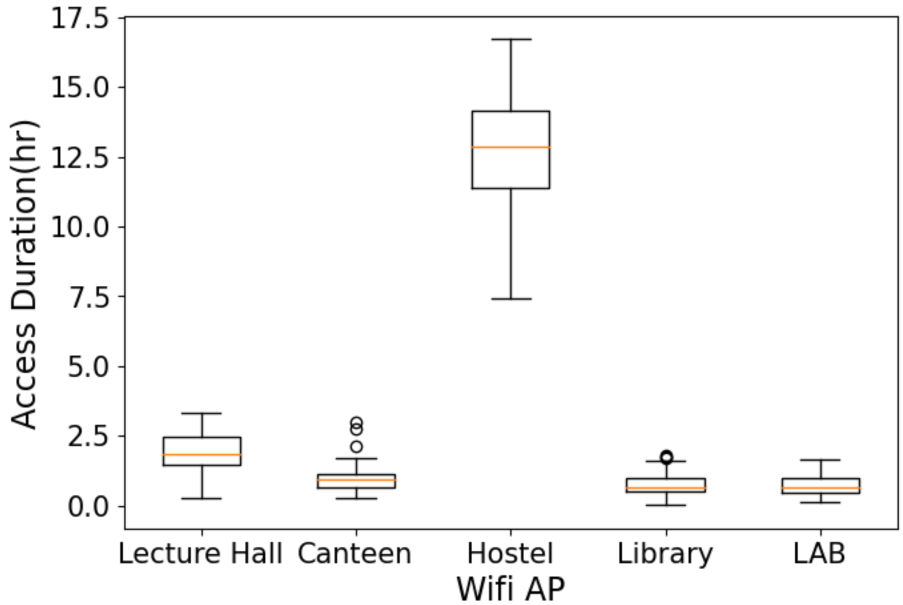


Figure 6.5: Access Duration of each Wi-Fi AP

6.5.3 Patterns of crowd density

Analyzing crowd density across different events or categorizations based on Wi-Fi AP is very essential to understand their behavior. Fig 6.6, shows four different events such as regular days, exams,tech-fest, and weekends and analyzes crowd density at four different Wi-Fi APs during our experiments. We analyzed that Cep, labs, and faculty block are the least visited places by the crowd on weekends. During the technical festival organized at our institute canteen becomes one of the most useful places and then followed by the hostel and library. And during the exams, CEP, Hostel, and library are one of the most utilized places.

Using this one of the most useful insights is that different places will be utilized by crowds on different occasions.

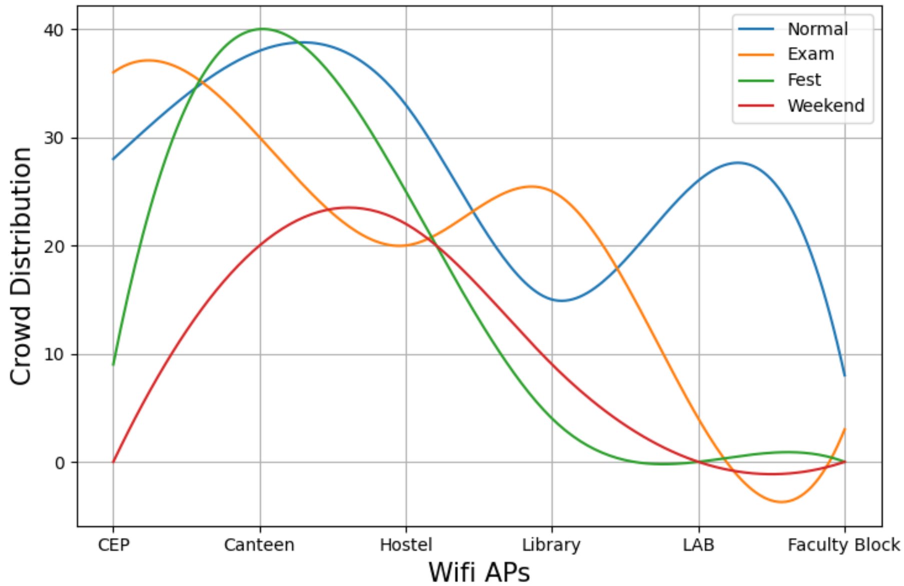
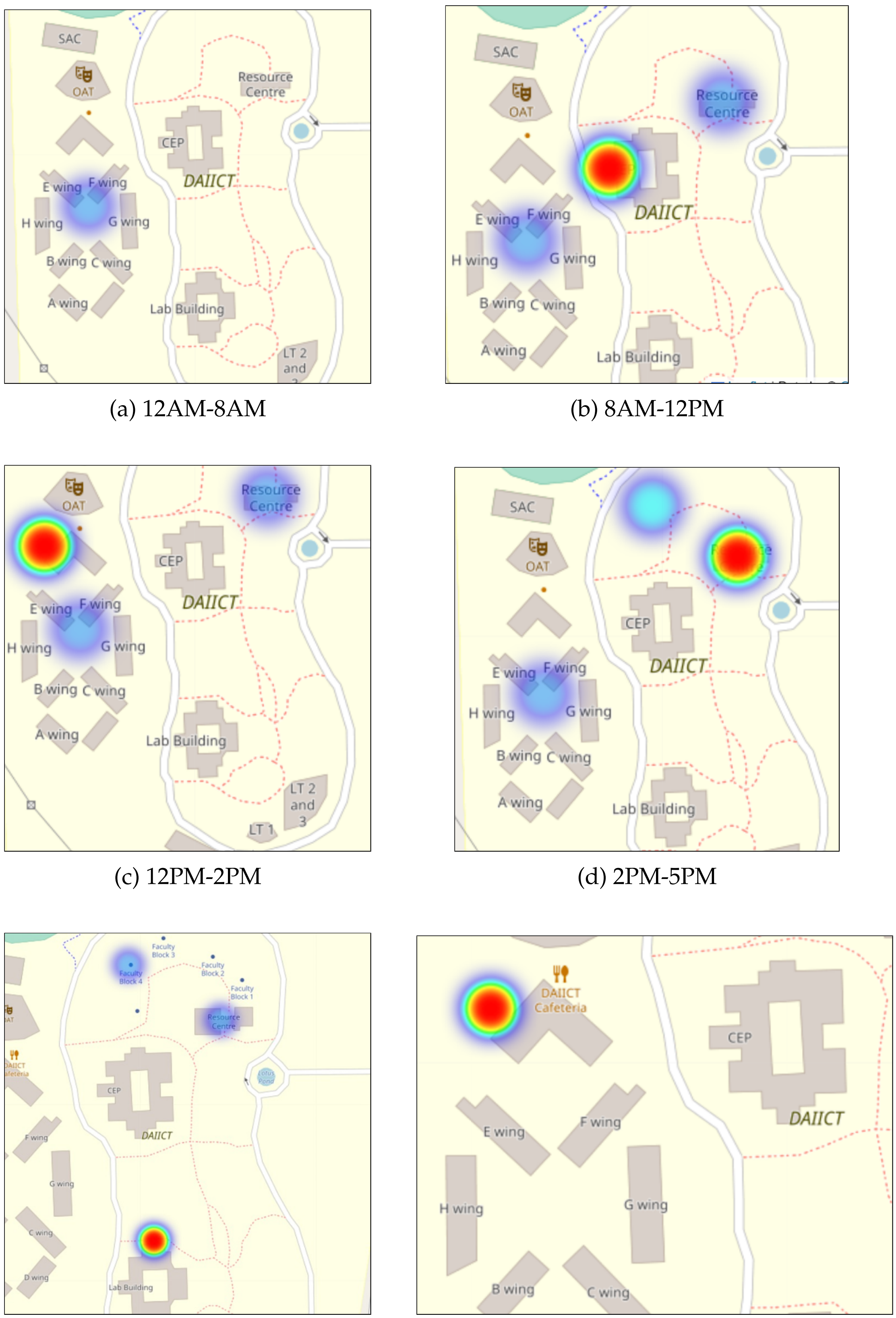


Figure 6.6: Crowd patterns during different events

6.5.4 Crowd Heatmap

The analysis of the crowd density heatmap reveals interesting patterns during different time slots on the DAIICT campus. Fig 6.7a, represents the duration between 12 am to 8 am, it is observed that students predominantly stay in their hostels during this period. Moving on to the next time slot, from 8 am to 12 pm, Fig 6.7b depicts that students have their lectures in the CEP block, resulting in a high concentration of students in that area. Additionally, a smaller number of students can be seen in the hostels and resource-center during this time. The afternoon time slots are divided into two intervals. The first interval, from 12 pm to 2 pm, shows a significant crowd density in the canteen, indicating that students gather there for meals. Meanwhile, a few students can still be found in the hostels and resource center as shown in the corresponding heatmap in Fig 6.7c. The second interval, from 2 pm to 5 pm, highlights a shift in student presence at the resource center. The heatmap in Fig 6.7d demonstrates a high concentration of students in the resource center, while some students remain in the hostels and faculty block during this time. During the evening slot from 5 pm to 8 pm, which generally corresponds to lab timings, students are heavily concentrated in the lab building. The heatmap in Fig 6.7e vividly displays this trend, with only a few students visiting the resource center and faculty block. Lastly, in the night time slot from 8 pm to 10 pm, Fig 6.7f exhibits that almost all students gather in the cafeteria, likely for dinner or socializing.



(a) 12AM-8AM

(b) 8AM-12PM

(c) 12PM-2PM

(d) 2PM-5PM

(e) 5PM-8PM

(f) 8PM-10PM

Figure 6.7: Heat Maps for crowd density visualization

6.5.5 Frequency of Movements CEP-Hostel

The Fig 6.8 provides insights into the frequency of movements occurring from the lecture hall to the hostel, allowing for analysis of the time duration within which students make this transition. By examining the data collected over a 30-day experiment period, it was observed that different percentages of students complete their movements within specific time intervals. Specifically, the data reveals that 25% of the total student movements from the lecture hall to the hostel occur within a time frame of 50 minutes. This information indicates that a significant portion of students can efficiently complete their journey within this relatively short duration.

And almost around 45% of total students' movements happen within 1 hour. In addition to that, 60% of the movements happen within 85 minutes. And also, approx. 85% of students move within 2 hours of the time period. This will help us in analyzing how movements happen from the cep to the hostel. It allows for a better understanding of the patterns and trends associated with student movements. Institutions can leverage this data to streamline operations, such as adjusting class schedules and optimizing lecture hall capacities.

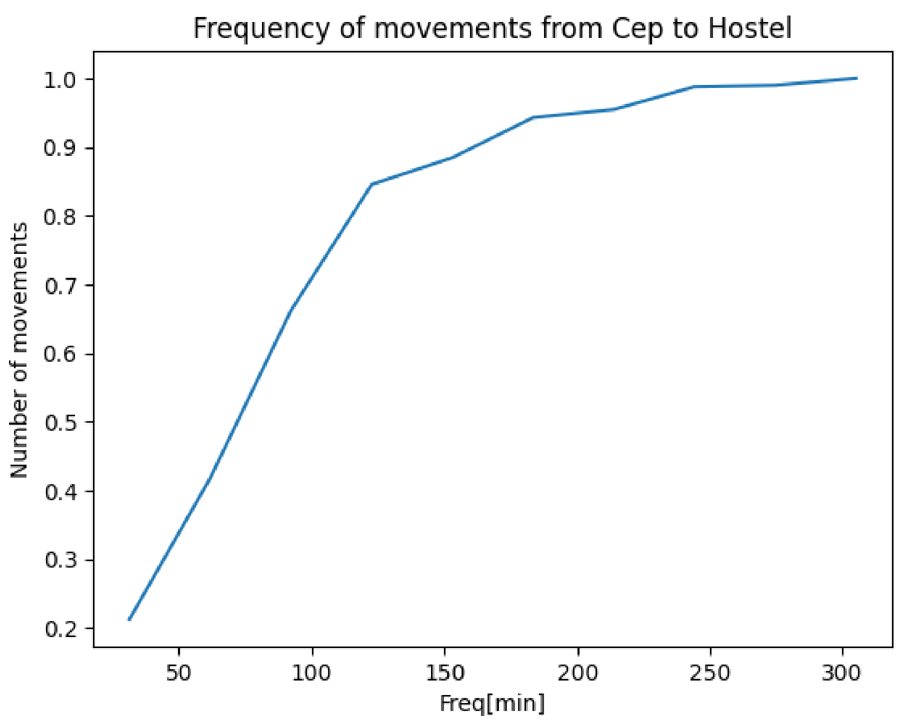


Figure 6.8: Frequency of Movements CEP-Hostel

6.5.6 Frequency of Movements Canteen-Hostel

The Fig 6.9 shows the frequency of movements that are happening from the cafeteria to the hostel. It helps us to analyze, within a specific time, amount of students moving from the cafeteria to the hostel. Here based on the approx duration in which students move from the canteen to the hostel during the entire data experiment period of 30 days and by 40 students, we found that within 20-25 minutes 30% of students move from the hostel to the canteen. Another thing is 30-35 minutes 50% of the total movements of students will happen from one hostel to the canteen. Also, approx. 70% of total student's movements happen from the canteen - hostel within 1 hr and moreover, 80% will move within 75 minutes of time intervals. Also, approx. 90% of the movements happen within 90 minutes of the time interval. This will help us in analyzing how movement happens from the canteen to the hostel. This finding emphasizes the significance of this duration and indicates that the majority of students successfully complete the moves journey within this time frame. The frequency of student transitions from the cafeteria to the hostel provides valuable insights for crowd management and resource optimization.

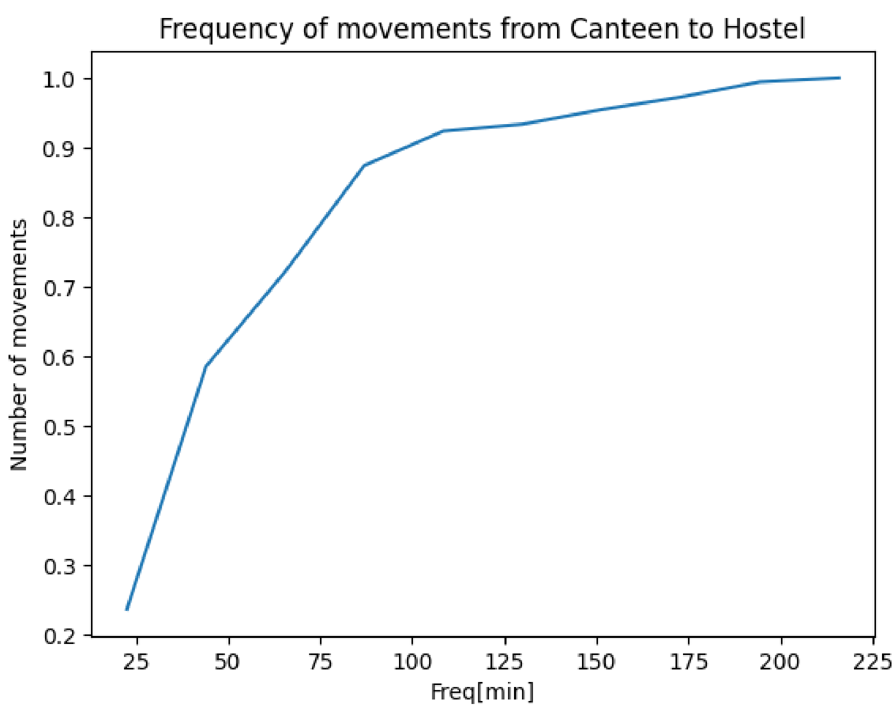


Figure 6.9: Frequency of Movements Canteen-Hostel

CHAPTER 7
Conclusion And Future Work

Conclusion

In this thesis, we have presented a sensing system to collect sensor data through the smartphones of students at DAIICT. This sensing system infers student behavior in terms of conversation level, smartphone usage, unlock count, and visited locations. Data from the microphone sensor and Wi-Fi sensor has been used to find the conversation level of students. The average conversion of students ranges between ± 3.4 and 2.5. By combining Bluetooth data and conversation level we found that 70% of students are more involved in smaller group discussions rather than large group involvement. Extension to indoor localization system using Wi-Fi infrastructure was carried out to quantify crowd density-based estimation in real-time. We observed crowd density at different Wi-Fi APs of our institute using the usage tracker app's collected data. We identified prime time slots, event-wise crowds at various Wi-Fi APs, and student crowd movement from one location to another location using dynamic changes in crowd flow. This will help both administrators and visitors, it is feasible to evaluate population density, identify congested regions, and crowd flow patterns in crowded places such as stadiums, malls, museums, etc.

Future Work

In future work, we want to implement a dashboard integrated within our application usage tracker itself, which will help user to get insights into their social circle and dynamics. The app module will show each user their daily and weekly social interaction duration patterns and average interpersonal communication performed over a week period of time, the number of people with whom they have interacted, and the type of interaction they got involved in more groups or single. Another in-app module related to crowd density for administration-side management, the amount of crowd present at different sublocations of specific areas, and user density estimation information within different subareas in a public place such as a museum or shopping mall at specific locations. And dynamic real-time people flow numbers within sublocations.

References

[1] J. E. Bardram and A. Matic. A decade of ubiquitous computing research in mental health. *IEEE Pervasive Computing*, 19(1):62–72, 2020.

[2] D. Ben-Zeev, E. A. Scherer, R. Wang, H. Xie, and A. T. Campbell. Next-generation psychiatric assessment: Using smartphone sensors to monitor behavior and mental health. *Psychiatric rehabilitation journal*, 38(3):218, 2015.

[3] Demasi, Aguilera, and Recht. Detecting change in depressive symptoms from daily wellbeing questions, personality, and activity. In *2016 IEEE Wireless Health (WH)*, pages 1–8, 2016.

[4] S. Georgievska. Detecting high indoor crowd density with wi-fi localization: a statistical mechanics approach. *Journal of Big Data*, 6, 03 2019.

[5] J. Johnstone. Clarence : A mobile application for behavior and wellness tracking. 2015.

[6] K. Katevas, K. Hänsel, R. Clegg, I. Leontiadis, H. Haddadi, and L. Tokarchuk. Finding dory in the crowd: Detecting social interactions using multi-modal mobile sensing. In *Proceedings of the 1st Workshop on Machine Learning on Edge in Sensor Systems*, SenSys-ML 2019, page 37–42, New York, NY, USA, 2019. Association for Computing Machinery.

[7] N. Lane, M. Rabbi, M. Lin, X. Yang, H. lu, S. Ali, A. Doryab, E. Berke, T. Choudhury, and A. Campbell. Bewell: A smartphone application to monitor, model and promote wellbeing. *Proceedings of the 5th International ICST Conference on Pervasive Computing Technologies for Healthcare*, 01 2011.

[8] M. Li, H. Wang, B. Guo, and Z. Yu. Extraction of human social behavior from mobile phone sensing. In R. Huang, A. A. Ghorbani, G. Pasi, T. Yamaguchi, N. Y. Yen, and B. Jin, editors, *Active Media Technology*, pages 63–72, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[9] C. Lijun and H. Kaiqi. Video-based crowd density estimation and prediction system for wide-area surveillance. *China Communications*, 10(5):79–88, 2013.

[10] C. Martella, M. Cattani, and M. van Steen. Exploiting density to track human behavior in crowded environments. *IEEE Communications Magazine*, 55(2):48–54, 2017.

[11] A. Matic, V. Osmani, A. Maxhuni, and O. Mayora. Multi-modal mobile sensing of social interactions. In *2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*, pages 105–114, 2012.

[12] M. Mohandes. A case study of an rfid-based system for pilgrims identification and tracking. In C. Turcu, editor, *Sustainable Radio Frequency Identification Solutions*, chapter 6. IntechOpen, Rijeka, 2010.

[13] M. Nakatsuka, H. Iwatani, and J. Katto. A study on passive crowd density estimation using wireless sensors. In *The 4th Intl. Conf. on Mobile Computing and Ubiquitous Networking (ICMU 2008)*. Citeseer, 2008.

[14] H. S. A. L. A. Nastasia Griffioen, Nastasia Griffioen and A. L. Aschoff. Everyone does it—differently: A window into emerging adults' smartphone use. *July 2021Humanities and Social Sciences Communications 8(1)*, 2014.

[15] S. Patel, K. Sasidhar, M. Vaghela, P. Katrodia, and R. Wagani. An m-health app that goes beyond screen time applications. In *2022 IEEE 10th Region 10 Humanitarian Technology Conference (R10-HTC)*, pages 94–99, 2022.

[16] S. Patel, K. Sasidhar, M. Vaghela, P. Katrodia, and R. Wagani. An m-health app that goes beyond screen time applications. In *2022 IEEE 10th Region 10 Humanitarian Technology Conference (R10-HTC)*, pages 94–99, 2022.

[17] L. Pei, M. Zhang, D. Zou, R. Chen, and Y. Chen. A survey of crowd sensing opportunistic signals for indoor localization. *Mobile Information Systems*, 2016, 2016.

[18] W. Raad, A. Hussein, M. Mohandes, B. Liu, and A. Al-Shaikhi. Crowd anomaly detection systems using rfid and wsn review. In *2021 4th International Symposium on Advanced Electrical and Communication Technologies (ISAECT)*, pages 1–5, 2021.

[19] M. Raento, A. Oulasvirta, and N. Eagle. Smartphones: An emerging tool for social scientists. *Sociological Methods & Research*, 37(3):426–454, 2009.

[20] G. Solmaz and D. Turgut. A survey of human mobility models. *IEEE Access*, 7:125711–125731, 2019.

[21] X. Tang, B. Xiao, and K. Li. Indoor crowd density estimation through mobile smartphone wi-fi probes. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 50(7):2638–2649, 2020.

[22] K. Vaghela, M. Sasidhar. Assessing mobile usage, physical activity and sleep through smartphone sensing: A digital phenotype study. *Digital Phenotype Study. SN COMPUT. SCI. 3, 331*, 2022.

[23] M. Vaghela and K. Sasidhar. Analyzing the human behavior using pervasive sensing system. In *Proceedings of the 21st International Conference on Distributed Computing and Networking*, pages 1–5, 2020.

[24] M. Vaghela and K. Sasidhar. Analyzing the human behavior using pervasive sensing system. In *Proceedings of the 21st International Conference on Distributed Computing and Networking*, ICDCN '20, New York, NY, USA, 2020. Association for Computing Machinery.

[25] M. Vaghela, K. Sasidhar, A. Parikh, and R. Wagani. Assessing mobile usage, physical activity and sleep through smartphone sensing: a digital phenotype study. *SN Computer Science*, 3(5):331, 2020.

[26] M. Versichele, T. Neutens, M. Delafontaine, and N. Van de Weghe. The use of bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the ghent festivities. *Applied Geography*, 32:208–220, 03 2012.

[27] R. Wang, F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, and A. T. Campbell. Studentlife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*, pages 3–14, 2014.

[28] R. Wang, W. Wang, A. daSilva, J. F. Huckins, W. M. Kelley, T. F. Heatherton, and A. T. Campbell. Tracking depression dynamics in college students using mobile phone and wearable sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(1), mar 2018.

[29] J. Weppner and P. Lukowicz. Bluetooth based collaborative crowd density estimation with mobile phones. In *2013 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 193–200, 2013.

[30] Y. Yoshimura, A. Krebs, and C. Ratti. An analysis of visitors' length of stay through noninvasive bluetooth monitoring in the louvre museum. 04 2016.

[31] Y. Yuan, C. Qiu, W. Xi, and J. Zhao. Crowd density estimation using wireless sensor networks. In *2011 Seventh International Conference on Mobile Ad-hoc and Sensor Networks*, pages 138–145, 2011.

[32] Y. Yuan, J. Zhao, C. Qiu, and W. Xi. Estimating crowd density in an rf-based dynamic environment. *IEEE Sensors Journal*, 13(10):3837–3845, 2013.

[33] X. Zhang, Y. Sun, Q. Li, X. Li, and X. Shi. Crowd density estimation and mapping method based on surveillance video and gis. *ISPRS International Journal of Geo-Information*, 12:56, 02 2023.

[34] X. Zhang, Y. Sun, Q. Li, X. Li, and X. Shi. Crowd density estimation and mapping method based on surveillance video and gis. *ISPRS International Journal of Geo-Information*, 12(2):56, 2023.