UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Ph.D. Dissertation

An Architecture for Crowd Density Estimation in Heterogenous Opportunistic Environment

BY

ADDEPALLI LAVANYA

ADVISOR

LLORET MAURI, JAIME

FEBRUARY, 2024

MOTIVATION

"IF YOU QUIT ONCE IT BECOMES A HABIT. NEVER QUIT!!!"

- MICHAEL JORDAN

"IT'S ALWAYS TOO SOON TO QUIT!"

- NORMAN VINCENT PEALE

WHEN I HEAR SOMEBODY SIGH, 'LIFE IS HARD,' I AM ALWAYS TEMPTED TO ASK, 'COMPARED TO WHAT?'

- SYDNEY J. HARRIS

ACKNOWLEDGEMENT

THERE IS ALWAYS SOMETHING AND SOMEONE TO BE GRATEFUL FOR…!

I begin by conveying my sincere gratitude to my advisor Prof. Jaime Lloret for all the backing until the extremes. He is such an amazing advisor who has always watched my best interest, the time when I have been in lag, he has always been patient for weeks and months at times. Every time when I have had anything to discuss about, he made it a point to reply my email within few hours. He is ace of the extraordinary advisor any doctoral student can ask for, always gave me my total space to decide on what and how to work. He always gave me the freedom to collaborate with anyone from anywhere and supported the direction of work. He is one of the strongest pillars who has stood by me with every part of my research. Only a thank you is not enough for such a wonderful academic advisor. He has taught me how an adviser should be defending the student irrespective of any setback no judging and no giving up on a student.

The next individual who taught me how a real researcher should think is Dr. Aline Viana Carneiro, from INRIA, Paris. The time when I served my internship in 2014, it was she who was the reason I got a substantial transition from a person who had an idea of research to a person who is able to do and experiment different dimensions of research. The time when I was supposed to wind up my research and frame a paper, I remember her words “Lavanya you should know when to stop, start is easy but to recognize when to end is what you need to learn”. It was nearly too late for me to realize what exactly she meant, but I did learn it the hard way. The team from INRIA were not only my colleagues, but also have become my friends and kept me motivated all through my time in INRIA.

With a particular mention to Dr. Ramesh Rasker, MIT, USA and his team Mr. Sandeep Shinde, Dr. Sachin Pachorkar and others from the team Kumbathon from Nashik, India. The team of Dr. Mohanty & Mr Mohan from Tata consultancy services, Ltd. Thane, India. They have given me an opportunity to present and showcase my research skills for finding a real-world solution for Kumbh Mela, 2015 in Nashik, India. While working with them I learned about how to be persistent in achieving when you believe in yourself and your idea. During the Kumbh Mela, Nilay Kulkarni a 16-year-old RnD, Mr Ramakrishna from google, Bangalore. India & his smart wife Mrs. Prabha from yahoo, Bangalore, India. Kept me awake for long hours and kept the working spirits alive. The phrase of “Burning Midnight oil” I understood while working on Kumbhaton Project and have equally experienced the magical feeling success after burning the midnight oil.

And lastly, to my family and friends who have been constantly supporting me and standing by side in thick & thin.

The journey of PhD is never easy and every bit of encouragements means a great deal...!!

ABSTRACT

This thesis presents a new framework called the "Dynamic Urban Crowd and Social Interaction Model (DUCSIM)," which is aimed at calculating crowd density and deciphering social networks in opportunistic environments. With the growing commonality of internet-linked electronic gadgets and the widespread influence of online social networks, an enormous digital trail has been created. The digital traces based on human mobility and the increased usage of wireless communication systems such as 3G, 4G, and 5G form a rich database to be analyzed.

These digital traces offer a unique way of modelling the crowd patterns within different contexts, like spontaneous assemblies in public spaces and planned scenarios, as in the case of mega-events. The study focuses on the challenge of opportunistic crowd gatherings, where people congregate for different reasons without planning; they manifest their motions dynamically and unexpectedly. The analysis of human behaviour in modern, developed cities requires that these gatherings occur in malls, road junctions, and flash mobs.

Macroscopic crowd density analysis based on data from MOBILE towers serves as the first stage in outlining the DUCSIM framework. The Median-of-Median (M-o-M) method is adopted for robustness as this analysis involves daily and weekly raw crowd count thresholds. Crowd densities are ranked in quartiles to show varying degrees of crowd distribution. Through the macroscopic analysis, the framework progresses to cumulative crowd mobility analysis. Crowd movement dynamics are measured by changing signals from MOBILE towers and formulating a crowd's density map to forecast its subsequent motions.

It examines the micro-analysis of individual movement and interpersonal relations on a smaller scale. It includes assigning people to MOBILE towers and forming social interaction graphs that infer and update social relationships.

The most important part of DUCSIM lies in its ability to dynamically learn and adapt to

create a novel representation model to suit the newly detected pattern. This flexibility helps to ensure the relevancy of the framework, which must be continually updated.

Custom predictive modelling combines with historical data that encompasses the thesis. The framework uses previous crowd densities and movement data to discover trends and predict upcoming crowd dynamics, thus improving urban planning efficiency, emergency response, or smart cities.

The DUCSIM framework provides a comprehensive, flexible and forecasting method of understanding and controlling urban crowd phenomena. A modern form of data analysis involving several data sources, supported by rigorous mathematics, makes this method unique for urban studies. Moreover, it gives impetus to the academic sphere and provides practical recommendations concerning the application of this methodology within modern city management and planning.

RESUMEN

Esta tesis presenta un nuevo modelo llamado "Modelo dinámico de interacción social y multitud urbana (DUCSIM)", que tiene como objetivo calcular la densidad de multitudes y descifrar las redes sociales en entornos oportunistas. Con la creciente similitud de los dispositivos electrónicos conectados a Internet y la influencia generalizada de las redes sociales en línea, se ha creado un enorme rastro digital. Las huellas digitales basadas en la movilidad humana y el mayor uso de sistemas de comunicación inalámbrica como 3G, 4G y 5G forman una rica base de datos que puede analiarse.

Estas huellas digitales ofrecen una forma única de modelar los patrones de multitud dentro de diferentes contextos, como asambleas espontáneas en espacios públicos y escenarios planificados, como en el caso de los megaeventos. El estudio se centra en el desafío de las reuniones multitudinarias oportunistas, donde las personas se congregan por diferentes motivos sin planificación; manifiestan sus movimientos de forma dinámica e inesperada. El análisis del comportamiento humano en las ciudades modernas y desarrolladas requiere que estas reuniones se produzcan en centros comerciales, cruces de carreteras y flash mobs.

El análisis macroscópico de la densidad de multitudes basado en datos de las torres de telefonía móvil sirve como primera etapa para delinear el marco DUCSIM. Se adopta el método Median-of-Median (M-o-M) para mayor solidez, ya que este análisis implica umbrales de conteo bruto de multitudes diario y semanal. Las densidades de multitud se clasifican en cuartiles para mostrar distintos grados de distribución de la multitud. A través del análisis macroscópico, el marco avanza hacia el análisis de movilidad acumulativa de multitudes. La dinámica del movimiento de multitudes se mide cambiando las señales de las torres de telefonía movil y formulando un mapa de densidad de multitudes para pronosticar sus movimientos posteriores.

Examina el microanálisis del movimiento individual y las relaciones interpersonales a menor escala. Incluye asignar personas a torres de telefonía móvil y formar gráficos de interacción social que infieren y actualizan las relaciones sociales.

La parte más importante de DUCSIM radica en su capacidad de aprender y adaptarse dinámicamente para crear un modelo de representación novedoso que se adapte al patrón recién detectado. Esta flexibilidad ayuda a garantizar la relevancia del marco, que debe actualizarse continuamente.

El modelado predictivo personalizado se combina con datos históricos que engloban la tesis. El marco utiliza densidades de multitudes anteriores y datos de movimiento para descubrir tendencias y predecir dinámicas de multitudes futuras, mejorando así la eficiencia de la planificación urbana, la respuesta a emergencias o las ciudades inteligentes.

El marco DUCSIM proporciona un método integral, flexible y de previsión para comprender y controlar los fenómenos de aglomeración urbana. Una forma moderna de análisis de datos que involucra varias fuentes de datos, respaldada por matemáticas rigurosas, hace que este método sea único para los estudios urbanos. Además, da impulso al ámbito académico y proporciona recomendaciones prácticas sobre la aplicación de esta metodología en la gestión y planificación de las ciudades modernas.

RESUM

Aquesta tesi presenta un nou model anomenat "Dynamic Urban Crowd and Social Interaction Model (DUCSIM)", que té com a objectiu calcular la densitat de multituds i desxifrar xarxes socials en entorns oportunistes. Amb la creixent comú d'aparells electrònics enllaçats a Internet i la influència generalitzada de les xarxes socials en línia, s'ha creat un enorme rastre digital. Les traces digitals basades en la mobilitat humana i l'augment de l'ús de sistemes de comunicació sense fils com 3G, 4G i 5G formen una base de dades rica per ser analitzada.

Aquestes traces digitals ofereixen una manera única de modelar els patrons de multituds en diferents contextos, com ara assemblees espontànies en espais públics i escenaris planificats, com en el cas dels megaesdeveniments. L'estudi se centra en el repte de les reunions multitudinàries oportunistes, on la gent es congrega per diferents motius sense planificació; manifesten els seus moviments de manera dinàmica i inesperada. L'anàlisi del comportament humà a les ciutats modernes i desenvolupades requereix que aquestes reunions es produeixin en centres comercials, cruïlles de carreteres i flash mobs.

L'anàlisi macroscòpic de la densitat de multituds basada en dades de les torres de telefonía mòbil serveix com a primera etapa per descriure el marc DUCSIM. El mètode M-o-M s'adopta per a la robustesa, ja que aquesta anàlisi implica umbrals de recompte de multituds diaris i setmanals. Les densitats de multitud es classifiquen en quartils per mostrar diferents graus de distribució de multitud. Mitjançant l'anàlisi macroscòpic, el marc avança cap a l'anàlisi de la mobilitat acumulat de multituds. La dinàmica del moviment de la multitud es mesura canviant els senyals de les torres de telefonía mòbil i formulant un mapa de densitat de la multitud per preveure els seus moviments posteriors.

Examina el microanàlisi del moviment individual i les relacions interpersonals a menor escala. Inclou assignar persones a torres de telefonía mòbil i formar gràfics d'interacció social que dedueixin i actualitzin les relacions socials.

La part més important de DUCSIM està en la seua capacitat per aprendre i adaptar-se de manera dinàmica per crear un model de representació nou que s'adapte al patró

recentment detectat. Aquesta flexibilitat ajuda a garantir la rellevància del marc, que s'ha d'actualitzar contínuament.

El modelatge predictiu personalitzat es combina amb les dades històriques que engloben la tesi. El marc utilitza dades de moviment i densitats de multitud anteriors per descobrir tendències i predir les properes dinàmiques de multituds, millorant així l'eficiència de la planificació urbana, la resposta d'emergència o les ciutats intel·ligents.

El marc DUCSIM proporciona un mètode complet, flexible i de previsió per entendre i controlar els fenòmens d'aglomeracions urbanes. Una forma moderna d'anàlisi de dades que inclou diverses fonts de dades, amb el suport de matemàtiques rigoroses, fa que aquest mètode sigui únic per als estudis urbans. A més, dóna un impuls a l'àmbit acadèmic i ofereix recomanacions pràctiques sobre l'aplicació d'aquesta metodologia en la gestió i planificació de la ciutat moderna.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ALGORITHMS

CHAPTER 1: INTRODUCTION

Modern-day urban areas are diverse living systems with significant human dynamics determining short-term and future growth patterns. Understanding and controlling these processes is essential for effective urban planning, emergency management, and the long-term success of smart city initiatives.

1.1 Background and Motivation

The trend in research on urban crowd dynamics has changed recently. Previously, qualitative research was used as the primary study method since no reliable computational techniques or rich big data sets were available [1]. However, using quantitative methods such as improved computational techniques is becoming a matter of concern with the rapidly increasing population in urban settings and promoting a green city [2]. Mobile communication and GPS tracking technology are advancements that allow the monitoring of crowd movements, hence the need for customized modelling methodologies [3]. This is the motivation of this study, given the need for informed decision-making in managing cities given the dynamics experienced in global urbanization [4].

1.1.1 The Emergence of Crowd Sensing in Urban Environments

Crowd sensing is revolutionary for data gathering from urban areas using the built-in sensors in numerous interconnected machines. As such, this approach quickly became one of the core elements in planning and managing smart cities, making it possible to estimate the nature of urban areas' physical traits automatically. Crowd sensing utilizes AI techniques to process large amounts of data derived from urban sensing, leading to better informed and reactive urban management [11]. Crowdsensing finds major use in urban traffic management in estimating vehicular density and distribution. Understanding and managing traffic jams planes and improving urban mobility is critical. Therefore, crowd sensing determines the fastest and safest automobile routes by looking at information collected from multiple locations, enhancing productivity and slashing traveling duration [6].

The mobile crowd-sensing system will improve traffic management and provide safe and efficient mobility across urban facilities [4]. Energy consumption, surveillance, safety, emergency preparedness, and so on are especially useful in crowded cities. [5,6] The traditional systems tend to deteriorate. Data is collected for driving style analysis and environmental details for traffic management and air improvement through vehicular crowd sensing (VCS) , targeting information from intelligent automobiles [5]. VCS innovations such as ROUTR propose a budget-aware, distributed strategy for enhancing VCS efficiency without central coordination. These innovations also increase the reachability of urban crowd sensing [7]. Figure 1-1 represents the flow of information and processes in crowd sensing within urban environments:

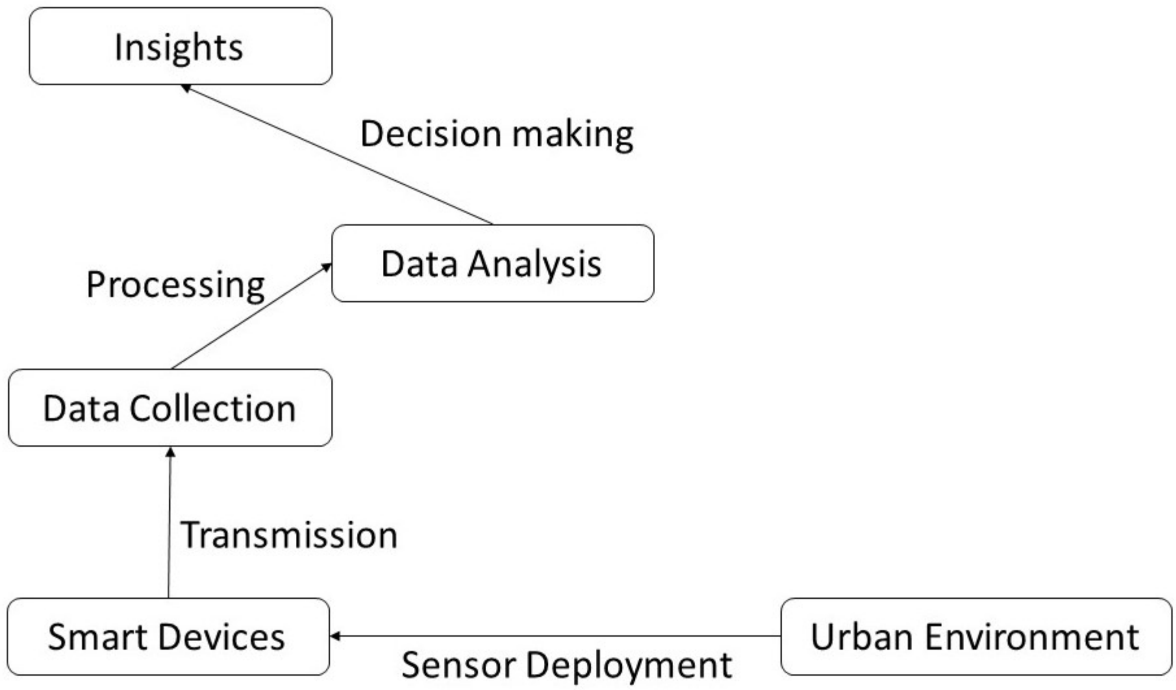


Figure 1-1: Urban Environment Crowd Sensing Scenario

- **Urban Environment**: Urban crowding may see many people participating in urban crowd sensing because many activities occur within urban centres.

- **Smart Devices:** An urban landscape uses smartphones and other devices like sensors. Such tools may help supervise different parameters of city development processes, such as urban flows, including transport situation assessment.
- **Data Collection:** The data that feeds them is provided through these smart devices. Data is transferred from all devices or systems to one centralized system or database. Data transmission is done exactly and reliably.
- Data Analysis: The concept then processes and analyses its data. Data analysis techniques are used to discover the significance of smart device data at this stage.
- Insights: The last step is deriving insights for the analyzed data. Such observations and perceptions could aid decision-making in a city developed for enhanced public services and improved living standards.

Urban crowd sensing involves key procedures such as sensor deployment, communication data handling, analysis, and decisions highlighted in the figure above. Smart cities and crowd sensing have become key technologies in smart city programs. A huge step in collecting and analysing urban data is its capacity to tap into a network of sensors sprawling across numerous linked devices. The city, however, should become more adaptive by using ICTs to achieve a smart environment.

1.1.2 Accurate Crowd Density Estimation

Application areas such as surveillance, traffic management, and public safety heavily rely on crowd estimation. Because crowd density is complex and important, many inventive techniques have been developed to improve the accuracy and dependability of those estimations. One strategy is the Selective Ensemble Deep Network Architecture, which aims to increase population census map densities and precision. Thus, this approach alters the balance of the density and counting loss weighting factor toward a better crowd estimate [8]. Another development incorporates RGB images and RSS signals from Unmanned Aerial Vehicles (UAVs). Hence, this coupling reaches a new record level of accuracy for understanding complex mob scenes [9]. Multi-task learning using depth information of a novel Crowd density estimation framework based on a Unified System. The technique acquires a further dimension, increasing its sensitivity toward crowd studies [10].

Another edge computing application is crowd density estimation, which uses residual bottleneck blocks and dilated convolutional layers to achieve high accuracies with low computing/storage overheads suitable for real-time applications requiring minimum delays [11]. A multi-scale Feature Adaptation Network is proposed to take precise crowd counts with variable scales. The described method prevails over conventional crowd-counting algorithms dealing with scales [12]. This demonstrates continual improvement toward perfecting crowd density estimates using advanced neural network models and modern tools, including robots (UAVs) and mobile devices based on edge computing. The studies above help understand the intricate urban crowd milieu and are pivotal to urban planning and security.

1.2 Evolution of Opportunistic Sensing Technologies

In the context of Cognitive Radio Networks and other wireless systems, opportunistic sensing technologies have achieved noteworthy efficiency, accuracy, and flexibility improvements by applying sampling techniques that use resources sparingly and adaptable. For instance, simple spectrum sensing employing DFT random sampling in conjunction with the Energy Detector model in CRSs provides an optimized ROC curve, minimal false alarm probability, and maximized signal-to-noise ratio [13]. Using Random Spectral Sampling instead of traditional Compressed Sensing (CS) in wide-band spectrum sensing is possible for compliance monitoring. This option can help simplify processes related to collecting data on compliance in broad-spectrum spaces [14]. Recent improvements have been in energy-efficient VLSI signal processing for wide-band spectrum sensing. For example, applying different multitap windows and fast Fourier transform (FFT) processing allows us to minimize spectral leakage and detection time. Spectral leakage on a channel-specific basis allows for exact wide-band observation under short sampling intervals, which is advantageous in cognitive radios. Extended opportunistic sensing is also applicable in smart cities where people-centric computing and communications are present; this will be through an application called “SmartCitizen.” The proposed approach includes time series analysis, opportunistic sensing, deep learning, and transfer learning, thus achieving an overall F1-score improvement of about 5.8% compared to the baseline [14][15]. Acceptance of such opportunistic sensing techniques is important while dealing with spectrum scarcity, ensuring conformity, and improving versatility throughout several wireless services. These portray the contemporary communication model's beginning and reflect a people-cantered approach [16].

1.3 Heterogeneous Data Integration

Heterogeneous data integration, a complex process merging diverse datasets, faces challenges due to the vast quantity, varying quality, and high data arrival speed in the era of big data. These complexities arise from multidimensionality, intertwinement, nonlinear dynamics, and inherent unpredictability of complex, uncertain, and high-dimensional data. Innovative approaches like GANs, with their generator and discriminator architecture, effectively handle data disparities, even in label less or structureless data [18]. The peak clustering algorithm excels in analysing data with numerous peaks, such as biological signals and high-throughput data. Clustering these peaks enhances the interpretation of underlying patterns and relationships [19]. The Heterogeneous Multi-Layered Network (HMLN) model adeptly processes multi-dimensional data, which is vital in bioinformatics, where data originates from various biological layers or systems. HMLNs enable the seamless integration of diverse biological data, revealing the hierarchical nature of organismic systems [20]. Matrix factorization, random walks, knowledge graphs, and deep learning are pivotal in biological information integration. These approaches facilitate the discovery of new biological associations, obscure relationships, and trends within HMLNs [21]. Heterogeneous data integration finds applications in optimizing power grid effectiveness, gathering consumer behaviour, weather, and grid data [22], and aggregating wearable, medical, and genetic data in health and fitness. Advances in these techniques significantly contribute to handling the modern data deluge, benefiting diverse industries with informed decisions and effective problem-solving mechanisms [23].

1.4 Computational Methods for Social Analysis

New computational methods have revolutionized social studies, impacting mental health, finance, and behaviour through big data and advanced algorithms. Social media data analysis can uncover insights into mental health, detecting patterns in language usage and interactions that signal conditions like depression and anxiety, enabling early intervention

[23]. Reinforcement learning algorithms have transformed decision-making in finance, adapting strategies based on market data to optimize investment decisions, portfolio management, and risk assessment [24]. In behavioural and social impairments, including autism spectrum disorder (ASD), computer vision integrated with mobile technology captures behavioural patterns and social interactions, aiding objective assessment, early diagnosis, and personalized care support [25]. These computational methods provide new perspectives on analysing social activities and trends, harnessing big data trends, machine learning, and advanced algorithms [26]. They enhance understanding of social processes and positively affect diverse fields such as health, finance, and behavioural sciences [27].

1.5 Mobile Network

A number of studies have looked into crowd density estimation and prediction through using mobile networks. A study was conducted using a sub-GHz WSN for crowd density estimation and the results proved that it had an excellent estimation performance [28]. Another research has suggested lightweight density estimation network architectures like the GAEnet for highly accurate and real-time applications. Such solutions address the problem of excess parameters and redundant structures in conventional crowd-counting networks [29]. There have also been studies that use light Wi-fi signals from sparse infrastructure to estimate crowds with respect to privacy of individuals monitored via traditional network. This research has proposed the use of multifeatured convolution neural network (MFNet) technique for crowd density counting that incorporate several crowd data sources such as hog, lbp, and canny. This shows how the technologies used in these approaches can be applied in calculating the densities, designing better communication services, controlling traffic, and others [30].

Analysis of Mobile Network for Crowd density estimation and mobility, is based exclusively on logs data, which provides an efficient way of revealing population movement. However, this approach simply explores the fundamental characteristic of mobiles as devices that can communicate with neighbouring transmitters regardless of the brand (2G, 3G, 4G, 5G, and 6G), call making, SMS and other features. The size of data that is generated per hour is massive to handle, thus the purpose of this research is to explore more by using less data in number of parameters. Cellular phones constantly ping (communicate) with the cell towers in order to stay connected on a mobile network. The position of the individual in a crowd, and even the direction towards which it is moving can be assessed by monitoring which mobile device towers they are connecting with, and determining how these connections develop during the course of time. This information is very important to understand the daily movement of people in a given region while they are going about their daily activities, conducting special events, and even evacuating in case of an emergency. It is a continuous and real time location source data since the system traces device's interactions with cell tower; even in an idle mode.

The proposed system estimates cumulative crowd densities and distinguishes among single individuals and grouped movements while assuming every person carries a phone. The data is of great importance to the town planners, the event managers, and emergency service providers as it offers insights towards creating appropriate and reliable schedules that lead to efficient traffic control, and improved and reliable urban infrastructure. It uses anonymous, aggregated data that is based on network connection and this helps to protect the individual's privacy by giving a general picture of the dynamics in the population. For the thesis, the MOBILE network focuses on the number of Mobile Phone devices connected to the BSC. Figure 1-2 illustrates an overview of a MOBILE Network.

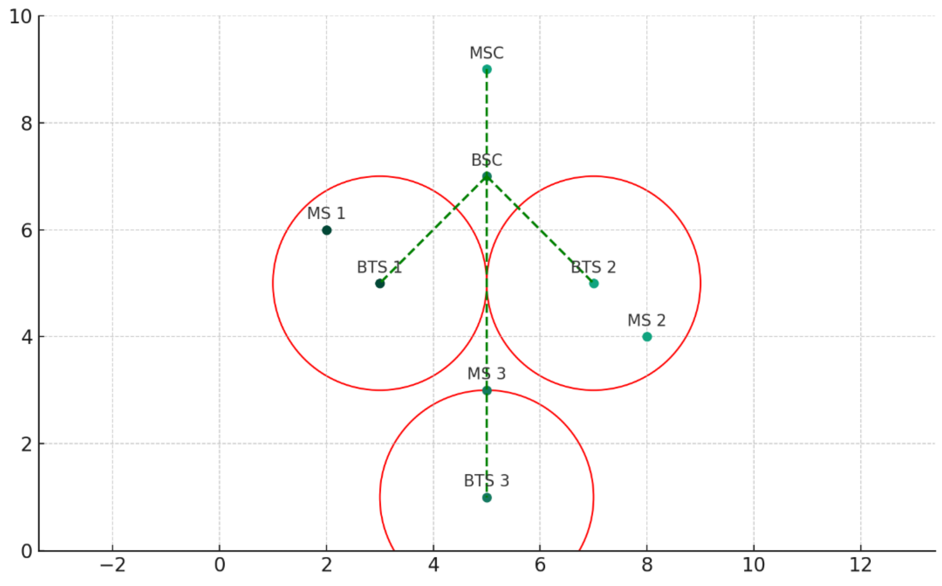


Figure 1-2: Over view of Mobile Network

Here is an illustrated overview of a basic MOBILE network:

- **Red Dots (BTS - Base Transceiver Stations):** These are the base stations distributed across the area, each with its coverage radius (indicated by the red circles). They are responsible for communicating directly with mobile stations (MS) within their range.
- **Blue Dots (MS - Mobile Stations):** Represent mobile devices or users within the network. They connect to the network through the nearest BTS.
- **Green Dot (BSC - Base Station Controller):** This controller manages several BTSs. It handles the set-up, frequency allocation, and handovers (when a mobile station moves from one BTS to another).
- **Yellow Dot (MSC - Mobile Switching Center):** This is the central component of the network. It connects to the BSCs and manages the routing of calls and data and interfacing with other networks.
- **Dashed Lines:** Indicate the communication links. The green dashed lines show the connection between the BTSs and the BSC, and the yellow dashed line represents the link from the BSC to the MSC.

1.6 Statement of the Problem

1.6.1 Defining Crowd Density Estimation Challenges

Crowd estimation poses several intricate issues, especially when it has to do with changing environments. Accurate assessment for crowd formation is very difficult as it is spontaneous with an unexpected group action. Adding to the complexities is that data must be collected and analyzed within varied environmental contexts, from open spaces to indoor venues. Additionally, crowd dynamics can be affected by various unexpected developments such as sociological occasions, emergencies, and weather events, making real-time prediction even harder.

1.6.2 Limitations of Existing Methodologies

The existing techniques of crowd density analysis mostly depend on old-school approaches, which include CCTVs and person counts. However, these methods have a short range as well as limitations of scalability—these impediments in real-time processing, privacy problems, and demand for a large physical facility. Furthermore, the existing traditional algorithmic ways might fail to utilize completely the vast range of heterogeneous data spectrum we have nowadays in this integrated digital world, giving rise to the gap between data possibility and its efficient usage for crowd analysis.

1.6.3 Need for a New Algorithmic Approach

For a new algorithmic paradigm to address the shortcomings of current methods. Therefore, The novel approach should combine diverse datasets such as MOBILET tower information, online activity on social networks, and other tracking information to provide a holistic and instantaneous picture of crowd behaviour. The software must also be flexible enough to fit different situations and adept at handling large volumes of data prevalent in cities.

1.6.4 Challenges in Opportunistic Environments

On the other hand, unplanned or opportunistic gathering situations are peculiar challenges. This encompasses fleeting gatherings with changing compositions, unstructured data on these events, and fast-moving crowds. For instance, such cases require flexible and strong systems to change rapidly to changing situations while making accurate estimations with no need for prior installation or extensive data-gathering campaigns.

1.6.5 Role of Synthetic and Real-world Data

Synthetic and real data represent an efficient approach to address these issues. Synthetic data allows the creation of crowds and different behaviour models in an artificial environment. It makes it possible to test algorithms on them and fine-tune them before real use of algorithms is possible when real-life data gathering would be impossible. On the one hand, facts obtained from cellular telephone records and the IoT or social networking sites reveal real-world information about crowd behaviouralU patterns and trends. The proposed algorithmic approach integrates various data sources that balance theoretical robustness with applied relevance to achieve precision and reliability of crowd density estimates in different and dynamic situations.

1.7 Research Objectives

1.7.1 Primary Goal: Development of the DUCSIM Algorithm

One main objective of this study pertains to the creation of the Dynamic Urban Crowd and Social Interaction Model (DUCSIM). This algorithm has been designed for this thesis to change how the crowd densities could be estimated within any diverse situation of chance. Aim for crafting an algorithm that will capture crowd motion's intricacies and adjust to the rapidly shifting cityscapes. DUCSIM will use advanced analytical procedures and data inputs for instantaneous, precise crowd behaviour and density.

1.7.2 Objective 1: Validation with Synthetic Data

The main goal is to validate the DUCSIM algorithm based on synthetic data. It entails developing virtual crowds and situations that emulate actual crowd behaviour. This validation process ensures that DUCSIM can provide reliable modelling and predicting of crowds' behaviour in structured environments. This is an important step to evaluate the algorithm's effectiveness and modify it appropriately before it can be applied to real-life data to ensure the results are trustable and sustainable.

1.7.3 1.3.3 Objective 2: Comparative Analysis with MOBILE Data

The second goal considers implementing DUCSIM's simulation efficiency with MOBILE (Global System for Mobile Communications) information. This is where different crowd density estimations from DUCSIM are compared to the data retrieved from the MOBILE towers. The main objective of this research is to evaluate the application's performance in practical settings and its compliance with the extant information sources. This step is vital in establishing the algorithm's practical applicability and effectiveness in urban settings.

1.7.4 Objective 3: Algorithm Generalization Using Multiple Data Sources

This study aims to generalize the DUCSIM algorithm and its utilization in various datasets. Data from social media, IoT devices, and others should be integrated into MOBILE algorithms to increase the algorithm`s precision and flexibility. This objective seeks to develop an all-encompassing framework that accommodates different data configurations, consequently rendering DUCSIM applicable for the studies of any city.

1.7.5 Objective 4: Real-world Applications and Implications

The last objective entails investigating and exhibiting actual field uses and impacts of the DUCSIM. Applying the model to simulate practical urban settings and scenarios to validate its efficiency and applicability in modern-day case studies, where it is used in urban planning and public safety, including event management. Hopefully, this study will show how DUCSIM can contribute to smarter city projects and help manage space within cities to make people safer and happier. This stage is key in presenting the real application and consequences of the research, signifying that it is more than a theoretical concept.

1.8 Research Scope and Limitations

1.8.1 Scope of the Study: Data Types and Sources

The research incorporates different data sets and sources majoring in electronic trails emanating from metropolitan communities, such as crowd mobility and density inferred through the usage of mobile phones indicated by MOBILE Tower Data. Moreover, the study uses information retrieved from social media, which depicts a live scenario of crowds and sentiments of different groups. Additional data streams, such as IoT devices and online traffic, are also important. The goal is to integrate different data sources to gain an overview of the behaviour of crowds that is common in cities.

1.8.2 Technological and Methodological Boundaries

State-of-the-art technologies, data analysis, and crowd simulations constrain the research. The program operates on self-learning using modern data science algorithms and processes massive heterogeneous databases. Nevertheless, the analysis recognizes the limitations dictated by present technology capacity; this includes computational proficiency, information systems’ storage space, and analytical equipment capacity. In terms of methodology, the research follows accepted practices for information science; however, it is restricted by the existing knowledge and models in crowd movement analysis.

1.8.3 Geographical and Temporal Scope

Research is mainly carried out in dense, populous urban areas, and the technological framework ensures ample information data. Nevertheless, this research does not focus on a specific city; hence, it may apply to different urban settings and cultures. As for temporality, the study is based on modern situations mirroring today's cities' reality and technologies. However, this temporal approach will give the study a current flavor that may render it irrelevant for future situations likely to involve new technology and social environment.

1.8.4 Limitations in Data Availability and Quality

The reliability and quality of the data used in this study are among the serious limitations of this research. Variations in data quality are brought about by inconsistent data collection means or approaches applied in different regions and platforms. Additionally, problems with data, such as a failure in technology or lack of information on some regions, would reduce the completeness of the analysis. The research also admits that the truthfulness of its results depends on the data quality and its representative nature.

1.8.5 Ethical Considerations and Privacy Concerns

This research abides by ethics that protect people's privacy regarding using their private information during the study. This study includes anonymization of data to hide the identity of particular people. Nonetheless, the work acknowledges that it is impossible to deanonymize high-resolution databases even through full deanonymization of data. On the other hand, this study examines moral concerns associated with crowd booking and its consequences regarding how people view fairness and trust. These factors determine how the research should be undertaken and the scope of use of specific data to maintain ethics and respect for privacy rights.

1.9 Methodological Framework and Innovations

This part of the thesis discusses the theoretical basis for constructing and implementing the DUCSIM algorithm and how this study is unique compared to others in the field.

Innovative Approach to Data Integration: This study's cornerstone is using various datasets to estimate the crowd densities correctly but accurately. DUCSIM is an unusual algorithm that combines information from cellular companies' towers, social media networks, IoT devices, and numerous digital systems. The multi-source approach that integrates several types of data on crowd dynamics takes this understanding one step further. Algorithmic Development: Developing the DUCSIM algorithm constitutes a revolutionary methodologic step. The algorithm has been designed for high-speed processing of vast data using sophisticated computing procedures. It analyses and forecasts the behaviours of crowds through machine-learning models and statistics. It stands out as one of the most notable aspects of this algorithm in that it performs well in various urban set-ups and can be expanded to many locations.

Synthetic Data Validation: Synthetic data also plays a role in algorithm validation as another innovation. The simulation creates the ideal environment and situation for testing and improving the algorithm under controlled circumstances. Firstly, this way makes DUCSIM more reliable before its use for real-world data. This is contrary to conventional crowd density estimation. Real-world Data Analysis: The thesis also raises a new bar regarding utilizing advanced algorithms on the actual data. Comparative analysis with MOBILE data gives an insight into the practicality of DUCSIM. The study also demonstrates that the algorithms can be effective in urban settings as they compare the algorithm-generated estimations and actual MOBILE data.

Algorithm Generalization: Generalizing the DUCSIM algorithm across different data sources marks an essential methodological milestone. This is a useful approach since this algorithm is flexible enough to handle many forms of data and street systems.

Ethical and Privacy Considerations: Another important element of the methodological framework is addressing ethical and privacy issues. The study employs strict measures for data anonymization and ethical practices to safeguard participants' privacy and ensure that these issues are sufficiently catered for. A careful approach to ethical data management provides a positive model for future research.

Application and Societal Impact: Lastly, the approach goes beyond mere analysis to include the implementations of the algorithm's DUCSIM and its societal implications. The practical aspect of this research is demonstrated by the fact that it covers all the implications of crowd density estimation for city planning and safety. This theoretical orientation comprises newly developed concepts concerning data harmonization, algorithm creation and verification, and application. This helps advance crowd density estimation and greatly contributes to urban studies data analyses.

1.10 Expected Contributions to the Field

This study will contribute to urban studies, data analysis, and crowd control. The contributions are diverse as they go beyond theories of practice.

Advancement in Crowd Density Estimation Techniques: This research has contributed significantly to developing methods for estimating crowd density. Unlike most current single-data-source methods, the DUCSIM algorithm presents a modern solution for integrating various dataset types such as MOBILE information, social networking data, and Internet of Things sensors' measurements. A methodological approach will allow researchers to create a model closer to reality.

Methodological Innovations: There is a set of innovative methods, especially related to data processing and analysis, at different stages of the research process. Applying machine learning techniques and statistical models to manage huge and divergent data sets is one of the milestones made over the years. Furthermore, artificial data used to validate algorithms has improved the credibility and integrity of the results.

Enhanced Predictive Capabilities: Another key contribution of DUCSIM is its predictive modelling capabilities. The algorithm offers vital means of forecasting crowd movements and densities, thereby being useful for planning and urban management, particularly because of developing a more efficient urban environment where issues of emergency management response planning, events coordination, public space design, and others should incorporate these predicting features.

Scalability and Adaptability of the Algorithm: Notably, the scalability and adaptability of the DUCSIM algorithm in various city contexts and types of data is an important breakthrough. Such flexibility provides room for the application of the algorithm in various urban set-ups, making it a useful tool for urban scholars and designers worldwide.

Addressing Ethical and Privacy Concerns: Additionally, the study helps contribute to the domain by tackling ethics and privacy issues in analytics. The study establishes ethically acceptable means of using urban data through strict data anonymization practices and ethical guidelines.

Practical Implications for Urban Management: These highlight critical contributions towards its practical application in urban management and planning. The computer program is vital as it offers the city planners, policymakers, and respondents to emergencies with genuine and latest details about the flow of the crowd. Contribution to Academic Discourse: The thesis adds to the scholarly debate by thoroughly investigating crowd density estimation. This work contributes to current research by expanding knowledge on urban dynamics and crowd-related phenomena, thus adding value to the academic understanding of these matters. Therefore, the expected contribution of this thesis is very broad, having several theoretical and applied consequences for urban management. The contributions will be long-lasting and provide a foundation for follow-up work in urban studies and data analytics.

1.11 Significance for Future Research

A new method, “the DUCSIM algorithm,” is proposed in this thesis. It can be used for further research in many areas. Application of the algorithm in crowd density estimation is a basis for thorough studies in crowd behaviour and dynamics that involve various data sources and developed analytical methodologies. As such, this study can extend the scope of predictive modelling practices in urban planning, disaster management, and urban redevelopment initiatives, leading to better urban administration schemes. Future work must increase the algorithm’s flexibility, scalability, and agility to varying cities and datasets. More research is required regarding the ethical use of data in urban locations with increased surveillance and information-gathering levels. The DUCSIM algorithm has a multidisciplinary feature, allowing multi-faceted research that bridges data mining, social behaviour, and urbanism. In the long run, it has smart city applications such as urban security, traffic control, and resource utilization. Moreover, it adds to urban studies' theoretical frameworks and data analysis on human mobility in different urban conditions. Ultimately, the DUCSIM algorithm can form a reference framework supporting further data mining attempts within urban studies and planning.

1.12 Thesis Structure and Contribution

The highly systematic thesis examines the evolution and justification of the DUCSIM population estimation model in a dynamic distributed environment. They are made sequentially to ensure that the contents in each chapter are logical and relevant to what has come before and the whole research findings.

Chapter 1 is the introductory chapter, forming a basis for the whole work. The study also lays down the research topic, the objectives it aims to resolve, and the relevant boundaries under urban studies and data analysis.

Chapter 2 provides an extensive review of methods related to crowd density estimation, including various technologies already in use and other gaps that call for DUCSIM.

Chapter 3: The technique of development and testing of the DUCSIM algorithm is disclosed, along with information on the data sources collected and analyzed. It outlines the method of validation of synthetic and observed data.

Chapter 4 focuses on constructing the DUCSIM algorithm, which consists of design principles, theory bases, and computer simulations. It offers knowledge of data processing and crowd density calculation.

Chapter 5 is on validations using synthetic datasets by comparing the performances of algorithms with real-time MOBILE data and measuring the precisions, accuracy, and robustness.

Chapter 6: the practical usage and implications of the DUCSIM model are highlighted, demonstrating its applicability to urban planning, security, and smart cities. The ethics of advanced data analytics in crowd surveillance is another issue addressed.

Chapter 7 summarizes the main ideas, explains how our results contribute to urban data analytics and pedestrian density estimate, and suggests future investigations to improve DUCSIM's potential uses in different urban situations.

This thesis-based study extends the frontiers of urban data analytics and population estimations, adding new knowledge and implementation practicality concerning several urban contexts.

1.13 Conclusion

The chapter of introduction lays down an appropriate basis for examining use of connection logs with cell tower in estimating crowd density as well as other measures linked closely with it. This thesis will explore how mobile devices not only can be used to gather current geographical information but also continuous location data. Such information forms an integral part of development of the urban plan, traffic control as well as assisting in management of emergencies. Using anonymous and aggregated data ensures privacy for people; meanwhile, it gives valuable information about the urban system operation. In the following chapters, illustrate how analysis of mobile network data can contribute to better urban planning and governance, reinforcing the need for more comprehensive use of telecoms for public good.

CHAPTER 2: LITERATURE REVIEW

Modern urban governance and safety hinge on crowd density estimation, including technology, sociology, and urban planning. The subsequent section briefly discusses its history, background, and the change from traditional to present technology. This provides the theoretical basis and important terminologies for creating a common ground or understanding of this area. We discuss technological changes in methodology by comparing old and new approaches in terms of improvement. Modern methods remain constantly challenged as they evolve with new challenges and techniques. The case studies demonstrate the use and reveal important implications. Other chapters discuss crowd sensing with opportunistic environment settings and utilize both synthetic and real-world data, look at social network analysis with crowd sensing in mind, consider ubiquitous computing, and review existing models and algorithms. We have thus adopted a comprehensive approach to ensure that scholars and practitioners fully understand the topic in question. This brings together historical perspectives, underlying issues, and emerging issues, which help us understand this crucial topic.

2.1.1 Overview of Crowd Density Estimation

Safety in a crowd requires addressing factors such as occlusion, resolution, and varying lighting conditions [31]. Crowd density estimation has been significantly upgraded through convolutional neural networks with high-resolution density maps [32]. Crowdedness estimation is a fundamental aspect of subsections such as crowd counting, tracking, and behaviour recognition. Crowd counting estimates the number of people via either direct (number-based) or indirect (trait-based) means, necessary in crowd management and for ensuring the public's safety [33]. Crowd tracking involves noting how people disperse over some time through counting. The procedure supports the management and detection of movement abnormalities. Crowd behaviour recognition determines and assesses crowd activities needed in various scenarios, such as managing events, responding to emergencies, and making proper decisions [34]. Problems associated with environmental conditions and the type of crowd result in challenges such as counting crowds, crowd density estimation, crowd tracking, and incident detection for datasets like PETS 2009[36]. For example, performance evaluation compares bipartite graphs, adjacency matrices, and node-link diagrams on different suitability levels. Although improvement has occurred, obstacles like occlusions, low-resolution variation, and light still occur, calling for more advanced systems to enhance safety and security among many people [37, 38].

2.1.2 Historical Evolution

Deep learning has revolutionized crowd counting and density estimation because of computer vision's advancements. Previous work traditionally used representation-based approaches and regressions, but these were not flexible enough in particular cases, such as those requiring large crowds. This move to newer techniques represented a major leap forward for the industry. A breakthrough in 2017 gave rise to a stacked CNN that effectively learns the number of people for counting concurrently, together with estimating density map information. The method yielded fewer inaccurate counts than present-day best practices, indicating that precision is preferable when estimating density maps [40]. In 2019, another advance was made with the advent of TEDnet aimed at improving density estimation maps – one of the major challenges in the domain. Among others, TEDnet exceeds other models in number correctness and density map performance, proving it an effective model for different crowd number estimation situations. Some recent studies are directed at using DRL for crowd number estimation and adding elements from learning-driven reasoning. DRL can help improve MAE, an essential measure of crowd density's accuracy. The evolution of this process represents modern crowd analysis methodologies that shift to adaptive and responsive models. Continued refinements show we are improving crowd counting and density estimation in preparation for new crowd behaviour analysis trends using deeper learning models. The field is still alive; hence, more groundbreaking can be made.

2.1.3 Theoretical Foundations and Key Definitions

Crowd counting is an important research topic in computer vision that can be used for traffic monitoring, urban planning, event management, and disaster preparation [38] [39]. Deep learning techniques have advanced crowd counting. In the past, counting heads in a congested area or an agile crowd was challenging [40]. Density estimation, an extended version of counting, will help understand the general movement patterns of the crowds and zones prone to overcrowding, thus enabling efficient control of the crowds. These advancements are greatly related to Deep Learning Image Processing, indicating the importance of the resolution of source images. This leads to specific assessments for crowd density in high-resolution images [41, 42]. Several crowd-counting approaches have emerged: Deep Learning-Based Single Image Crowd Counting: Here, a new approach to estimating crowds' numbers is used. In this novel method, these modern deep-learning models can extract difficult-to-understand details from one photograph [43]. Spatiotemporal Attention Convolutional Neural Network for Video Crowd Counting: Regarding the changing crowd information, it considers spatial and temporal dimensions, relying on attention mechanisms inside CNNs to classify different sizes and dynamics of crowds [43]. Crowd Counting Using DRL-Based Segmentation and RL-Based Density Estimation: A combination of DRL for segmentation and RL for density estimation allows the selection of appropriate models or methods based on the complexity of scenes, image quality, and requirements of the particular implementation task [44]. Crowd analysis uses the latest algorithms, like encoder-decoder algorithms and attention modules, which are custom-made depending on different crowd analysis needs. The technique selection is based on issues like scene intricacy, information quality input, and particular use case necessities [45]. These improvements allow a better understanding of crowd psychology, which is useful for safe crowd control.

2.1.4 Traditional Methods vs. Contemporary Approaches

Traditional crowd counting and density estimation approach was based on manual technique. However, modern deep learning-based approaches have replaced these methods since they are slow, labour-intensive, and error-prone. Some present-day techniques, especially using CNNs, provide scalability, increased accuracy, and better performance. CNN-based algorithms are efficient in image processing, facilitating accurate calculation of a crowd count and a density-based representation derived from videographic or photographic information [44]. However, recent research has improved these CNN-based methods by using Bayesian crowd-counting losses that provide a probabilistic perspective on crowd estimation. Spatial resolution is improved by pixel modelling in density estimation, while real-time video data analysis gives information on the dynamics of people's movements [45]. Some of these improvements are important in scenarios like mass events for monitoring crowds to guarantee accuracy and efficacy, for instance, calculating capacities in public places or addressing public health issues [46]. Precise estimation of crowds and their densities is essential to efficient crowd management, disaster readiness, and town planning. This is a revolution to crowd counting, from manual techniques to deep learning approaches, which are still used today [47]. Today, these sophisticated computing techniques are more relevant in addressing societal issues. Research shows that new versions of these methods will improve their accuracy, practicality, and effectiveness in different fields.

2.1.5 Challenges and Advancements in Current Techniques

The latest IoT and complex algorithms techniques have greatly improved crowd density estimation, crowd estimation, and steer/crowd but come with new problems. Crowd analysis has been modernized with an innovative model, making it more efficient and accurate. One innovative method utilizes a based system for estimating the crowd density as it gathers necessary data to project the crowd intensity. Using techniques such as the GRAY LEVEL CO-OCCURRENCE MATRIX improves the extraction of features and regression, enabling continuous data monitoring and analytics in dynamic crowd management situations [47]. Using Convolutional Neural Networks and computer vision, Crowd Density Estimation based on Auto-Encoders computes density maps through Gaussian filters. So far, the auto-encoder model has proven successful, better than the traditional MSE approach. They are very good at handling sophisticated visual data and producing precise space density representations. The Visual framework provides a new tool for dense crowds' estimation without any physical device and uses CSI of Wi-Fi signals. It improves local data differentiation, enhancing crowd density estimate resolution at locations. It provides a reliable prediction of crowd density and displays migrating people, which is an effective aid in tracking movements indoors [48]. Crowd Density Estimation from Autonomous Drones Using Deep Learning employs deep learning frameworks and autonomous drone pictures for crowd density estimation. Such are complexity changes, scale, intensity variations, density changes, and height differences. Applying this strategy is appropriate when dealing with open-ended and non-static scenarios, providing an opportunity for other studies on crowing behaviour. The advances show the possibility of using current technologies to analyse crowds. However, it is still challenging to obtain reliable camera platforms; occlusion and efficient pose estimation are also among the existing hindrances. Further research and development would improve the accuracy and efficiency of crowd analysis and management systems [49].

2.2 Opportunistic Environments in Crowd Sensing

Crowdsourcing is a new method for collecting information that utilizes the computing devices we have everywhere nowadays [50]. By definition, this idea encompasses the ability to network the capability of individual desktops for collecting, sorting, and sending information using devices such as smartphones, wireless wearable sensors, and moving car sensors as network nodes [51]. Crowdsourcing improves the relevancy and correctness of data on such parameters as road monitoring and environment [52, 53] in a vehicle sensor network. In particular, cloud-based mobile crowd-sensing systems collect sensory information from widely accessible mobile phones for numerous uses, including weather tracking, activity recognition, sensory cost minimization, and delivery efficiency maximizing, among others [54, 55]. Mobile sensing has a high potential across various areas, such as social networking, health care, environment monitoring, and road safety. For instance, wearable devices in healthcare track vital signs and send data via telehealth systems. At the same time, environmental monitoring indicates air quality measurements and noise levels. However, challenges must be overcome, such as protecting participant data confidentiality and sensor validity when new designs emerge. The future steps might involve optimizing data collection techniques and additional applications. Crowdsensing for opportunistic purposes as opposed to traditional means of collecting and using data [56, 57].

2.2.1 Defining Opportunistic Sensing Environments

Opportunistic sensing uses autonomous mobile devices to get information in large-scale environments. The information spans through numerous sectors [58, 59]. These environments include various forms of data like monitoring office workers, using heat to understand temperature sensations, and even detecting poisonous substances in the air using mobile devices [59]. It has applications in smart cities, facilitating traffic management, environment monitoring, safe work, and increased productivity in the workplace. It helps collaborative spectrum sensing for opportunistic Access, improving

Network efficiency in Telecommunication [60]. Although the information was collected at several remote sites, data accuracy problems and noise are required for authenticity. Future studies will refine data gathering/analytics techniques that include application for urban planning and, for example, environmental surveillance. Researchers have been working on opportunistic sensing for over three years, a new step toward utilizing a large-scale sensor system for increasing intelligence, efficiency, and security [61].

2.2.2 The Role of Opportunistic Data Collection in Crowd Estimation

Much benefit is gained in the crowd estimation processes within mobile crowdsensing and the Internet of Things (IoT) by deploying commonly available mobile devices and IoT sensors [62]. Opportunistic data collection in intelligent transportation generates current traffic information for wise decisions—mobile-based smart sensing technologies like infrastructure-assisted on-demand crowdsourcing gather contextualized accurate data from specific locations [63, 64]. The use of crowdsensing and mobile network functionalities promotes the dissemination of spatial information for enhanced situation awareness of critical incidents in traffic and traffic flow [65, 66]. Utilizing opportunistic sensors in real-time, traffic prediction in urban areas incorporates information from drivers and mobile phone users with other source data to support intelligent management of highways and construction sites [67, 68, 69]. Modern crowd-counting techniques can draw upon opportunistic data acquisition and IoT-enabled mobile devices to advance traffic and crowd monitoring in intelligent transportation and urban IoT systems.

2.2.3 Technologies Enabling Opportunistic Sensing

The concept of opportunistic sensing is relatively new and applies modern management methods for data acquisition in different environments. Fog-based semantic and energy-efficient sensing management; Social opportunistic sensing of a smart city; Participatory and opportunistic mobile sensing for various applications [70]. A fog-based semantic approach for intelligent sensing network management improves energy efficiency by using fog computing to process data near sources, thus lowering the network’s traffic to better handle performance, analyse, and transmit data. Smart cities exploit social opportunistic sensing that generates massive information on urban dynamics, public utilities, and behaviour, enabling urban plans, traffic management, and police work [71, 72]. Mobile sensing can be participatory or opportunistic, with widespread use in environmental monitoring, health care, urban development, etc [73]. The opportunistic sensor is in line with the progress of communication networks, especially in the emerging 6G system, where it offers increased power efficiency for devices, integrated social networking, and user-cantered mobile applications for enhanced data gathering and management, improving efficiency and timely in several tasks as [74][75].

2.2.4 Analysis of Data Quality and Reliability in Opportunistic Settings

Data quality and reliability are crucial in opportunistic settings, i.e., network communications and sensor networking. Several studies address these challenges from different angles: Reliability Enhancement in Mobile Ad-hoc Networks: The selection of a forwarder with high path reliability, low routing load, and minimum distance to the destination makes RE-oR reliable [76]. The message's reliability is guaranteed, especially during the transitory-natured ad hoc networks. Opportunistic Citizen Science Data: Massive data on species distribution is generated by citizen science programs like eButterfly. It is important that any ad hoc data used can be confirmed as being both valid and accurate. Citizen science data is useful but may not reflect regional species richness precisely, hence the need for an extra testing procedure [77].

Reinforcement-Learning-Based Opportunistic Routing Protocol for UASNs: The ROEVA routing efficiently reduces power consumption and increases the reliable delivery of messages besides addressing uncovered path issues in an underwater acoustic network [78]. The reason for developing such protocols is because of underwater communication challenges like energy, delay, link quality, and depth information. Maritime Search and Rescue Wireless Sensor Networks: The opportunistically routed protocol with maritime SAR optimizations delivers low latency, affordable costs, and minimizes power expenditure. The problem focuses on maritime particulars, including low node power and mobility (very time-sensitive) [79]. Energy-Efficient Clustering and Cross-Layer-Based Opportunistic Routing Protocol for Wireless Sensor Network: CORP is a protocol that uses clustering, routing, and filtering techniques to improve data-gathering efficiency for wireless sensor networks. This enhances data accuracy while minimizing energy costs, making it sustainable and efficient. These studies show how difficult it can be to ensure reliable data on different networks and applications in opportunistic scenarios. This field continues to grow to improve network reliability and security.

2.2.5 Case Studies: Crowd Sensing in Opportunistic Environments

Opportunistic networks have seen a new concept, namely, crowd sensing, emerge in pervasive computing and wireless networking. It uses the connectivity characteristics of personal digital devices such as smartphones and tablets to share and compile information across different environments [79]. This method multiplies the crowds' collective power, resulting in a richer data source. Applications Spanning Domains: Indoor localization finds applications of crowd sensing using mobile phones to map locations and track movements, mainly in places with poor GPS. The tool is very useful for traffic data collection and allows the monitoring of traffic in urban areas in live mode from the gadgets of commuters and pedestrians involved. In IoT environments, crowd sensing improves the performance of connected devices [80].

Challenges in Crowd Sensing: The veracity of data should be the first thing to consider in a situation where there is more than one. Such data must be reliable and exhaustive since it comes from various sources [81]. Addressing data quality concerns, user participation, etc, are part of enhanced mobile crowd-sensing environments. Sophisticated methods of assessing data reliability using the experience and integrity of contributors, including voting-based trustworthiness models and game theory [82]. Future Prospects and Research Directions: Research seeks better sensors with efficiency, quality, precision, and scalability. These include sophisticated data validation methodologies, improved collection methods, and incentive-based reward structures encouraging active participation [83]. The technological developments include 5G and beyond. This provides better quality, up-to-date information [84-85]. Further reading: Opportunistic crowdsensing is a platform that combines information technology, data sciences, and massive engagements for large-scale crowd sensing. Nevertheless, dealing with basic issues should be a priority.

2.3 Advances in Synthetic and Real-world Data Utilization

Opportunistic sensing is used for crowd estimation as it relies on synthetic, empirical, and real-world data, leading to higher precision in practice [86]. Synthetic Data for Enhanced Crowd Counting: It uses synthetic data to enhance real-world crowd counting. Trained by a segmentation network on synthetic data for distinguishing persons from the background, crowd counting networks become more fit in real crowd data and increase accuracy. Training on Synthetic Data for Pose Estimation: On CAD and mainly for training synthetic deep learning models to estimate poses. This helps reduce the process of creating data and makes it possible to deploy this model in the real world [87]. Challenges in Real-world Data Application: Using learning-based approaches on genuine data presents difficulties mainly because of the distribution discrepancies in synthesized and actual-world data features. As a result, innovations such as match normalization and nonlinear least squares fitting are proposed. These innovations introduce new loss functions for 3D object registration using actual measurements [88]. Synergistic Training Methodologies: New approaches include cross-training methods where large amounts of unlabelled monocular footage are combined with supervised learning from synthetic images. They are better than most conventional supervised techniques and can be practically useful [89]. Crowd counting and pose estimation combine to create an opportunistic crowd estimation scheme that is more accurate, efficient, and practical for real-life applications. The ongoing innovations in crowd analysis will lead to constant development as well [90].

2.3.1 Comparative Analysis: Synthetic Data in Opportunistic Contexts

The comparison of synthetic data and actual data in opportunistic settings is fascinating. Recently, synthetic data has become a reliable data set that is easier to control in unpredictable opportunistic environments [91]. In this case, it outperforms controlled experimentation and modelling whereby unique circumstances that cannot be recreated in real life are simulated, assisting applications such as research and disaster management [92]. Overcoming Real-World Data Limitations: Synthetic data has no real-world constraints like privacy or resource restrictions and thus yields meaningful information about healthcare practices and town planning, among others. Enhancing Machine Learning Models: Artificial data plays a crucial role in machine learning, where algorithms require massive and varied source sets to create solid future precursors, mainly in deep learning [93]. Challenges and Considerations: Nevertheless, the authenticity of artificial samples remains challenging, and data are hard to synthesize.

It should be noted that these problems mean the need to correspond synthetic data to real-life situations. Future Directions: GANs and better simulation methods could make synthetically generated data essential in many contexts [94]. Comparison of synthetic data within opportunist contexts can yield meaningful results for a data analyst and machine learning.

2.3.2 Real-world Data Challenges in Opportunistic Environments

Real-world data cannot be employed for opportunistic mobile sensing, as it is collected in an unstable manner across space and time. An advanced spatial indexing strategy like the Geohash-grid tree can support efficient querying and analysis in opportunistic sensing scenarios, exploiting the spatiotemporal behaviour of mobile sensor networks [95, 96]. Handling Streaming Data: Streaming data is common in many real-world applications, making it difficult to have a benchmark comparable to the actual data, thus highlighting the need for relevant benchmarks for measuring and improving the algorithmic performance regarding the management of streaming data [97, 98]. Addressing Real-World Complexities: It is important to note that utilizing real-world data, especially in opportunistic settings, mandates sophisticated techniques and tools attuned to the variable quality of such data. This necessitates ongoing research and development. Space indexing algorithms and in-stream analytics are some methods devised despite real-world data problems, which are priceless information sources [99, 100].

2.3.3 Methodologies for Data Generation and Validation

Several data collection and testing strategies used across various fields are useful in dealing with complicated issues. The automotive industry adopts randomized synthetic data generated using a modular parameter representation involving ranges and distributions. L2 and L3, satellite validation data, reveal good consistency but some uncertainty regarding ocean colours. For quality assurance purposes, standardized protocol validation is needed for next-generation sequencing in veterinary infection biology [101-104]. Analysing photovoltaic-induced harmonic disturbances in low-voltage network safety utilizing measurement and simulation methodology. PLS-SEMs are used by composite-based predictive tools focusing on out-of-sample assessments and offering more sophisticated ways to facilitate empirical estimation [105]. These revolutionary advances in data sciences include self-driving cars, ecological exploration, and bio-medicine.

2.3.4 Successes and Limitations in Data Utilization

This is because of individual considerations specific to different data utilization realms. Among many monitoring systems, they benefit the HPC System's health and performance in operation [106]. Nevertheless, gaps in data collection limit their full power of use, urging us to identify and highlight a set of necessary elements for a functioning monitoring system [107]. There are limitations to its use in daily medical work. Healthcare's financial alignment initiative is characterized by variable results that are influenced by resource allocation and customer engagement. Savings also depend on proper planning; hence, significant emphasis is placed on Medicare reform [108]. These illustrate how complex it is to use data, with certain pros and cons in various conditions, limits, and potentials of data to maximize its utility in different spheres of life.

2.3.5 Future Trends in Data Acquisition and Analysis

Several aspects of new technology in connection with enhanced methods facilitate data acquisition and analysis in such diverse areas as healthcare, physics, or rock engineering. Such mobile healthcare apps as HealthTracker allow patients to participate in managing their health through tracking, storing, and analysing data related to their well-being [109]. These apps promote engagement in patients' health and enhance health outcomes using live feedback. Sophisticated data acquisition systems ensure that physics experiments make accurate and fast analyses of large quantities of data to allow for complicated research and data extraction [110]. Modern numerical modelling improves hydro-mechanical understanding of rock masses for tunnel and dam engineering in rock engineering [111]. Technological advancements underpin these trends through improved data collection and analysis of fields where real-time analytics, extracted signals, and numeric modelling flexibility are emphasized.

2.4 Social Network Analysis in Opportunistic Crowded Environments

SNA is a highly applicable analytical technique that could be used to understand interactions among people in different disciplines. SNA has brought out the Asian Leadership Gap, which shows East Asians tend to have few circles of friendship, limiting their access to leadership [111]. The social network analysis (SNA) categorized user groups and their interactions as they influenced people's perspectives about vaccines in social media debates [112-113]. SNA, when applied in virtual learning contexts, enhances the understanding of social development and growth among educational groups across different online channels. SNA emphasizes strong links within hacker communities for hacking ecosystems [114]. SNA supports recommendation system development in community-discovering work environments like CERN. SNA offers a new perspective on various societal aspects of leadership, public health, and digital learning, to mention a few.

2.4.1 Fundamentals and Role of SNA in Opportunistic Settings

In opportunistic situations, SNA aims to understand complicated structures and interlinks in dynamic social networks [115-116]. It reveals network structure characteristics and communication patterns in informal/opportunistic environments. The first function identifies core actors and influencers that understand network dynamics [117-118]. SNA shows the information passageways pointing at the health status of the given network [119].

Businesses use opportunistic SNA in consumer research, mapping disease transmission pathways and social determinants of health [120]. It provides information on community dynamics and resource sharing, thus informing urban planning. Data-related challenges such as volume and accuracy can be tackled using modern machine-learning techniques [121]. Opportunistic SNA is also a powerful analysis technique to understand complex sociological processes applicable to businesses, marketing, health issues, and urban planning.

2.4.2 Integrating SNA with Opportunistic Crowd Sensing

Social Network Analysis (SNA) integrated with Opportunistic Crowdsensing (OCS) presents novel solutions for information acquisition and processing [122]. The SONATA protocol enhances crowdsensing platform credibility through dynamic vote-based trustworthiness analysis within a social network setting, ensuring data accuracy and authenticity [123]. Crowdsensing improves traffic management systems by enabling real-time content distribution, enhancing response times, and optimizing urban mobility [124]. Diverse studies leverage social and user differences to enhance sensing quality and user engagement, recognizing the importance of user backgrounds and socio-cultural dimensions [125]. Incorporating SNA into OCS extends user profiles through social media platforms like Facebook, allowing for more targeted tasks and increased participation [126]. Social media platforms like Twitter facilitate public crowdsourcing of environmental events like heat waves, assessing society's response to environmental occurrences [127]. These applications demonstrate the potential of SNA and OCS in improving information reliability and urban planning, among others, highlighting the growing research interest in leveraging social networking dynamics for enhanced opportunistic crowdsensing capabilities.

2.4.3 Case Studies and Applications

Applying social network analysis in opportunistic networks provides solutions for various domains. Notable case studies include Professional Fraud Detection in Automobile Insurance: Through identifying loops in the networks' structure, investigating potential collusion between parties' interactions, and assisting researchers with fraud cases, SNA discovers fraud ringing in vehicle insurance [128]. Wireless Contacts, Facebook Friendships, and Interests: This paper investigates multi-dimensional, complex human interaction by tracing the social connections of all campus-wide Wi-Fi, personal profiles, friendships on Facebook, and interests. The result is a multi-layered social network, looking at how the interaction between social layers impacts human relationships [129] Information Flow in Disconnected Delay-Tolerant Networks: This paper uses a naive Bayesian model to counter proximity malware to stop infections via casual contact in dispersed areas. Malicious nodes causing untrue information and poisoning are also issues the model addresses [130]. Modelling Social Gauss-Markov Mobility for Opportunistic

Networks: Scientists present the HNGM mobility scheme, which combines a few societal attributes within a Gaussian-Markov model for the network of opportunity. HNGM results better than the conventional RWP methodology used in other network simulations because it faithfully reproduces real-world movement traits [131]. SNA has been deployed into opportunistic networks with various objectives, including fighting insurance fraud, spreading malware, understanding complex composition analysis, and interacting with physical and digital social dynamics to different challenges posed by such types of networks [132].

2.4.4 Ethical Considerations and Data Privacy in SNA

Ethical considerations and privacy issues are paramount when employing Social Network Analysis (SNA) in learning analytics. Informed Consent and Privacy: Therefore, informed consent is vital for the confidentiality of student information. It is also important that students are properly told why, how wide, and for what purpose the data will be analyzed so that they participate voluntarily [133]. Anonymization and Data Protection: Finally, sensitive information has to be protected by removing all identifiers from collected data via SNA. It is important to set strong security barriers to prevent unauthorized access. Addressing Potential Biases: Knowledge of embedded biases within analytical instruments such as SNA should be important [134]. Therefore, to avoid faulty notions about their interactions with some students, it is essential to critically evaluate these biases arising from algorithms or using research processes. Holistic Decision Making: Other data and information sources should be used cautiously alongside SNA insights if you consider making round and rational decisions within a complex educational setting. Ethical Use and Responsible Interpretation: Ethical use goes beyond data acquisition, analysis, and interpretation. It also implies implementing positive actions that contribute to ensuring educational rights and the right to be free among students. SNA may be informative about educational settings, though SNA should strictly comply with ethical standards to avoid compromising student confidence for better instruction and academic advancement [135].

2.4.5 Opportunities and Challenges in SNA for Opportunistic Environments

Integrating modern multidimensional networks presents opportunities and challenges: Low Earth Orbit (LEO) Mega Satellite Constellations: The case of LEO mega satellite constellations makes it possible to achieve almost complete global coverage, particularly in underserved regions. They occupy niche roles in ground networks but are very fragile in case of a lack of communication synchronization between terrestrial and satellite systems [136]. Terahertz (THz) Bands for UAVs in 6G Ecosystem: Unmanned aerial vehicles will use terahertz bands for ultrahigh throughput communication, a breakthrough innovation to data transmission speeds in 6G environments. On the other hand, they imply signal range and power trade-offs in their design. Based on 6G New Radio: 6G NR-U wireless infrastructure, UAVs can increase coverage capabilities for different terrain types. Implementing regulatory and standardization requirements for smooth integration with existing cellular networks [137-138]. Blockchain Technology in Healthcare: While blockchain technology offers a solution for better health information management as well as medical care, it also has its challenges, such as issues with privacy, trustworthiness of data, and security threats. Tackling these challenges has been found pivotal for effective utilization of in healthcare [139]. However, developing a comprehensive system for addressing the multi-faceted issues related to this technology will require the participation of techno-political leaders and legislature to achieve success.

2.5 The Role of Ubiquitous Computing in Opportunistic Crowd Sensing

Ubiquitous Opportunistic crowd sensing also utilizes ubiquitous computing, which relies on mobile devices and edge computing resources for data capture and processing. Several innovative approaches and frameworks highlight the potential of this paradigm: CrowdSense@Place (CSP) Framework: Users' visits can be tracked for each smartphone, and places can be accurately separated using this CSP framework. Through smartphone sensing, CSP provides environmental data, making it feasible for environment monitoring and classification [140]. Blockchain-Based Incentive Mechanism for Mobile Crowd Sensing: This approach brings into place a blockchain-based incentive scheme for mobile crowd sensing and edge-computer-aided networks. On data storage and credibility, it ensures that the data are true and secure, as well as establishes a decentralized reward system for contributors [141]. Honeybee-T Collaborative Mobile Crowd Computing

Framework: The work-stealing approach of Honeybee-T helps design efficient message exchanges between devices, considering both their dependencies and computational power requirements. This approach aids with the efficient and effective administration of large volumes of data flow from numerous devices in the context of mobile crowd sensing. As such, these studies and frameworks highlight a possibility that ubiquitous computing may find application in opportunistic crowd sensing for environmental monitoring and urban planning. The study shows emerging solutions concerning data integrity issues, storage requirements, processing efficiency, and participant payment schemes.

2.5.1 Emergence of Ubiquitous Technologies in Opportunistic Settings

Ubiquitous technologies are making significant inroads into various unpredictable environments, such as DTN and social media, with implications for diverse domains. Ubiquitous Technologies in Peer-to-Peer Content Sharing: Wireless technologies like Wi-Fi Direct and NFC revolutionize content sharing in infrastructure-less settings [142]. This is particularly valuable in academia, enabling students to share diverse content among devices seamlessly. Adapting these technologies for educational purposes fosters individualized learning and research approaches in university settings [143]. Remote Work in the COVID-19 Pandemic: The COVID-19 pandemic has underscored the need for novel metrics to assess and enhance remote work effectiveness.

Social media and pervasive technologies enable communication, collaboration, and productivity in remote work environments. Overcoming the unexpected challenges posed by the pandemic requires effective digital platforms for employee interaction and business continuity [144]. Challenges in Real-World Health Monitoring: Ubiquitous technologies are shaping the landscape of health monitoring, especially in complex real-world scenarios. Customized health monitoring systems address the intricacies of daily life and diverse health contexts, emphasizing intelligent designs that cater to various settings and user needs in a comprehensive yet user-friendly manner [145]. Modelling Domestic Activities for Context-Sensitive Technologies: Research efforts are focused on understanding domestic activities to develop context-aware technologies aligned with users' needs. This research ensures that technologies seamlessly integrate into users' daily routines and home environments, enhancing their usability and effectiveness [146]. The versatile applications of ubiquitous technologies in opportunistic contexts highlight their growing importance in facilitating collaboration in education, supporting remote work, improving healthcare monitoring, and creating context-aware domestic equipment.

2.5.2 Impact of IoT and Mobile Devices on Opportunistic Crowd Sensing

Opportunistic crowd sensing using mobiles has been expanded to mass and real-time data collection by integrating with the Internet of Things (IoT). This integrated environment leverages the ubiquity and connectivity of mobile devices to create efficient applications. Connected Cars and IoT: The collaboration between IoT and mobile devices is evidenced by connected cars. The cars receive updates via neighbouring sensors and add feedback to distant networks thanks to IoT connection. The sympathetic relationship between these bodies improves vehicle performance and transport reliability in general [147].

Efficient Task Allocation and Energy Consumption: Context-aware task allocations and artificial intelligence-based frameworks are used to enhance the capabilities of mobile crowdsourcing applications for better results. Context-sensitive processors make it possible to save energy by intelligently selecting work for non-functioning or distant devices. This refers to sophisticated systems in AI-enabled architecture geared towards resolving data security issues and optimized tasks [148]. Optimization of Multitask Assignment with Reinforcement Learning: Reinforcement learning is leveraged for the multi-action of assigning IoT devices in mobile crowdsourcing. They distribute various roles into IoT components to increase the performance and efficiency of data collection in networks [149]. The developments in IoT and the inclusion of mobile devices in opportunistic crowdsourcing provide possibilities for better data gathering, delegation of chores, and more information security. The technological advancement of this industry will persist as more devices go beyond simple stationary communication systems into everyday activities.

2.5.3 Challenges in Ubiquitous Data Collection and Analysis

Ubiquitous data collection and analysis, essential in modern technology, confront diverse challenges spanning different domains and applications. Heterogeneity of Data Sources: This involves the management, storage, and processing of data generated from different sources, such as smartphones IoT/IoS, among others [150]. However, integrating different data types and formats into business intelligence (BI) applications is challenging. Accuracy and Reliability: Recording correct or reliable health data and government or public health statistics is vital for treating patients. Automatically detecting errors is critical because when it comes to data quality, errors result in confusing conclusions with consequent poor decisions [151]. Privacy and Security Concerns: The emergence of modern smart homes with advanced tech tools raises issues related to users' security and personal information protection. Security of these devices and responsibilities relating to data collection and usage, such as self-management for chronic disease, are extremely important. Intergovernmental Relations: In this context, effective coordination among the different government levels must exist in federal systems. Information sharing and policy formulation of large-scale public projects is only possible through a consistent framework for intergovernmental relations.[152] Data Collection in Challenging Environments: Achieving connectivity to collect data, especially in remote or constrained areas, poses privacy concerns. Thus, this situation dictates that novel data collection procedures appropriate under varying environmental constraints must be found.

Monitoring Large-Scale Events: It is difficult to monitor and analyse the crowd's behaviour in events such as pilgrimage during Hajj. Data collection and analysis are important in activity recognition, group behaviour analysis, detection of stress among participants, and health monitoring of participants. To deal with those mentioned challenges, strategies must be aimed at improving the ways to collect data, ensuring its quality, addressing confidentiality and security issues, and developing more advanced analytical instruments [153]. A multi-faceted approach is required to fully exploit ubiquitous data collection and analysis with minimal risk and limitations. This will involve reviewing data collection methods, ensuring the data remains valid, and addressing safety and confidentiality issues.

2.5.4 Future Directions and Innovations

Cutting-edge technologies like RFID are poised to reshape the future of ubiquitous computing, altering our interactions with technology and daily life. Federated Learning and Edge Computing: FL and Edge computing integration in the UI and 6G network domain shows great potential. Besides this, FL allows modelling training in collaboration and sharing data privately between devices, thereby improving UX design. In tandem with an advanced strategy dubbed “edge computing” that positions a few computing elements close to the network edge, it creates an effective yet distributed computational structure that substantially increases the efficacy of FL [154]. Technology and social media in Dementia Care: Dealing with Dementia: Dementia Care involves smartphones, tablets, or the Internet as health education means improving living standards and care [155]. Energy Optimization in Smart Cities: Sustainable solutions are necessary for green IoE-based applications in smart cities. Energy utilization efficiency influences these cities' transport, administration, and quality of life [156]. Context-Aware Security for Smart Homes: Context-Aware Security Mechanisms for innovative smart home security focus on cost-effectiveness and higher threat prevention and detection capabilities than conventional security systems. Apps for Linguistic Data [157]. Collection: Increasingly, smartphone apps such as Dialäkt Äpp and Voice Äpp are being employed for public involvement and survey work on language variation and evolution [158]. 5G Networks and Beyond: Due to the massive implementation of Smart Devices, there is a large market for 5G networks with greater capacity for bigger networks [159]. Such a high-capacity network enriched with improved operation characteristics would facilitate wireless communication services for numerous users. This illustrates the impact that ubiquitous devices have had on various industries. Technology evolves daily as an agent of healthcare, communication, energy management, and security. Its usefulness and applications will be incorporated into our daily routine.

2.5.5 Case Studies: Ubiquitous Computing in Dynamic Environments

Ubiquitous computing in dynamic environments is a rapidly evolving field with diverse applications and challenges. A3C-Based Real-Time Scheduler: An AA3C-based advanced scheduler for the Edge-Cloud under uncertainties. It is, therefore, an effective way of processing computationally intensive dynamic systems that require timely decision-making. Framework for Human Behaviour Monitoring: Another research concentrates on tracking patients’ acts during ADLs. It promotes observing, supervisory, and evaluative practice in everyday activities. Such knowledge can improve patients’ care and provide elderly persons with chronic diseases [160].

Chaotic Maps-Based User Authentication Scheme: A new user authentication and privacy-preserved extended chaotic maps mechanism for distributed, pervasive, and heterogeneous environments. The scheme offers DoS resistance and ensures data safety for users [161]. Denotational Mathematics for Home Hospitalization Support: Using Denotational Mathematics, researchers have created a program that generates medically appropriate contextual content for home hospitals. In particular, telemedicine must embrace this methodology to provide pertinent assistance with personal instructions for each individual's household and health care requirements [162]. Context Information Interoperability: Literature-based solutions of the information interoperability in ubiquitous computing. It is critical in enhancing interoperable ubiquitous computing in disparate environments and contexts [163]. Ubiquitous computing case studies illustrate different uses in health care, patient tracking, cyber security, homes, hospitals, etc. This emphasizes the current trends regarding solutions to the issues using contemporary technology-based interventions.

2.6 Existing Models and Algorithms for Opportunistic Crowd Density Estimation

Opportunistic crowd density estimation has been evolving through different models and algorithms developed to address the unique challenges of estimating accurate crowd densities. Lightweight Dense Crowd Density Estimation Network: This approach involves presenting a more efficient and lighter convolution block engineered for crowd characteristic extraction. It provides an approach to dealing with efficiency in crowd density estimation by lowering the values of various networking and computing parameters. The model also employs spatial group normalization designed to deal with biases from variable crowd distribution, which is common in crowded areas [163].

Crowd Density Estimation using Imperfect Labels: This paper presents a system for creating imperfect labels using a deep learning model to measure the effect of label mistakes on crowd counting precision. This innovation shows that even if the crowd models are built using imperfect labels, they have strong resilience against annotation errors and can model as well as those trained with perfect labels [164]. DroneNet for Crowd Density Estimation using Drones: This work presents a new system called DroneNet that builds on Self-organised operational neural networks (Self-ONN) to estimate crowd density using drone data. This model has efficient learning capabilities and less computational complexity than conventional CNN-based models, increasing performance in drone-based crowd-monitoring applications [165].

Parallel Multi-Size Receptive Fields and Stack Ensemble Meta-Learning: The approach utilizes a paralleled multi-size receptive fields unit model drawing on a noteworthy amount of the CNN layer features. The system can scale varying-sized elements of an object for any chosen plane within the framework. This approach is especially suitable for dynamic crowd settings that normally involve spatial and temporal deformations [166]. Attention-Based Capsule Network and Multi-Column CNN: This deep neural network contains two columns containing CNN and capsule network-based attention modules. This approach focuses on estimating crowd numbers by taking static or video images and appears promising in several benchmark datasets; hence, it could be applicable in crowd density estimation [167]. Such models and algorithms are important in various applications such as autonomous driving, visual surveillance, crowd control, public space planning, and increasing traffic warnings caused by crowd density. The current research and development continue to push the limits of accurately and efficiently estimating crowd densities using advanced computational techniques.

2.6.1 Comparative Analysis of Algorithms

The table in this section provides a complete synopsis of some algorithms and data for crowd density estimation, an indispensable element in comprehending crowd behaviour and control. It includes details on model names, data-specific features, types, associated constraints, and referenced resources. The table 2-1 contents shed light on key findings within this domain:

- Model Names and Datasets: In most cases, the models are not accompanied by information on the data used for training or if another independent dataset is involved.
- Type of Datasets: The most notable fact is that high use of visual sets, mainly pictures and videos, can be observed. These datasets mostly come from surveillance cameras or drones, demonstrating that visual data is important in estimating crowd density.
- Limitations: Interestingly, the table demonstrates the endogenous shortcomings inherent in the models and datasets. These include a lack of effectiveness in dense crowds, the impact of annotation errors on the model accuracy, and the complex balance between model depth and inference time.

Table 2-1: Systematic Comparison of Crowd Models

Model Name	Dataset Name	Type of Dataset	Limitations	References
Parallel Multi-Size Receptive Fields and Stack Ensemble Meta-Learning	Not specified	Not specified	Including most layer features in the prediction model negatively affects the prediction's outcome.	[163]
DroneNet – Self ONN	Benchmark crowd datasets containing images taken with drones	Images taken with drones	Deeper CNN models improve accuracy at the cost of increased inference time	[164]
Imperfect Labels	Not specified	Not specified	Impact of annotation errors on the model accuracy	[165]
Lightweight Dense Crowd Density Estimation Network	Not specified	Not specified	Efficient compression models	[166]
CSRNet	ShanghaiTech dataset	Images taken from surveillance cameras	Limited performance in crowded scenes	[168]
MCNN	ShanghaiTech dataset	Images taken from surveillance cameras	Limited performance in crowded scenes	[169]
SANet	ShanghaiTech dataset	Images taken from surveillance cameras	Limited performance in crowded scenes	[170]
Generalized Loss Function	Not specified	Not specified	The proposed loss function may not be optimal for all datasets	[171]
Learning to Count via Unbalanced Optimal Transport	Not specified	Not specified	The proposed method may not be optimal for all datasets	[172]
HARNet	Not specified	Not specified	Uneven crowd distribution may lead to inaccurate counting	[173]
Annotator Model	Not specified	Not specified	Impact of annotation errors on the model accuracy	[174]

Lookup-Table Recurrent Language Models	Not specified	Not specified	Scaling up the size of RNN language models with only a constant increase in floating-point operations	[175]
CascadeTabNet	Not specified	Not specified	An end-to-end deep learning model that exploits the interdependence between the twin tasks of table detection and table structure recognition to segment out the table and column regions	[176]
LCDet	Not specified	Not specified	Real-time crowd counting in surveillance videos	[177]
CrowdNet	Not specified	Not specified	Real-time crowd counting in surveillance videos	[178]

2.6.2 Identifying Gaps and Potential Areas for Future Research

1. Dataset Diversity and Representation: Many models such as “Parallel Multi-size Receptive fields and stack ensemble meta-learning,” “lightweight dense crowd density estimation network,” LCDnet,” “UEPNet,” “generalized loss function,” “learning to count via unbalanced optimal transport,” “HARNet The absence of data on this matter suggests there may be a lacuna in diverse and representational datasets employed herein. Such models must be tested with many data sets to establish their generalization properties (robustness) in different scenarios and environments

2. Exploring the Impact of Imperfect Labels: Thus, the “Imperfect Labels” model was assessed on the DroneRGBT dataset comprising RGB and IR imaging pairs. The significance of this approach involves dealing with the influence of inaccurate annotations on the model’s reliability. However, there appears to be an opportunity to research which kind of imperfect data (label imprecision, noise, or missing values) would impact which specific crowd density estimation model. The field of research warrants further studies as resilient and exact models for imperfect data are essential even in most real-world situations.

3. Real-Time Analysis and Computational Efficiency: However, some models, such as "DroneNet" and one with "Self-ONN," are necessary due to the need for lower computational complexity in real-time applications. However, achieving this balance between accuracy and computational speed is still unresolved. The "DroneNet" limitation mentions that deeper CNN models could improve accuracy at the expense of longer inference cycles. This issue is critical for real-time analysis, such as public safety monitoring and event management.

4. Adaptability to Varied Environments: Although models such as MCNN, FCRN, SANet, and CSRNet performed moderately well concerning ShanghaiTech data set in most congested situations, they only captured sporadic events. The study shows an existing research gap concerning model development for crowding at different densities, especially in places where there are high volumes of pedestrian traffic. However, models must be flexible enough to accommodate varying crowd sizes and density levels.

5. Ethical and Privacy Considerations: Crowd density estimation models pose ethical and privacy issues, particularly if implemented in public areas. However, the literature does not focus much on how these models handle privacy, consent, and other ethics concerning the use of surveillance data. However, research combining privacy-protecting methods and ethical considerations in creating and using such models should be undertaken.

6. Long-Term and Large-Scale Deployments: Such evaluations are mostly limited to controlled or short-term stages. However, the performance of these models is not well understood in the long-term large-scale deployment of such devices. Such as knowing how they should be maintained, if they can easily change environments, and if they are durable for a long period.

7. Comprehensive Evaluation Metrics: However, the existing assessment metrics, such as Mean Absolute Error (MAE) or Grid Average Mean Error (GAME), are limited, as they do not take into account other aspects important for comprehensive modelling, including adaptability, computational speed in real-time settings, and moral factors.

The literature review reveals several gaps in crowd density estimation: diversity of datasets, effects of imperfect labels, real-time analysis, computational efficiency, adjustment for different environments, legal and privacy issues, long-run performance, and broad performance measurements. Closing these gaps would yield more comprehensive, effective, and ethical models for crowd density estimates.

2.7 Conclusion

The literature review chapter encompasses various aspects around crowd density estimation ranging from conventional means to modern techniques using technology. It scrutinized the subtleties surrounding opportunistic environments in crowd sensing, the centrality of both artificial and genuine dataset, and the incorporation of social network analysis within these contexts. Ubiquitous computing was also highlighted in the chapter; IoT and mobile phones were identified as being vital components of opportunistic crowd sensing. It also offered a critical review of different models and algorithms, outlining their advantages and disadvantages under varying circumstances. This review presents different dimensions of the topic and establishes a platform for addressing subsequent issues that will be examined in the succeeding chapters and their recommendations. These insights form a basis towards improving crowd density estimation, as well as in predicting movements using mobile network data.

CHAPTER 3: THEORETICAL FRAMEWORK

3.1 Conceptual Framework of the Study

3.1.1 Definition and Scope of Opportunistic Environments

Opportunistic settings exhibit rapid changes in which random crowds emerge suddenly without plans as they converge. Crowd density estimation varies from place to place since such environments may happen on busy city streets or at special locations (such as sports facilities or other special events) or even occur accidentally when a group of people unexpectedly gathers. It is important to recognize that using the term opportunistic in these contexts denotes the randomness and unpredictability of mass formation, which is quite hard to contemplate beforehand. Examples of opportunistic environments include urban areas with many people walking around. People flock to these busy urban centres, tourism places, marketplaces, and transport junctions [179]. They also exhibit varied patterns of crowd dynamics. For example, street performances, public protests, or social media viral might prompt a spontaneous gathering of crowds that vary in numbers and attributes. Such environments do not suit traditional crowd management techniques since they are typically unpredictable and spontaneous.

Similarly, special event venues like stadiums, concert halls, and convention centres may turn into opportunistic places during unplanned meetings. For instance, the appearance of a celebrity in a surprising manner at a music show or a big sports victory can cause people to gather not only within but also outside the place. Such crowd control is conventional and will not be suitable for dynamic situations in different situations [180].

Adaptive crowd density estimation is the key challenge faced by opportunistic environments. These environments can hardly be monitored using conventional methods like manual counting or fixed sensors because they change rapidly. Instead, innovative approaches are required. These include real-time video analysis, mobile crowd-sourcing apps, and machine learning algorithms. Such technologies can also quickly adapt in case of changes within people’s movement and formation, supplying reliable and current data concerning the size of groups and people’s activities and behaviour [181]. Opportunism requires a strategy of people’s density adjustment towards management success. Such a monitoring system enables officials and event planners to react swiftly to crowd-sourced problems and provide safety measures using the best management techniques. Using technology and data analytics, one can anticipate and manage unexpected assemblies of event attendants and other citizens, promoting safety during the event and improving public satisfaction. The opportunistic contexts refer to situations characterized by a high degree of uncertainty and sudden gatherings of crowds. Crowd density estimation in such an environment is demanding as it must be handled based on the nature of the situation with a highly adaptable crowd control strategy. To overcome such challenges in an opportunistic environment and ensure safety and satisfaction, adopting advanced technologies or data-driven solutions [182].

3.1.2 Role of Opportunistic Sensing in Crowd Density Estimation

Opportunistic sensing constitutes an essential aspect of crowd density estimation that focuses on places where normal detection or lack of feasibility prevails. Using various data sources, including personal device statistics, social media activity, and digital trails, this approach provides quick updates on crowds' location, direction, and extent. It is an agile solution to the unpredictability of people in crowds. The WiCount system is an example of opportunistic sensing [183]. It uses Wi-Fi traffic data, easily available for estimating crowd density. With each passing second, it gathers communication packets sent out by nearby Wi-Fi access points and uses up-to-date machine learning classifiers with high precision to determine how many individuals have run into specific zones. The performance of WiCount is admirable since it can count 99 percent of persons within a crowd or people entering at different times. Such a technique is of immense significance during mass surveillance in public gatherings and places such as shopping malls and stadiums [184].

The other innovative procedure uses deep learning sense of the radio wave in cellular communication. Deep learning models can estimate the crowd density and their motion pattern by analysing radio wave signals emitted by mobile phones and other wireless equipment, thus avoiding camera monitoring. It solves camera-based privacy issues and is useful when visual surveillance has been deemed unworkable. The opportunistic sensing can include CSI measurements obtained by decoding the cellular network's unencrypted sync. Individuals in the surrounding area reflect these signals transmitted by cellular eNodeB [185]. Crowd density estimation can be done quite accurately if you learn how the differences between CSIs are influenced by people passing by or walking around. This research shows great potential; it averages an accuracy of about 84% points. The method demonstrates an ability for CSI-based population sensing to estimate crowds.

There are numerous advantages associated with opportunistic sensing. Firstly, it enhances data precision using multiple information sources and sophisticated data processing mechanisms. In most cases, they exceed traditional techniques to estimate finer figures of people in a crowd. Additionally, opportunistic sensing tackles privacy issues, an important point for people living in modern times and under constant threats of being monitored [186]. Less invasive techniques, such as radio wave sensing and Wi-Fi traffic analysis, could be applicable in an environment where privacy rights come first. Opportunistic sensing is also the last step that ensures timely response to incidents related to crowds, efficient events coordination, and improved public security. Opportunistic sensing is an advanced and adaptable way of measuring crowd density. This utilizes different information sources and modern technologies to overcome the limitations of traditional monitoring systems. Opportunistic sensing helps manage crowd dynamics by offering accuracy, addressing privacy issues, and providing real-time insights.

3.1.3 Various Data Sources in Opportunistic Settings

Data heterogeneity is often challenging when integrating different data sources in opportunistic settings. However, this analysis must be combined with findings from other data sources to enhance its precision and validity. This section combines different data sources like MOBILE data, Wi-Fi signals, and social media activities in the context of crowd density estimation. This research discusses the difficulties and risks of merging the different data streams and the approaches and algorithms required to integrate these data into a single, comprehensible set [187].

The process uses Fisher information to evaluate the informativeness of each data source's sources. This more sophisticated data fusion framework results in straightforward parameter estimates, sound tests, and confidence intervals in the setting of generalized linear models. In addition, dial selection as per the guidelines results in better estimations than other binary integration methods. An alternative way is to adopt an open-access system for merging different data sources to perform a safety-related analysis of drugs. This platform combines information from locations like ICSCRs, RWD, social network information, and literature. Each workspace contains data-intensive analyses developed, and the end user sets up "investigation scenarios" by drug-event combinations [188].

The data about sequences of viruses must be easily accessible and available for integration, which will most probably include data on the host's genetic structure and patient history to understand the disease mechanisms and its prevention strategies. While few host-pathogen integrated datasets exist until now, a number of them should be developed when additional knowledge on the disease emerges [189]. Recognizing where the common variants are temporally and spatially distributed and linking them to the phenotype reported in the literature could develop useful integrative surveillance mechanisms. Some strategies and algorithms can allow the integration of multiple data sources in an opportunistic setting. This way, such different data streams are combined to achieve more reliable analysis output. Considering the informativeness of the data sources and employing proper integration techniques will allow researchers to combine and analyse multi-source data to comprehend complicated processes, for instance, better crowd density approximation.

3.1.4 Framework for Analysing and Interpreting Data

A solid framework is instrumental in analysing and interpreting opportunistic data, especially in crowd density prediction and diverted crowds. Such a framework employs various data sources for data acquisition, including video feeds, cellular telephone signals, and social media. Secondly, this data is analyzed using methods borrowed from machine learning, computer vision, and signal processing. This aims to measure the crowd's density and predict crowd behaviour to control the same where appropriate [190].

Constructing a robust framework is important when opportunistic situations make data collection tough because of the unplanned nature and spontaneity at which crowds are formed. This framework comprises five steps: data collection, data preprocessing, feature engineering, modelling and analysis, and determining the significance of results in formulating informed policies on crowd management. The data collection involves getting information from several sources like CCTV, mobile calls, or data harvested from social media sites. This usually involves setting up sensors, cameras, data streams, etc., to collect appropriate information [191].

Having collected relevant data, pre-processing is key in ensuring that all the collated data is cleaned up and prepared to be considered before any critical analysis. In this stage, procedures include data fusion, noise reduction, and feature extraction to confirm that the info is appropriate for subsequent processing. Feature engineering is imperative in detecting and determining appropriate features among the data. These attributes form a basis for assessing crowd density, predicting crowd movement, and facilitating crowd diversion where necessary. Analysis of modelling follows in this case where techniques used in signal processing, machine learning, and computer vision on the data.

Consequently, it is possible to construct models for crowd density estimation, crowd steering, and crowd mobility prediction. The last one is the assessment and analysis of the models and results. These models are stringently evaluated, and their outcomes are used interpretatively for making relevant decisions on issues relating to crowd management [192].

3.2 Theories of Crowd Dynamics in Opportunistic Environments

3.2.1 Fundamental Theories Underpinning Crowd Behaviour

Gustave Le Bon’s “theory of the crowd mind” is one classical theory concerning individuals in a crowd's behaviour. This theory suggests that individuals cease being unique as they are assimilated into the collective mind. The first perspective believes that crowd behaviour is mostly determined by the common consciousness and not by the individual identity of the persons in the crowd. However, many modern theories nowadays explain crowds' characteristics under different conditions. A modern theory is “collective resilience in crowds” by John Drury. His theory is mainly concerned with the cooperative behaviour of crowds when faced with emergencies and the ability of such situations to stimulate self-organizing among crowds [193]. It notes the importance of social contact and communication between crowd people, which may result in mutual resistance and partnership during misfortune. Such theories should apply to opportunism’s highly variable and uncertain situations. It is important to understand the psychological and cognitive dimensions of human behaviour for predictive and managerial purposes during crowd behaviour forecasting for these situations. Several approaches have emerged to detect and predict crowd behaviour, including:

- Cognitive deep learning frameworks integrated with Psychological Fuzzy Computational Models based upon theories such as Occupational Arousal Theory, Fuzzy set, Five Factor Model, and Visual Attention. Using these approaches, one can easily spot different crowd behaviours [194].
- Psychological perspectives on risk management in public-private partnerships (PPP), including psychometric advice and better ways to share risks.
- Development of unmanned aerial systems based on drone technologies, artificial intelligence, and machine learning to help monitor and analyse crowd behaviours before, during, and after peaceful and non-peaceful events.
- A crowd behaviour simulation model proposed under the information-gap theory considers curiosity an intrinsic motivator influencing individual and crowd behaviour [195].

Therefore, the multiplicity of these methods reveals the necessity to estimate people’s actions under different circumstances, such as opportunistic ones. By forecasting behaviours and management, considering psychological and cognitive aspects of crowd behaviour may help improve safety in crowd-related issues.

3.2.2 Modelling Crowd Dynamics in Unstructured Environments

Crowd dynamics are modelled in unstructured, opportunistic environments through different methodologies and models representing crowd behaviour. There are two main strategies, including those focusing on the behaviour of agents in a group as a set of microscopic objects (agent-based models) and treating multitudes as an averaged mass motion (fluid-dynamic models) [196]. However, these models will have to be flexible to consider random gatherings of crowds, varying size composition of crowds, and quick changes in crowd density. Models for describing crowd dynamics have been recently developed. Using hyperbolic elliptical equations, one can give a formal description of the movements of a crowd moving [196]. It is a microscopic description of an individual's behaviours and ascertains that the evacuation time from a bounded domain is finite and the model is well-posed. A second study suggests a two-dimensional kinetic model for disease contagion in crowds, considering the spatial spreading of diseases in confined areas with an average-sized population [197].

For automated driving in unstructured, dynamic environments, a learning-based model predictive trajectory planning controller solves the absence of prior knowledge in unstructured environments and introduces a risk map that maps density and motions' obstacles and the road with risk occupancy [198]. The proposed controller bridges the gap between planning and control by learning the residual model uncertainty through Gaussian Process regression to improve the model's accuracy. The modelling of crowds in unorganized and random settings is a challenging area gaining significance as researchers attempt to create flexible models reflecting the fluctuating nature of crowd behaviour within those contexts [199].

3.2.3 Impact of Opportunistic Data on Crowd Dynamics

Using opportunistic data such as mobile phone signals, social media activity, and sensor data has revolutionized how people understand crowd dynamics, especially in unplanned areas. Such information is invaluable for our appreciation of the collective dynamics of crowd behaviours, which are often considered unknown. In a nutshell, let us now examine the meaning of crowd data concerning crowd dynamics, the problematic aspects, and research perspectives on this subject [200].

Real-Time Data Collection: Real-time data collection becomes possible using opportunistic data sources, including mobile phone signals and social media activity. Unlike other data collection approaches, such as hand counts or static sensor-based approaches, it involves real-time monitoring of crowd movements and behaviours. In addition, regarding rapid changes that may occur in a crowd, this real-time element will be very useful as it allows for monitoring such situations, thereby enhancing public safety and effective crowd control management. Granularity of Data: Data is generated at an atomic level in personal devices such as smartphones [201]. Therefore, we can collect specific details of single activities or locations in the crowd. It enables better insight into crowd dynamics like density, movement patterns, preferences, and individual choices [202].

Identification of New Patterns and Behaviours: These opportunistic data can disclose hidden patterns of behaviour and trends that one cannot see through other data sources. Analysing digital data trails from people's mobile phones and activities on social media allows researchers to understand why crowds act collectively as spontaneous gatherings and how viral trends and unforeseen happenings affect them [203]. Despite the evident advantages of leveraging opportunistic data for understanding crowd dynamics, several challenges need to be addressed:

- Data Sparsity: There are regions and M-o-Ments in a case of opportunistic data that can be very short on data points. However, this may lead to gaps in knowledge about crowd behaviour and reduce the precision of crowd density estimates and behaviour forecasts [204].
- Data Noise: Any error in opportunist data, including wrongness and insignificant input, maybe a source of noise that will frustrate analysis efforts. The quality and reliability of insights drawn from the data depend on noise reduction and data cleaning techniques [205].
- Privacy Concerns: Privacy issues arise on an individual level due to the collection of personal data from mobile devices and social media. Striking a balance is essential while using opportunity data as it infringes individual privacy rights regarding research data collection [206, 207].
- The use of opportunistic data in studying crowd dynamics has been highlighted in several research papers, which discussed both the chances and obstacles connected with this approach. Crowd participation in mobile computing (MC), multimedia big data visualization techniques for a scale-free network in the smartphone, a hybrid multi-componential model framework for data-driven crowd dynamics, integrated fusion crowd simulation methods for big data smart cities, etc. [207].

The opportunistic data has enabled us to perceive crowd behaviours in real-time with detailed descriptions. On the other hand, though, data sparse, noise, and privacy issues should be handled with great caution. The studies carried out in this area will create possibilities that will help with crowd control, security, and better knowledge concerning the group behaviour of humans in undifferentiated open spaces.

3.2.4 Comparative Analysis of Models

The extensive search results were unsuccessful in comparing the theories of crowd movements. Hence, based on current knowledge, this paper looks at some of the theoretical models used for modelling crowd behaviour and the extent to which they apply under opportunistic situations. Crowd dynamics theoretical models represent various options or approaches that vary in their specific features and purposes. This model is an essential tool providing the basis for predictive crowd behaviour used to monitor and control large groups. Here, we delve into three prominent categories of crowd dynamics models: social force, continuum, and agent-based models [208]. The social force model emphasizes the interaction of an individual with their surrounding within a crowd. The assumption is based on the idea that different kinds of social forces control the movement of people in a crowd. The need to get to a specific point, avoiding the collision of their movement with the crowds, and other factors determine the character's behaviour. Social force models are good at understanding how emergent crowds' behaviours and macroscopic happenings can be easily understood, but they may exaggerate the complex decision-making process of each crowd member [209].

Unlike the continuum model, where a crowd is treated as a continuous fluid, partial differential equations describe their movement. These models especially suit studies of crowding phenomenon in open spaces and reveal information about crowd density parameters and flow structures. Despite that, it may be problematic for the continuum models to explain the impact of obstacles, structures, or other environmental factors that can greatly influence the behaviour of the crowds. The most diverse of the three is the so-called agent-based model that simulates the activity of individuals in the crowd. This way, there is no limitation regarding behaviours, interactions, and decision-making processes that can be expressed. Unlike other approaches, agent-based models are good at describing how different crowds behave and have individual attributes that can be altered to fit situations such as evacuation or public gatherings. They may also be computationally intensive and use large amounts of data for calibration and validation. The comparative analysis presented in the table 3-1 discusses the strength and limitation of several models in context to the proposed research context of opportunistic environment:

Table 3-1: Crowd Density Models Comparison

Model Type	**Key Features**	**Purpose**	**Applicability in Opportunistic Situations**	**Limitations**
Social Force Models	Focus on individual interactions within a crowd	Understanding emergent crowd behaviours and macroscopic events	Suitable for open, clear areas with straightforward movement goals	May oversimplify complex individual decision-making processes
Continuum Models	Treat crowds as continuous fluid, use partial differential equations	Suited for studying crowd dynamics in open spaces	Effective in analysing crowd density and flow structures	Struggle to account for environmental factors like obstacles, which can influence crowd behaviour
Agent-Based Models	Simulate activities of individuals, no limits on behaviour expression	Describe diverse crowd behaviours, adaptable to specific scenarios	Ideal for complex situations with intricate social networks and decision-making, like evacuations	Computationally intensive and require substantial data for calibration and validation

Every model has its own merits and applies in particular circumstances. The social force mode can be applied when a crowd traverses an area free from crashes, collisions, and other encounters. There are more efficient continuum models for analysing the flow and density of crowds over large open spaces. Agent-based models with highly sophisticated individual-level simulations are more suitable for complex environments where individual behaviours, interactions, and decision-making play a significant role in crowd dynamics. While selecting the suitable model for a particular situation, especially in emergency evacuation, the choice must depend upon whether the model can represent these variables sufficiently and offer useful information for the specific case [210].

The choice of the appropriate model is dependent upon the nature of the situation in which it is applied in an opportunistic environment like an emergency evacuation or unplanned events. For example, social force models could simulate individuals' movements in open, clear areas where the leading motive is getting there and not hitting other people. On the contrary, agent-based models might suit such situations as complex social networks, difficult decision-making, and various group behaviours, to mention but a few. Finally, in this chapter on the model theory, crowd comparison should consider each model's capability in specific situations related to the features and needs of the crowd. The suitability of a given opportunistic environment requires that factors such as individual behaviours, group interactions, environmental variables, and the emergence of crowd properties be examined with care to identify the proper model of opportunity identification [211].

3.2.5 Application of Theories in Diverse Crowd Scenarios

Crowd dynamic theories have a huge potential in developing practical crowd management strategies in different opportunistic environments to improve crowd management strategies, emergency response plans, and urban design policies. These theories are useful for explaining and acting on impromptu gatherings in urban spaces, spontaneous crowds of unexpected events, or emergencies. Theoretical ideas in real-life situations are integrated towards practical solutions for improving safety, proper allocation of resources, and urban planning [212].

Crowd dynamic theories are also used to develop a dynamic disturbance-propagation model predicting crowd panic behaviours. This model is based on fluid dynamic conservation laws and aims to depict how crowds move when disturbed. Firstly, this model provides a theoretical basis for understanding crowd panic, and secondly, it enables the conduct of Lyapunov-type analysis on the stability of crowds under disturbances. The crowd panic model allows urban planners and emergency responders to understand how panic crowd behaviours spread and develop effective measures for controlling crisis occurrence in reality [213].

Moreover, it can be applied to optimizing opportunities in traffic flow management. For example, there is a model that reallocates traffic by charging different prices for roads. The idea behind this approach is that drivers intend to use a cheap way with the least distance traveled. Optimization of road-based variable cost factors redirects people to less busy routes, allowing traffic to be unlocked and mobility to be enhanced. The application of theories of crowd dynamics for traffic management shows how theory can be turned into specific plans for better city design and allocation of resources [214].

Such examples illustrate the usefulness of the theories of crowd dynamics, as they may serve to deal with exploiting environments. By applying findings drawn from theoretical models, practitioners and policymakers can make the right decisions, have appropriate crowd-controlling mechanisms in place, and improve security measures. Finally, when theory is transferred to practice, authorities can better react to the diverse ways crowds act unpredictably in different circumstances. This, in turn, leads to improved crowd management, safety, and urban planning outcomes [215].

3.3 Principles of Synthetic Data Generation and Validation in Opportunistic Contexts

3.3.1 Rationale for Using Synthetic Data in Opportunistic Environments

The analysis, which underpins synthetic data for modelling crowd dynamics and opportunism, is provided here. There are varied reasons why artificial information would be useful in this case, which highlights its immense benefits. Secondly, synthetic data provides a controlled testing ground for models and algorithms whereby there can be limited or no actual-world data available [216]. These are opportunistic conditions that pose difficulty in data collection, thus making use of synthetic data useful for initiating exploratory work. In addition, synthetic data is a more ethical and privacy-friendly substitute for personal real-world data. Using artificial data will ensure it is done ethically without infringing privacy rights, yet researchers can still conduct relevant studies. This section will also discuss wider aspects of artificial data, such as the fact that it enables researchers to get an in-depth comprehension of intricate crowd processes without the restrictions of collecting real-world data.” Through it, researchers can run experiments on simulations that are impossible or unethical in real life but help them understand crowds’ behaviours in opportunistic settings [217].

3.3.2 Methodologies for Creating Realistic Synthetic Data

Synthetic data is created by combining various statistical models and algorithms. This makes the generated data as close to reality as possible. Such approaches may include Monte Carlo simulations, agent-based modelling, and other probabilistic models. This synthetically generated data can consider crowd size, movement patterns, and environmental variables to make them as real and applicable as possible.[217]

Studies in many areas, like human analysis or medical informatics, discuss creating artificial data as a viable substitute for actual data collection. Synthetic data generation has already been applied to mitigate the problems associated with a lack of real patients' data for use in machine learning-supported decision-making in medicine[218].

Although many studies have addressed the issue of data-free model compression in the deep learning setting, it remains unresolved. Some free distillation procedures are susceptible to catastrophic forgetting and inconsistencies between synthetic and practical statistics. Recently, a framework has been proposed for data-free knowledge distillation that keeps an updated database of generated examples and imposes the constraint that each sampling strategy matches the real data distribution [219].

A study examined externally evaluating clustering documents via synthesized data for a different setting. The closeness measure was applied for each pair of measures. The value change of any measure while considering an increased or reduced number of generated clusters was used as an assessment tool for measuring the effectiveness of the respective measure [220].

3.3.3 Validating Synthetic Data Against Real-world Opportunistic Data

The next important step will involve validating artificial data for authenticity, suitability, and relevance to research and decision-making purposes. In this regard, various ways and standards exist with which we can rely on artificial data as a sample of real data. Another approach compares synthetic data with real data collected from opportunistic settings. Such a comparison enables scientists to examine whether or not there is an association between the two data sets. Using approaches like cross-validation, statistical testing, and error analysis also helps measure the correctness and appropriateness of synthetic data [221].

The study ranked different methods for synthesizing health data based on their ability to support a particular analytic workload using a " utility " metric.” This study concluded that multivariate Hellinger distance is appropriate for evaluating various utility metrics as it best compares the actual and synthetic joint distributions expressed by a multivariate Gaussian copula model [222]. A flexible validation approach proved indispensable in evaluating synthetic solar irradiation data. The appreciation arises from varied utilization of artificial solar irradiance data with different emphases and necessities. This, therefore, requires a flexible validation approach for the data generated through the synthetic solar irradiance to be valid and applicable within varying circumstances [223].

Another study used machine learning algorithms that were developed and validated in predicting HIV infection in men who have sex with men. The models were evaluated using the measures of accuracy, precision, recall, F-measure, and AUC. As such, these measures ensured the researcher could completely justify that the machine learning models would work in practice [224]. These instances highlight the variety of ways and measures in which synthetic data can be verified using actual opportunistic datasets. The credibility of synthetic data as an analogy to real-world data should be established across various domains to allow researchers and practitioners to successfully apply synthetic data in conditions when the real-world counterparts are difficult or impossible to gather.

3.3.4 Challenges and Solutions in Synthetic Data Generation

Challenges and opportunities in the murky landscape of synthetic data creation. Creating artificial data that accurately depicts the complexities of opportunistic circumstances in real life is no easy feat, requiring extensive thought processes and ingenuity. Ensure the synthetic data is representative. The synthetic dataset should reflect authentic environment trends and represent the data collected in spontaneous environments [225]. This can be accomplished through an integrated process of consideration of the multi-scenario, multi-context nature of these settings. The other intricate problem involves attempting to describe the random behaviour of crowds; one of them is crowd dynamics. Human behaviour is highly dynamic and is affected by many variables, which may cause it to change suddenly. Consequently, imitation of human actions that display variability in opportunity scenarios should encompass synthetic data [226].

.

Synthetic data generation should include all important factors affecting dynamic crowd movement. Examples of these variables will involve individuals' environment and physical or psychological abilities. Insufficient inclusion of these factors in the synthetic dataset may jeopardize the reliability and applicability of the synthesized data [227]. These problems have their respective solutions at their disposal. One way to explore this includes improving computational models for artificial data production. Researchers can try improving models' accuracy, leading to more realistic synthetic data. Alternatively, developing synthetic data generation algorithms represents a viable option [228]. This involves using complex algorithms considering several variables and their interactions, greatly enhancing synthetic data's quality and realism. The algorithms should be flexible for various opportunistic surroundings to remain relevant and universal [229].

Crowd dynamic experts and those familiar with different subject matter areas can contribute tremendously to genuine synthetic data by bringing relevant knowledge necessary for its developmental improvement. Combining computational models with human judgment is essential for closing this gap between artificial and observable data sets and generating reliable and relevant databases for research or decision-making purposes [230]. Therefore, it highlights the need for representativity, randomness, and all significant variables in explaining a phenomenon or observable event. The part examines advanced computation-based designs, improved algorithms, and expertise based on expert input. This aims to generate artificial information mimicking the dynamics of opportunistic environments and provide useful planning knowledge [231].

3.3.5 Synthetic Data in Algorithm Development and Testing

Synthetic data is important in modelling and evaluating algorithms for crowd density estimation and behaviour predictability in opportunistic scenarios. Artificial data can be used at different levels for developing algorithms, from a first-stage idea to complete adjustment and tuning. For example, testing algorithms under different sets of controlled conditions and scenarios give more robust, accurate, and reliable algorithms before they are used in real-world situations [232]. The utility of synthetic data has been identified in several applications such as simulation and prediction research, hypothesis testing, algorithm testing, epidemiology/public health researchers, healthcare IT system development, education, and training, releasing open datasets, and linking of data sets.[233] Firstly, it allows the creation of a realistic dataset without fear of breaching privacy. Furthermore, it quickens the development cycle, saving effort and cost while shortening the development process.

On this note, synthetic data will help reduce the effort needed for data collection and labelling while working on computer vision tasks. Nonetheless, the artificial data must be realistic for practical purposes [234]. A decompose-wavelet algorithm was designed for seismic data processing; it was applied to synthetic and true data and proved better multiple attenuations similar to those in commercial programs. Moreover, a prediction model for scholarship receivers has been created in data mining using the kNN algorithm and SMOTE. The model's accuracy, precision, recall, and F1 scores were remarkable, suggesting its ability to forecast future beneficiaries. [235, 236, 237]

3.4 Theoretical Concept: Mobile Crowd Density Estimation

A methodology to assess crowd density using telecommunication data through BTS is shown in the following figure 3-1.

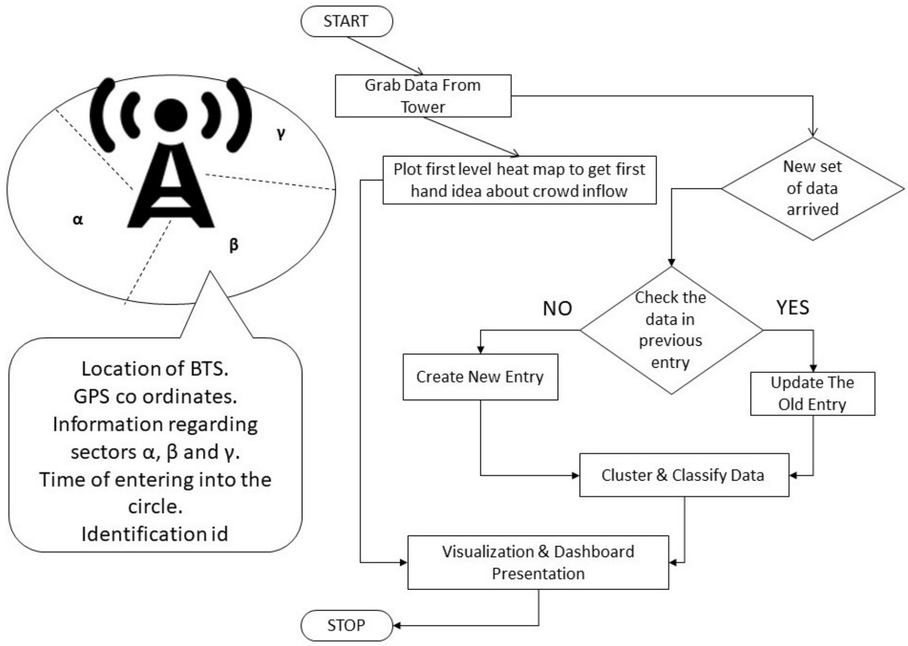


Figure 3-1: Conceptual Flow of Mobile Crowd Density Estimation

The raw data is generated and extracted through a network of strategically placed BTSs assigned sequential alphanumeric identifications. Alpha (A) is associated with the BTS units interacting with mobile devices in corresponding sectors: alpha, beta, and gamma. The combined effort of these sectors encompasses an overall supervision field, collecting necessary information critical to mobile movement analysis. The first step in analysis involves creating a heat map upon data attainment. This is a very direct way of showing the spatial distribution and the density of the population of the catchment area of the BTS. The study provides a preliminary insight into how crowds behave spatially and temporally. Incoming data continues, and the subsequent branching of the decision-making flow occurs after visualization. When a new data set is added, it updates the existing database on ground reality.

On the contrary, if no new data is available, a new data entry should be created, and this additional information should be added to the dataset. At this point, the core of the analysis is found in the clustering and classifying phase, where aggregate data is segregated according to defined criteria. This step is key because it divides the data into clusters that reflect different modes of crowd dynamics. Subsequently, each cluster undergoes a classification procedure to unravel the hidden patterns, possibly yielding valuable intelligence about behavioural trends and path movements over the specified areas, presenting marks the end of a theoretically built-up conceptual diagram. Data storytelling is depicted as carefully laid out data aggregated, clustered, and ranked on a dashboard that acts as a canvas. This dashboard, however, is more than just a data display; it is an interactive interface that allows meaningful information retrieval out of complex volumes.

3.5 Mathematical Foundations of Crowd Density Estimation in Opportunistic Environments

Complete algorithm of quartile classification of the crowd density analysis. In this regard, developing an algorithm based on all parameters necessary for effective urban crowd dynamics management is crucial. In this part of the chapter, an explanation of the mathematical constructs and principles informing the algorithm, particularly concerning the quartile category classification and implementation in opportunistic environment are discussed.

3.5.1 Initialization: Locations and Mobile Towers

An initial stage per a city's landscape is based on MOBILE towers and their positions. Assuming 'n' denotes all unique sites, then "MobileTowers (n)" designates the distribution of Mobile towers in such locations. First, the program plots every MOBILE tower and places it next to its geographic area where it is located. In the first stage, the algorithm involves a precise scenario where the urban terrain is expressed as MOBILE towers. This process is fundamental to the subsequent crowd density estimation and requires a rigorous mathematical approach, as presented below in table:

Table 3-2: Nomenclatures

n : Total number of distinct geographic locations in the city.
$L = \{L_1, L_2, \ldots, L_n\}$: Set representing distinct locations in the city.
$T = \{T_1, T_2, \ldots, T_n\}$: Set of MOBILE towers, where each T_i Corresponds to a MOBILE tower.
$f(T_i)$: Function mapping MOBILE tower T_i to its location L_j.
(x_i, y_i) : Geospatial coordinates of MOBILE tower T_i.
$D(L_i)$: Density of MOBILE towers at location L_i.
$C(T_i, t)$: Crowd count at MOBILE tower T_i At time t.
$A(L_i)$: Area of location L_i.

A. Mathematical Framework

1 Geographical Mapping of MOBILE Towers:

Function Definition:

Define a mapping function $f: T \rightarrow L$ such that each MOBILE tower T_i is associated with a unique location L_j.

For every T_5 is located at a central square represented by L_3, then $f(T_5) = L_3$.

Geospatial Coordinate Assignment:

Assign geospatial coordinates (x_i, y_i) to each MOBILE tower T_i, denoting its precise physical location.

2 Density Function for Mobile Towers:

Density Calculation: The density of MOBILE towers in location L_i Is given by:

$$D(L_i) = \frac{\sum_{T_j \in L_i} 1}{A(L_i)}$$

Eq: (3-1)

Here, $\sum_{T_j \in L_i} 1$ counts the number of towers in L_i, and $A(L_i)$ It is the area of that location.

3 Time-Dependent Crowd Data Collection:

Crowd Count Model: Model the crowd data collection at each tower as a time-dependent function: $C(T_i, t) =$ crowd count at tower T_i At time t. This function is essential for analyzing temporal variations in crowd density.

B. Diagrammatic Representation

Representation of MOBILE Towers in figure 3-2: Each blue dot on the diagram symbolizes a MOBILE tower. These towers are distributed across a simulated 100×100grid, representing different areas within a city. Location Identification: The annotations (e.g., L1/T1, L2/T2) provide a dual identification for each point, indicating both the location $(L1, L2, \ldots)$ and the corresponding MOBILE tower (T1, T2, ...). This dual labelling is essential for understanding the specific placement of each tower within the urban layout. Geospatial Layout: The X and Y axes represent the geographical coordinates of the MOBILE towers. This kind of coordinate system brings the perception of the locations of various towers with each other on the city map and, thus, an understanding of the spatial distribution of those towers through the cities. Urban Planning Insights: One could also infer something about the city's urban planning through the distribution pattern of the towers. A large number of towers in a given place would suggest it is a high-population density or commercial zone, whereas a few towers in another area could point to low occupation or residential areas. Strategic Importance for Crowd Density Estimation: This demonstrates that the location of Mobile towers is very important when calculating crowd densities in urban settings. Analysing these towers' locations could help determine areas for potential crowd gathering and crowd movement.

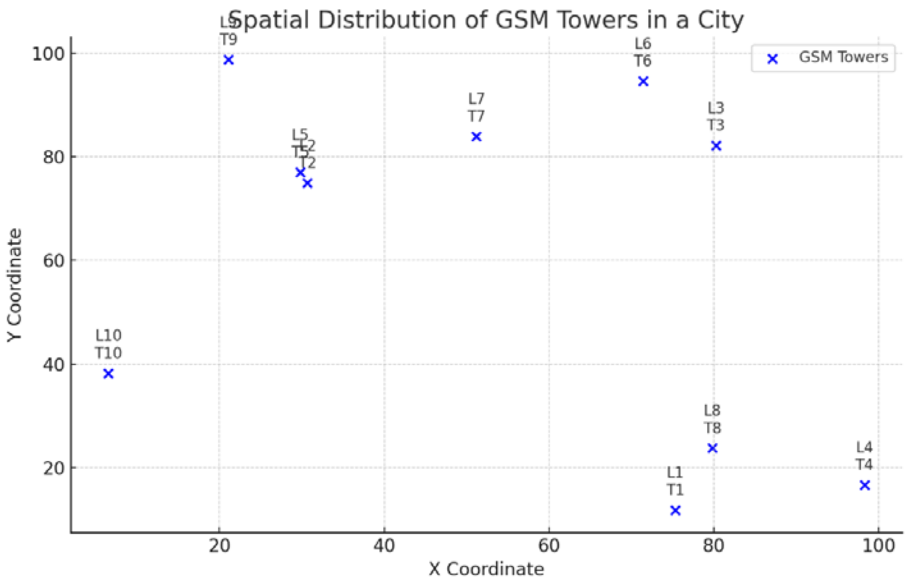


Figure 3-2: Scenario Presentation of Distribution of Mobile Towers

The diagram is key to the discussion in the thesis, especially since it connects the points where MOBILE towers are located and the crowd density data in managing cities and urban planning.

C. Analytical Significance

- The granularity of Analysis: This helps granularly analyse crowd dynamics in various urban settings because it is a precise mapping and density calculation undertaking.
- Resource Allocation: Tower distribution helps strategically deploy resources for effective crowd management.
- Dynamic Monitoring: Dynamic monitoring and forecasting of crowd movements becomes possible because of the temporal model C(Ti,t).

This part presents a complete mathematical and schematic foundation of the connection between MOBILE towers and urban regions. Such preparation is of great importance for fine-grained and evolving studies on crowd density. These are essential for successful urban planning as well as crowd management strategies. Using spatial mapping, density functions, and time-series-based data collection models provide an effective toolkit for handling urban crowd complexity.

3.5.2 Crowd Count Arrays

The ‘Ccount’ array is important for the crowd density estimate algorithm since it tracks the accumulated crowd information of each MOBILE tower hourly. This is made possible through these basic arrays on which comprehensive data about temporally changing crowd dynamics can be created. Intelligently, ‘Ccount’ arrays consider the crowd dynamics in different aspects. These matrixes or arrays are not simple repositories of information but sophisticated networks designed to reveal the subtle rhythms of urban motion.

A. Mathematical Framework

1 Advanced Array Structure:

- Definition of Multidimensional Array:

Let's define 'Ccount' as a four-dimensional array, represented mathematically as $Ccount_{i,j,k,l}$.

- Where:
  - i corresponds to the MOBILE tower index (1 to n towers),
  - j represents the hour of the day (1 to 24),
  - k denotes the day of the week (1 to 7), and
  - l accounts for additional parameters like event types or weather conditions.
- Data Entry Representation:
  - For each entry in the array, C count $t_{i,j,k,l}$, indicates the crowd counts at the i-th tower during the j-th hour, on the k-th day, under the l-th condition.

2 Comprehensive Data Collection Model:

- Detailed Accumulation Process:
- The crowd count data is gathered as per the following equation:

$$\text{Ccount}_{i,j,k,l} = \sum_{p=1}^{P} \delta_{p,i,j,k,l}$$

Eq: (3-2)

- Here, $\delta_{p,i,j,k,l}$ Represents the presence (1) or absence (0) of an individual p in the vicinity of tower i during hour j, on day k, under condition l.
- P denotes the total population under consideration.

3 Statistical Analysis within Arrays:

- Implementation of Statistical Measures:
- Apply statistical functions within each segment of the array for a granular analysis: Mean

$$(\mu): \mu_{i,j,k,l} = \frac{1}{H} \sum_{h=1}^{H}$$

Eq: (3-3)

Ccount $_{i,h,k,l}$

- Standard Deviation

$$(\sigma): \sigma_{i,j,k,l} = \sqrt{\frac{1}{H-1} \sum_{h=1}^{H} \left(\text{Ccount}_{i,h,k,l} - \mu_{i,j,k,l}\right)^2}$$

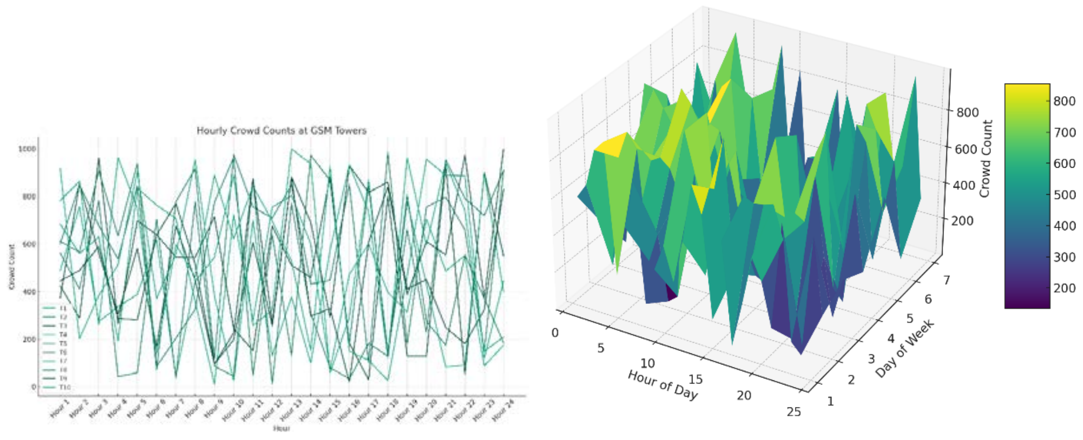
Eq: (3-4)

- H is the number of hours considered for each statistical calculation.

B. Diagrammatic Presentation

A three-dimensional surface plot illustrates in figure 3-3 (A & B) creates an effective and illustrative image picture of how population change occurred during specific times in a MOBILE tower. This plot depicts a cycle of 24 hours, where the X-axis describes each hour of the day. This is shown on the y-axis, where one can see the days of the week from Monday to Sunday. The vertical dimension of the plot is referred to as the Z-axis, and it shows the crowd count revealing increasing numbers of persons detected by MOBILE towers across different hours during a particular day. In particular, the peaks and troughs seen in the graph give important insights. Peaks were highlighted when most people were around the tower at one point. Such instances are vital for urban planners and authorities because they often indicate a need for more attention or resources.

On the contrary, the troughs, meaning where the surface goes down, embody the times of less-density crowds. Over time, analysis of such patterns may help understand the dynamics underlying daily and weekly city rhythms, thereby making informed decisions, including emergency preparedness and general urban planning. These variations are accented by a colour gradient, moving from cooler through warm colours, making the plot more than an operational device.



(A) (B)

Figure 3-3: Synthetic Presentation of the Crowd Density

‘Count’ is written in a much more advanced and complex manner that incorporates time dimension, variation based on location, and dependence on certain conditions. Sophisticated data structure with detailed mathematical modelling provides deeper analysis and forecasting needed in effective city governance.

3.5.3 Threshold Arrays

Threshold Arrays such as ‘T_{daily}’ and ‘T_{weekly}’ determine baseline densities for all days under study. Daily and weekly crowd patterns are compared with these baselines to spot normal trends and exceptional conditions.

A. Mathematical Framework

1 Calculation of Daily Thresholds:

'T_{daily}' represents the median crowd count for each MOBILE tower, calculated daily.

Mathematically, for a MOBILE tower i, the daily threshold on day d is defined as:

$$\mathrm{T}_{\text{daily } i,d} = \mathrm{median}\{\ \mathrm{Ccount}_{i,1,d}, \mathrm{C\ count}_{i,2,d}, \ldots, \mathrm{Ccount}_{i,H,d}\}$$

Eq: (3-5)

Where $\mathrm{Ccount}_{i,h,d}$ is the crowd count at tower i, hour h, on day d.

2 Calculation of Weekly Thresholds:

'$\mathrm{T}_{\text{weekly}}$' is calculated as the median of the daily thresholds for each MOBILE tower over a week.

For MOBILE tower i, the weekly threshold is defined as:

$$\mathrm{T\_weekly\ y}_i = \mathrm{median}\left\{\mathrm{T\_daily}_{i,1}, \mathrm{T\_daily}_{i,2}, \ldots, \mathrm{T\_daily}_{i,D}\right\}$$

Eq: (3-6)

Here, D is the number of days in the week.

B. Diagrammatic Presentation

A week-long graph illustrated in figure 3-4 for daily and weekly thresholds of crowd numbers at MOBILE towers is depicted as follows. This diagram comprises lines representing the median crowd count (Daily Threshold) for a particular MOBILE tower on diverse days from Monday through Sunday. These lines depict how the crowd density changes from peak levels at every tower and when it dips into a trough that indicates lower crowd presence. In the diagram above, the dashed horizontal lines mark each tower's average weekly threshold values. Weekly thresholds are vital in determining general patterns and any unusual manifestations of crowd movements. For example, a major variation from such trendlines would imply abnormal behaviour in the crowd and call for additional monitoring or action. As such, this detailed visual illustration depicts temporal patterns of crowds in urban settings for planning purposes and successful crowd control.

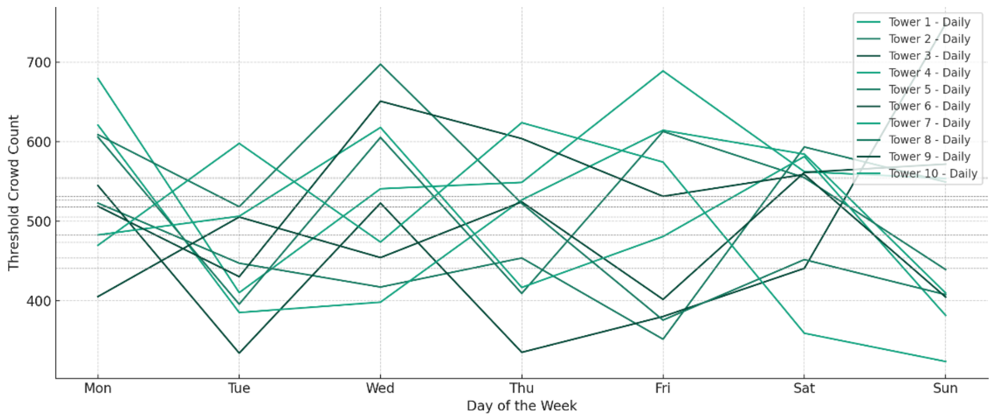


Figure 3-4: Synthetic Presentation of Threshold Computation

C. Analytical Significance

- Baseline Establishment: These thresholds give the basis for distinguishing average crowd behaviour from abnormalities.
- Anomaly Detection: Large variations away from the indicated levels indicate anomalies, allowing preventive crowd management as well as distribution of resources.
- Temporal Trends Analysis: Temporal trends in crowd behaviour can be understood by observing threshold variations on different days and at various times.

Baseline crowds' densities, captured by 'T_{daily}' and 'T_{weekly}' threshold arrays. The result of these calculations, coupled with the following analysis, is powerful in explaining crowd patterns and identifying the abnormalities. This makes the visual representation of these data more meaningful to urban planners and authorities working on the management of crowds.

3.5.4 Quartile Arrays

This analysis is based on quartile classifications. The arrays 'Q_{daily}' and 'Q_{weekly}' contain classifications of the quartiles of the crowd count data for days and weeks. This division makes it possible to stratify crowd densities into low, medium, high, and high categories.

A. Mathematical Framework

Crowd Counts Data Representation:

Let $C\ count_{h,i}$ represent the crowd counts at MOBILE tower i during hour h, with the dataset comprising hourly counts for each tower.

Flattening and Quartile Calculation:

The crowd counts are first flattened into a single array to apply quartile analysis across all towers and hours:

$$\text{Count}_{\text{flattened}} = \text{flatten}\left(C\ \text{count}_{h,i}\right)$$

Eq: (3-7)

Quartiles are calculated on this flattened data:

$Q1 = 25$ th percentile of $Ccount_{flattened}$

$Q2 = 50$ th percentile (Median) of Ccount flattened $_{(Men}$

$Q3 = 75$ th percentile of $C\ count_{flattened}$

Quartile Classification:

For each hourly crowd count, $count_{h,i}$ at MOBILE tower i during hour h, the classification based on the calculated quartiles is as follows:

If $Count_{h,i} < Q1$, then the classification is 'Least Crowded.'

If $Q1 \leq Count_{h,i} < Q2$, then the classification is 'Medium Crowded.'

If $Q2 \leq Count_{h,i} < Q3$, then the classification is 'Heavy Crowded.'

If $C\ count_{h,i} \geq Q3$, then the classification is 'Most Crowded.'

B. Diagrammatic Representation

The scatter plot of figures 3-5 and 3-6 presents a detailed graphical representation of the crowd density grading in every MOBILE. Crowd count, in hours, is plotted for each tower with one of the quartile-based labels distributed along the X-axis. Each point is coloured to depict its time of occurrence, thus facilitating comprehension of how crowd density fluctuates during various hours. Such a plot shows the patterns of crowd occurrences during certain hours and helps planners strategically manage the crowding of the various towers. For example, some towers are consistently classified as belonging to the “Most Crowded” category at particular hours of the day and may need more attention or even additional resources to ensure safety issues are addressed promptly. A plot becomes the holistic device that helps understand urban crowds' motions and provides urban planners and the authorities with a means of planning appropriate responses to them.

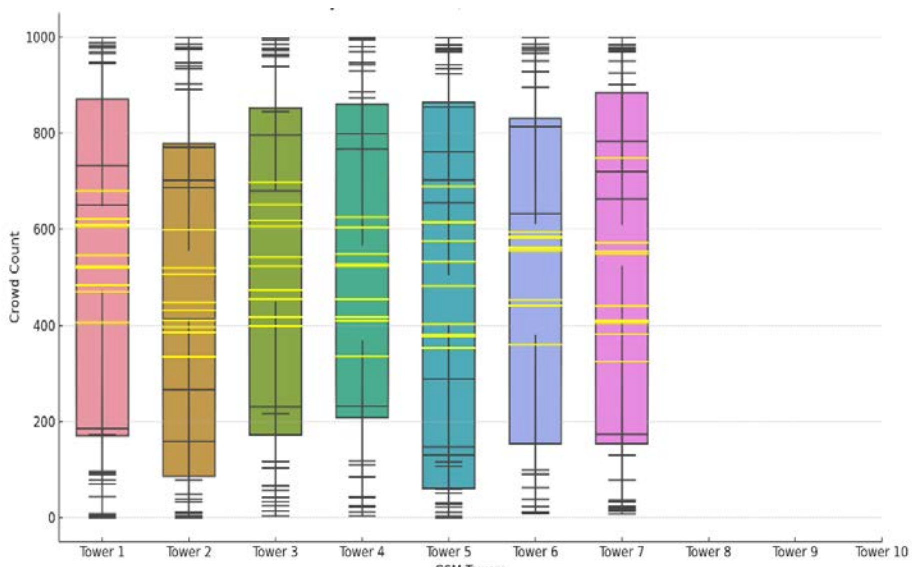


Figure 3-5: Synthetic MOBILE Tower level Crowd Density Distribution

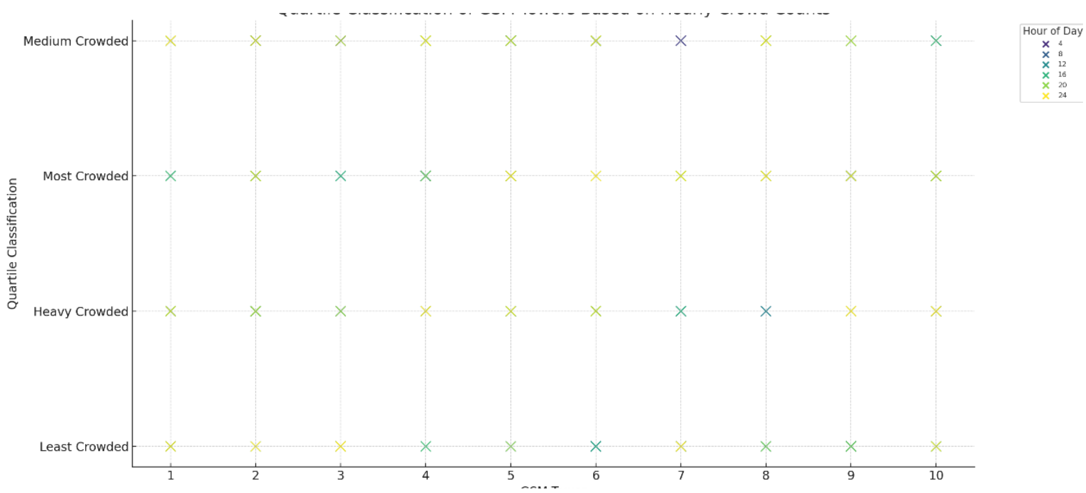


Figure 3-6: Synthetic Presentation of Quartile Classification of Crowd

C. Quartile Classification Scatter Plot

- The scatter plot shows different quartiles of the hourly crowding in the MOBILE tower.
- MOBILE towers (x axis), Classification to different categories (least crowded – most crowded).

- Color-coding each point to mark the hour of the day creates a time dimension.

The hourly crowd data is classified into low, medium, heavy, and super-heavy densities. Applying this classification to the tower by an hour, it will be possible to classify each tower in any of the four tower's density levels during a specific time frame. In this way, we get a detailed picture of the difference in crowd density between various towers and different times of the day for every tower.

3.5.5 Time Tracking Variables

Variables 'hour' and 'day' indicate the times their value occurred. The "hour" term embraces the daily/24-hour cycle and a week as "a day." Hour' and 'day' as time-tracking variables in a crowd density estimation context are critically important to appreciate the temporal aspects of the data. The variables allow us to look at the pattern of people across certain hours on a specific day or days of the week.

A. Mathematical Framework

Representation of Time Variables:

Let hour $\in \{1,2,\ldots,24\}$ represent the hours of the day, covering the entire 24-hour cycle.

Let day $\in \{1,2,\ldots,7\}$ denote the days of the week, addressing the weekly cycle.

Temporal Data Matrix:

Consider a matrix M where each element $M_{d,h}$ represents the crowd count data at hour h on day d.

The matrix dimensions are 7×24, corresponding to the days of the week and hours of the day.

B. Diagrammatic Representation

A weekday heatmap illustread in figure 3-7 offers a graphical view of variations in crowd density across a normal week. A heatmap, where each cell represents a certain hour of a particular day whose color intensity indicates the crowd density. This visualization makes it easier to note trends like peak crowd hours and densest or least dense days. The same could be used; for instance, bright colors can point at rush hours or weekends, while dark colors can refer to a night silence. Urban planners and authorities with information about these trends can better allocate resources, design the best crowd management strategies, and determine areas requiring additional infrastructure or improved safety measures.

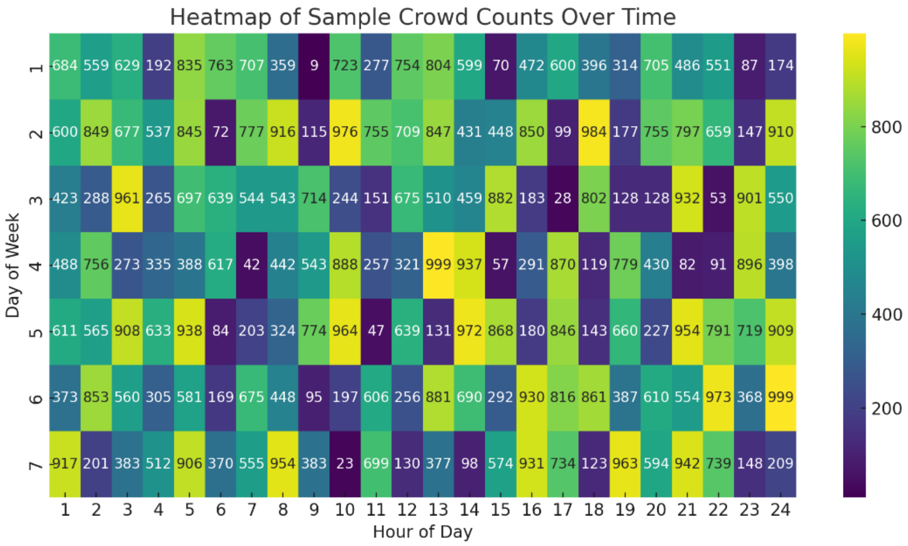


Figure 3-7: Synthetic Temporal Crowd Density Presentation

3.5.6 Data Collection for Each Hour and Each Day

During every hour of the day, 'Ccount' is collected by the algorithm. Currently, crowd size data signifies how the dynamics in the crowd density can be assessed in real-time. 'Ccount' data for each hour per day are essential in crowd density analysis. This information constitutes a key point in determining the change in crowd density along different temporal periods with real-time purposes.

A. Mathematical Framework

Hourly and Daily Data Collection:

Let Count cour,h, i $_{-}$ Represent the crowd count at MOBILE tower i during hour h on day d.

The data collection spans a week, hence $d \in \{1,2,\ldots,7\}$ and $h \in \{1,2,\ldots,24\}$.

Data Representation:

A three-dimensional matrix M is used to represent this data:

$$M_{d,h,i} = \text{Ccount}_{d,h,i}$$

Eq: (3-8)

Where the dimensions of matrix M are determined by the number of days, hours, and towers.

B. Diagrammatic Presentation

A three-dimensional graph illustrated in figure 3-8 shows hourly crowd counts at each MOBILE tower for one week. The days of the week are captured on the X-axis, and the Z-axis denotes the crowd count. The crowd counts for various towers every day, and their hour is indicated by each line in the plot respectively.

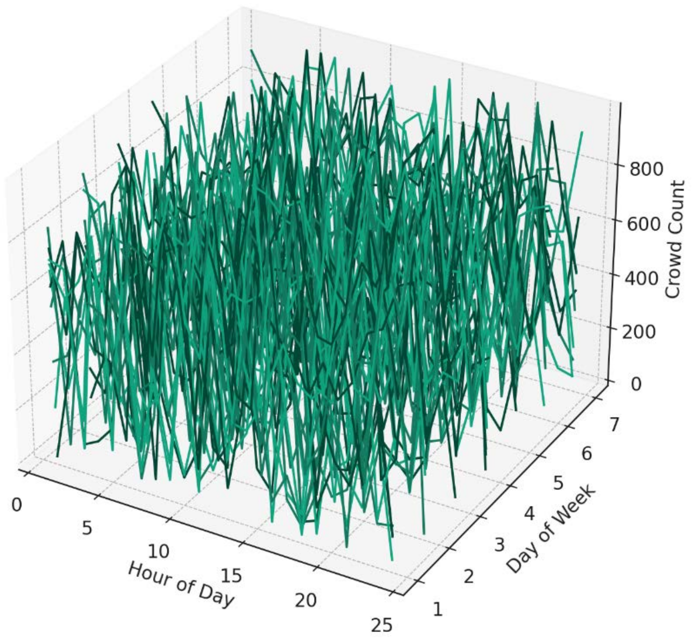


Figure 3-8: Synthetic Presentation of Hourly Crowd Density

A 3D plot comprehensively and vividly represents the crowd's density around MOBILEs during the week. The plot has each line showcasing the hourly counts of the crowd per a specific MOBILE tower so that the changes in crowd density over various days and hours are visible. Fluctuations in the line indicate high and low traffic hours for every tower. Such pattern identification lets one observe an upward trend on certain days, including rush hours and special events with many people. This analysis of the patterns provides an advantage to city planners and the related authorities in terms of proper optimization of resource distribution along with strategies for crowd control that vary with the time/place where this high crowd density occurs.

3.5.7 Calculating Daily and Weekly Threshold Values (Median-of-Median)

Daily and weekly thresholds should be calculated in crowd density analysis through the Median-of-Medians method to create credible standards. This technique offers a stronger measure of central tendency than other methods that do not adequately account for variability and outlier information. The Median-of-Medians procedure obtains the robust threshold for data sets containing outliers or skewed distribution. This way entails coming up with an arithmetic mean involving the median of the medians as a central point.

A. Detailed Calculation for Daily Thresholds:

For daily thresholds, the process is as follows:

First, for each hour h of day d, calculate the median crowd count across all towers, denoted as $\mathrm{M}_{\mathrm{d,h}}$.

$$\mathrm{M}_{\mathrm{d,h}} = \mathrm{median}\left(\left\{\mathrm{Count}_{\mathrm{d,h,1}}, \mathrm{Ccount}_{\mathrm{d,h,2}}, \ldots, \mathrm{count}_{\mathrm{d,h,n}}\right\}\right)$$

Eq: (3-9)

Then, for each day d, compute the threshold $\mathrm{T\_daily}_{\mathrm{d,h}}$ as the median of these hourly medians up to hour h.

$$\mathrm{T\_daily}_{\mathrm{d,h}} = \mathrm{median}\left(\left\{\mathrm{M}_{\mathrm{d,1}}, \mathrm{M}_{\mathrm{d,2}}, \ldots, \mathrm{M}_{\mathrm{d,h}}\right\}\right)$$

Eq: (3-10)

Weekly Threshold Calculation:

For weekly thresholds, the method involves aggregating daily medians:

Calculate the median for each hour h across different days.

$$\mathrm{M}_{\mathrm{weekly,h}} = \mathrm{median}\left(\left\{\mathrm{M}_{\mathrm{1,h}}, \mathrm{M}_{\mathrm{2,h}}, \ldots, \mathrm{M}_{\mathrm{7,h}}\right\}\right)$$

Eq: (3-11)

The weekly threshold for each hour h is then:

$$\mathrm{T\_weekly}_{\mathrm{h}} = \mathrm{M}_{\mathrm{weekly,h}}$$

Eq: (3-12)

B. Diagrammatic Presentation

These threshold values are represented in figure 3-9 in a complex line plot. They occur for daily thresholds and are shown consecutively as time goes on during one day. A separate line that shows weekly thresholds indicates an overall chart for one week.

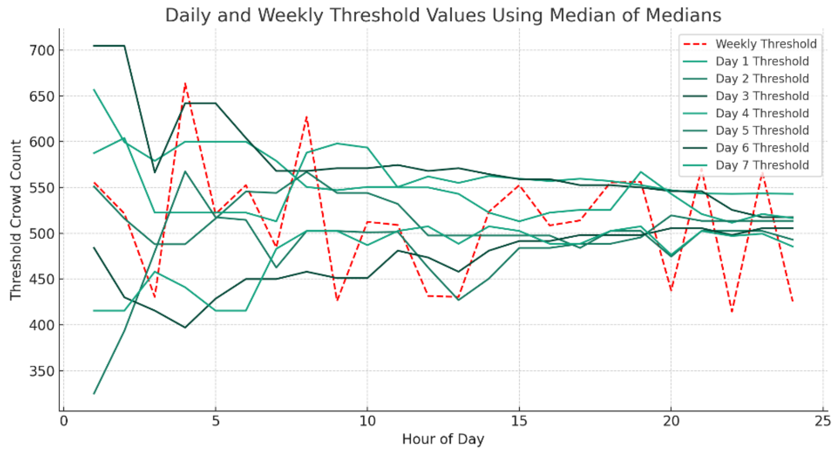


Figure 3-9: Synthetic Presentation of Median-of-Median Thresholds

The enhanced line plot provides a sophisticated visualization of the threshold values computed using the Median-of-Medians technique. The line thresholds daily demonstrate how the central value of crowd counts changes along the daily days in a usual crowd pattern to minimize the effect on outliers. The trendline for weekly thresholds is more general and means the typical trend of crowd density per hour throughout the week. The adjusted way of determining thresholds provides a better understanding of mob behaviour and helps plan for unforeseen events like terrorist attacks on the capital city.

3.5.8 Quartile Classification for Crowd Density (With Threshold)

The algorithm classifies each 'Ccount' entry into quartiles based on the distribution of crowd counts. The crowd density is then categorized into quartiles (Q1, Q2 [median], Q3, Q4), corresponding to density levels (Low, Medium, High, and Very High). This classification is relative to the corresponding threshold ('T_{daily}' or 'T_{weekly}') to ascertain if the crowd count exceeds these thresholds.

Each crowd count entry is denoted as $C_{count}(T_i, h, d)$, for a specific MOBILE tower T_i, hour h, and day d, are classified into quartiles. These quartiles are derived based on the statistical distribution of crowd counts. Mathematically, the quartiles $Q1, Q2, Q3$, and $Q4$ are calculated using the empirical distribution function $F_{C_{count}}$, where Q2 is the median. The quartiles correspond to density levels: Low, Medium, High, and Very High.

The quartile classification is formulated as follows:

$$\begin{cases} \text{Low} & \text{if } C_{count}\,(T_i, h, d) \leq F^{-1}_{C_{count}}\,(0.25) \\ \text{Medium} & \text{if } F^{-1}_{C_{count}}\,(0.25) < C_{count}\,(T_i, h, d) \leq F^{-1}_{C_{count}}\,(0.5) \\ \text{High} & \text{if } F^{-1}_{C_{count}}\,(0.5) < C_{count}\,(T_i, h, d) \leq F^{-1}_{C_{count}}\,(0.75) \\ \text{Very High} & \text{if } C_{count}\,(T_i, h, d) > F^{-1}_{C_{count}}\,(0.75) \end{cases}$$

Eq: (3-13)

This classification is relative to the corresponding thresholds, T_{daily} or T_{weekly}, to ascertain if the crowd count exceeds these thresholds.

A. Storing Quartile Classifications

Quartile classifications are stored in 'Q_{daily}' and 'Q_{weekly}' for future reference and analysis. Quartile classifications are systematically stored in matrices. Q_{daily} and Q_{weekly} for subsequent analysis and reference. This storage facilitates the tracking of crowd density trends over time.

B. Mapping Crowd Density for Prediction

A mapping of crowd density versus threshold values is derived using' historical 'Ccount' data and quartile classification. It is useful in estimating future crowds' densities and distributions. Predictive mapping uses historical crowd count data as well as quartile classifications. This mapping is denoted as $M_{density}\,(C_{count}\,, Q_k)$, correlates crowd density levels with threshold values, forming the basis for future crowd density predictions.

C. Crowd Steering and Management

The algorithm's predictions and classes go into forming the steering strategies for efficient crow control during peak hours or when major crowds gather, especially at places with many visitors. Crowd management strategies also involve planning how to steer or direct people based on the algorithm's predictions and classification of individuals. Dynamic crowd control becomes very critical at times of high traffic volumes and when large-scale events are taking place because it greatly enhances security and effectiveness. These mathematical expressions unequivocally define this quartile classification process.

D. Diagrammatic Presentation

The infographic illustrates in figure 3-10 about the Quartile Classification for Crowd Density in an urban environment. It includes various elements such as a graph showing the distribution of crowd counts with quartiles, a matrix representation of 'Q_{daily}' and 'Q_{weekly},' a flowchart for predictive mapping, and a diagram for crowd steering and management strategies. This visual representation should aid in understanding the classification process and its application in urban planning and crowd management.

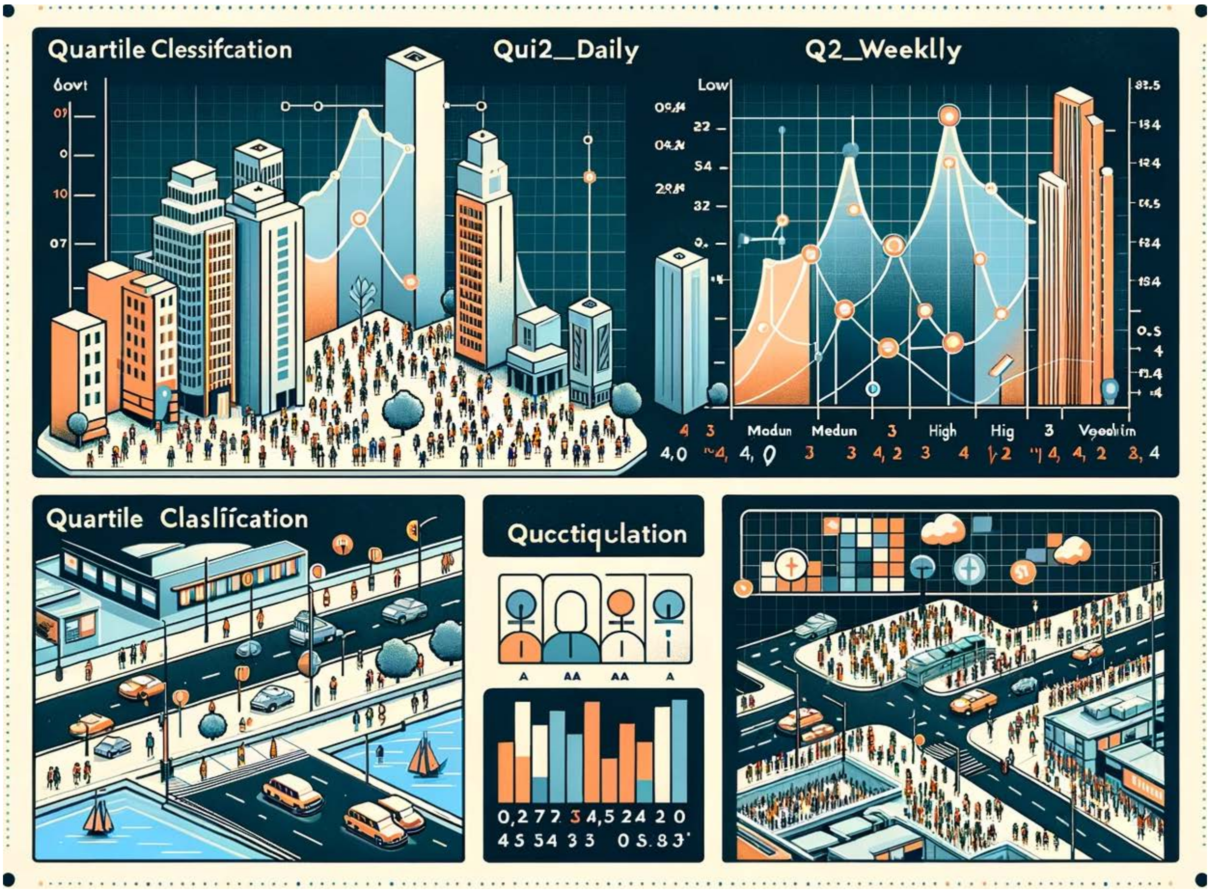


Figure 3-10: A Preview of Crowd Density || Source: Conceptually AI-Generated

3.6 Social Network Mobility Process

3.6.1 Individual Tracking

Each individual's connection to MOBILE towers is tracked hourly. This granular data collection is crucial for understanding individual mobility within the urban space. The process involves monitoring each individual's interactions with MOBILE towers within an urban space. These interactions are logged hourly, allowing for a detailed understanding of mobility patterns.

- **Individual-Tower Interaction**

Let $I(t, u, T_i)$ Represent the interaction of individual u with the MOBILE tower. T_i At time t.

$$\begin{cases} 1 & \text{if individual u is connected to a tower } T_i \text{ at time t} \\ 0 & \text{otherwise} \end{cases}$$

- **Hourly Data Collection:**

For each hour h of the day, the individual-tower interactions are aggregated to form a matrix representation, $M_{u,h}$.

$$M_{u,h} = [I(h, u, T_1), I(h, u, T_2), \dots, I(h, u, T_n)]$$

Eq: (3-14)

Here, T_n Represents the total number of MOBILE towers in the observed area.

- **Mobility Pattern Analysis:**

Over a period P, the individual mobility pattern $MP(u, P)$ is determined by analyzing the sequence of matrices. $M_{u,h}$.

$$MP(u, P) = \{M_{u,h_1}, M_{u,h_2}, \dots, M_{u,h_{|P}}\}$$

Eq: (3-15)

This sequence captures the Spatio-temporal mobility pattern of individual u across different hours and locations.

The granularity of the data collection allows for detailed modelling of individual mobility presented in figure 3-11. Time-series analysis and spatial modelling can identify individual movement trends and patterns. Statistical methods such as Markov chains or clustering algorithms may be employed to analyse these patterns.

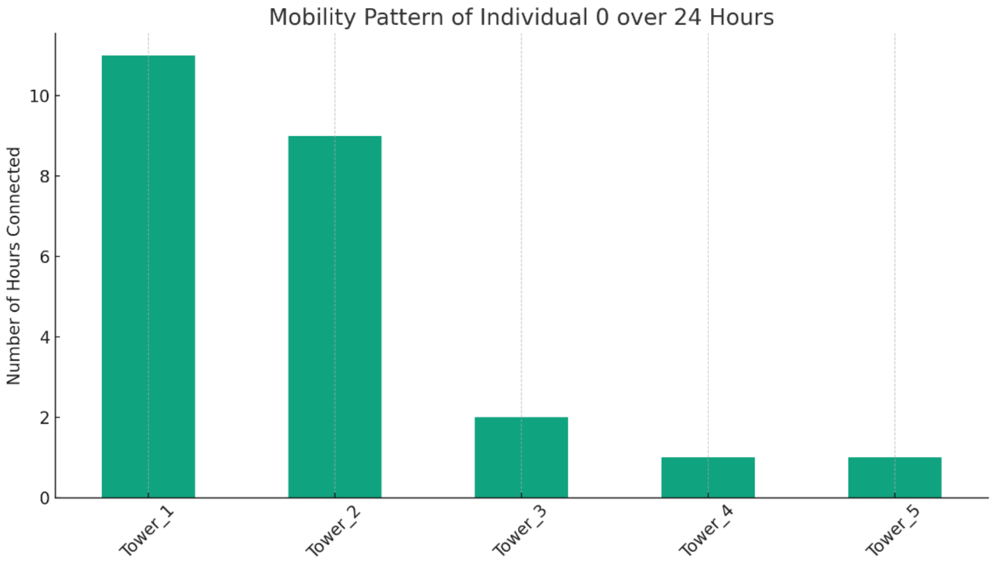


Figure 3-11: Conceptual Individual Mobility Pattern

3.6.2 Location Mapping

Every MOBILE tower is mapped to a specific location in the city, enabling a detailed analysis of movement patterns relative to geographical points. Location Mapping is pivotal in understanding the spatial dynamics of urban mobility. It involves assigning every MOBILE tower to a precise geographical coordinate within the city. This mapping facilitates the analysis of movement patterns about specific geographical points.

A. Geographical Coordinates:

Each MOBILE tower, denoted as T_i, is associated with a set of geographical coordinates (x_i, y_i) Representing its location.

$$T_i \leftrightarrow (x_i, y_i)$$

Eq: (3-16)

Here, x_i and y_i Are the latitude and longitude coordinates, respectively, which precisely locate T_i On the urban map.

B. Spatial Mapping Function:

A function $f: T \rightarrow \mathbb{R}^2$ is defined to map each tower to its corresponding geographical point.

$$f(T_i) = (x_i, y_i)$$

Eq: (3-17)

This function f translates the set of all towers T into a two-dimensional real space $\mathbb{R}^2$, effectively creating a spatial representation of the MOBILE network within the city.

3. Individual Movement Pattern Analysis:

With the spatial mapping established, individual movement patterns can be analyzed regarding their geographic trajectory. For an individual u moving across towers $\{T_{i1}, T_{i2}, \ldots, T_{ik}\}$ over time, their movement trajectory Γ_u It is represented as:

$$\Gamma_u = \{f(T_{i1}), f(T_{i2}), \ldots, f(T_{ik})\}$$

Eq: (3-18)

This trajectory is illustrated in figure 3-12 which provides a geographic path of individual u 's movement within the urban space.

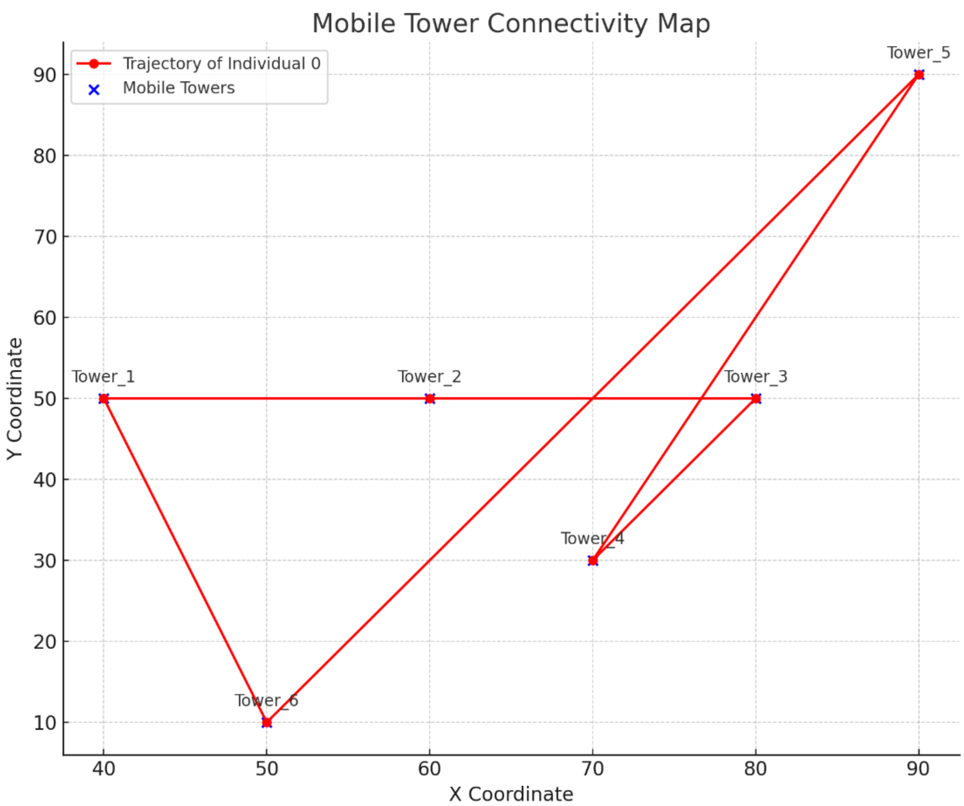


Figure 3-12: Conceptual Representation of Individual Mobility Pattern

3.6.3 Analysis of Urban Dynamics

Aggregating these individual trajectories across the population yields insights into urban mobility dynamics. Statistical spatial analysis tools like heat maps or spatial clustering can identify high-traffic areas, movement trends, and potential urban bottlenecks.

3.6.4 Mobility Pattern Analysis

Individual movement profiles are created based on their hourly connections to MOBILE towers. This data is used to trace movement paths throughout the day, providing insights into individual mobility patterns within the city. Mobility Pattern Analysis involves constructing individual movement profiles from hourly connections to MOBILE towers. This process involves tracking and analysing the sequential connection data of individuals to various MOBILE towers throughout the day.

- **Hourly Connection Data:**

Let $C(u, T_i, h)$ denote the connection of individual u to the MOBILE tower T_i During hour h.

$$\begin{cases} 1 & \text{if u is connected to } T_i \text{ at hour h} \\ 0 & \text{otherwise} \end{cases}$$

- **Construction of Movement Profiles:**

The movement profile for an individual u over a day (comprising H hours) is represented as a sequence of connections:

$$MP(u) = \{C(u, T_1, 1), C(u, T_2, 1), \dots, C(u, T_n, H)\}$$

Eq: (3-19)

Here, T_n Denotes the total number of MOBILE towers in the observation area.

- Path Tracing:

For each individual u, a path of movement $P(u)$ is traced using the sequence of their hourly connections.

$$P(u) = \{T_{i1}, T_{i2}, \dots, T_{ik}\}$$

Eq: (3-20)

where T_{ij} It is the MOBILE tower to which u is connected at the j-th hour.

3.6.5 Analysis of Mobility Patterns

Statistical analysis is applied to MP(u) to discern patterns in an individual's mobility. Techniques such as time-series analysis, clustering, and sequence pattern mining can be utilized to identify regular routes, frequent locations, and temporal trends in individual mobility.

- **City-Wide Mobility Insights**:

Aggregating individual movement profiles across a population offers a comprehensive view of urban mobility patterns. This aggregated data can be analyzed to understand peak movement hours, popular routes, and spatial distribution of population movement within the city.

3.6.6 Identifying Groups and Relationships

3.6.6.1 Group Identification

Groups of individuals connected to the same MOBILE tower simultaneously are identified the illustration presented in figure 3-13. This grouping helps in understanding collective movement patterns.

Let $G(t, T_i)$ Be the group of individuals connected to the MOBILE tower. T_i At time t.

$$G(t, T_i) = \{u \mid I(t, u, T_i) = 1\}$$

Eq: (3-21)

Here, $I(t, u, T_i)$ Is the indicator function defined previously representing whether individual u is connected to the tower T_i At time t.

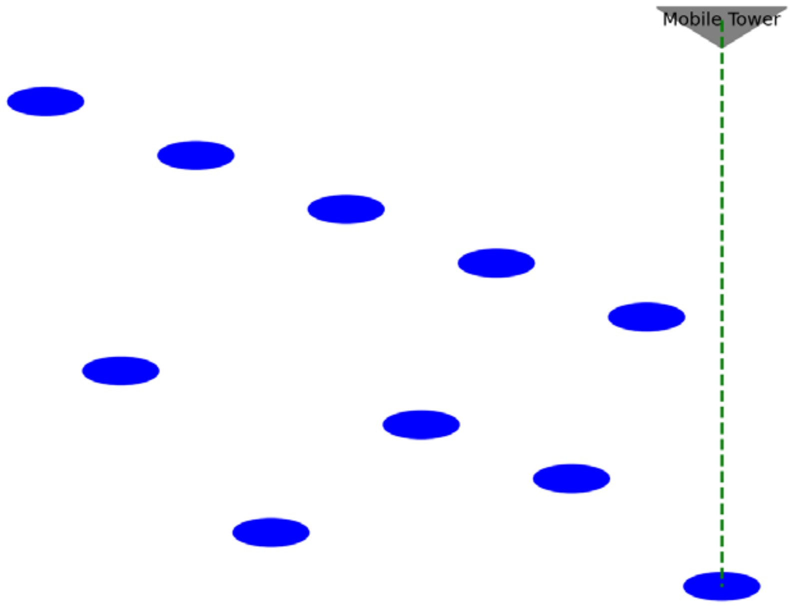


Figure 3-13: Conceptual Presentation of Group Identification Connected to MOBILE Tower

3.6.7 Relationship Inference

Figure 3-14 illustrates the social ties are inferred for groups that consistently appear together across various locations over time. These ties are classified as 'social' for frequent groupings and 'random' for sporadic occurrences.

- **Social Ties Definition:**

Define a social tie $S(u, v)$ between two individuals u and v based on their cooccurrences in groups across various locations over time.

$$S(u, v) = \sum_{t,T_i} 1_{\{u,v \in G(t,T_i)\}}$$

Eq: (3-22)

Where 1 is the indicator function that equals 1 when u and v are both in group $G(t, T_i)$, and 0 otherwise.

- **Classification of Ties:**

Social: If $S(u, v)$ exceeds a certain threshold.

Random: If $S(u, v)$ is below that threshold.

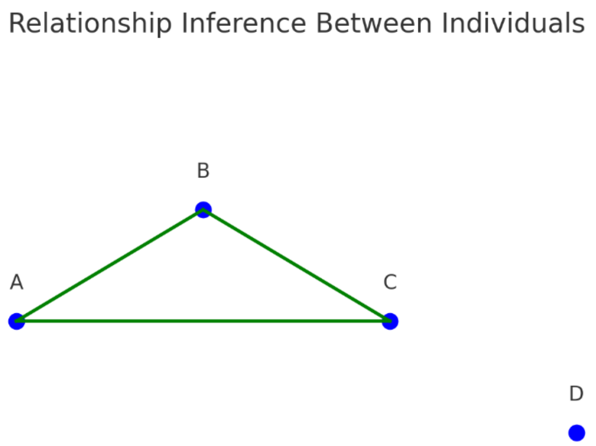


Figure 3-14: Conceptual Presentation of Individual Social Ties

3.6.8 Time and Location Analysis for Relationships

The duration and frequency of time spent together at the same MOBILE tower locations are analyzed to strengthen the inference of social ties.

- Duration and Frequency Analysis:

Analyse the total duration $D(u,v)$ and frequency $F(u,v)$ of co-occurrences of u and v at the same locations.

$$D(u,v) = \sum_{t,T_i} \Delta t \cdot 1_{\{u,v \in G(t,T_i)\}}$$
$$F(u,v) = \sum_{T_i} 1_{\{\exists t: u,v \in G(t,T_i)\}}$$

Eq: (3-23)

Here, Δt is the duration of each time interval.

3.6.9 Integration with Crowd Density Analysis

This individual and group movement data is integrated with the crowd density analysis from the algorithm's first part. This integration is key to understanding how individual and group movements contribute to crowd dynamics.

Integrate individual and group movement data $M(u,t)$ and $G(t,T_i)$ with crowd density $C(T_i,t)$.

$$C(T_i, t) = |G(t, T_i)|$$

Eq: (3-24)

$C(T_i, t)$ represents the number of individuals at the tower T_i At time t, reflecting crowd density.

3.6.10 Predictive Modelling for Social Dynamics and Crowd Movement

Historical data on individual movements and inferred social ties are used to build predictive models. These models can forecast crowd movements and social dynamics under various scenarios.

3.6.10.1 Historical Data Utilization:

Use historical data on individual movements and social ties to build predictive models.

$$P(Y_{future} \mid X_{historical})$$

Eq: (3-25)

Where Y_{future} represents future crowd movements and social dynamics, and $X_{historical}$ includes historical movement data and inferred social ties.

3.7 Conclusion

The conceptual framework chapter provides a detailed mathematical exposition of an algorithm for analysing crowd density and individual mobility patterns in opportunistic urban environments. The algorithm combines quartile-based crowd density analysis with granular tracking of individual movements and social ties, offering a robust framework for understanding and managing urban crowd dynamics. The mathematical formulations and principles detailed here are pivotal for implementing effective crowd management and urban planning strategies, especially in dynamic and unpredictable urban settings.

CHAPTER 4: PROPOSED METHODOLOGY

Urban mobility and social dynamics are increasingly complex, requiring us to examine how people and collectives relate in a city's MOBILE environment at a granular level. The approach here will involve minutely tracing both group and individual movements, inferring the social networks, and linking these findings to emerging city structures and trends. Our approach centres around high-level detailing of connections between each person and a specific MOBILE tower. This is done through hourly tracking of people's contacts with these towers to have a spatial-temporal representation concerning urban mobility. All activities are minutely recorded, which forms a strong basis for future analysis.

In the centre of the concept of 'Individual-Tower Interaction,' represented mathematically as $I(t, u, T_i)$, which indicates whether an individual u is connected to a specific tower T_i At a given time t. This forms the basis for constructing hourly matrices. $M_{(}u, h)$ individual-tower interactions are pivotal in discerning individual mobility patterns over a given period. The proposed technique then moves further and identifies clusters of people simultaneously associated with similar MOBILE towers. However, this part of the research sheds some light on the movement patterns as a group and constitutes the beginning of our social network analysis.

The research uses advanced inference algorithms to deduce the latent social structure within these movements based on direct observations and measurements. It is based on the assumption that social relations can be deduced from how often, frequently, and consistently different groups within several MOBILE towers coalesce over time. Using some defined thresholds, these relationships are referred to as social for frequent interactions while random for lesser cases. The analysis, of course, includes assessing the time and regularity of such communications. We, therefore, seek to strengthen the inference of social ties by using the information on how long callers interacted with each other at each of the MOBILE tower locations where they were. The proposed method includes crowd density analysis that combines personal and group motion information with the aggregated MOBILE network-based crowd density measures. This fusion is necessary for depicting an entire picture of urban movement and social process dynamics. Developing a predictive model is where methodology peaks. Using historical trends of individual movement and inferred social linkages, these models predict how crowds will move and interact in a host of hypothetical situations. Based on advanced statistical techniques and machine learning procedures, the models focus on the intricacy and diversity of urban social networks.

A systematic and data-oriented research methodology is used to learn more about MOBILE's complicated fabric surrounding urban mobilities and social relations. The foundation is based on intensive data collection, a strict analytic approach, and a sophisticated predicting model that penetrates subtle processes in the city lifestyle. The subsequent sections of this chapter talk about the methodology, the proposed algorithm, and the data for this study at the granular level.

4.1 Proposed Framework

The depicted framework in figure 4-1 outlines an integrated system for monitoring and predicting crowd dynamics using many data sources. It harnesses information from mobile devices, Wi-Fi networks, and vehicular networks alongside geospatial data from social networks. The system processes this data to compute crowd density, assess mobility patterns, and evaluate social connections, ultimately visualizing its predictions and assessments through a sophisticated dashboard.

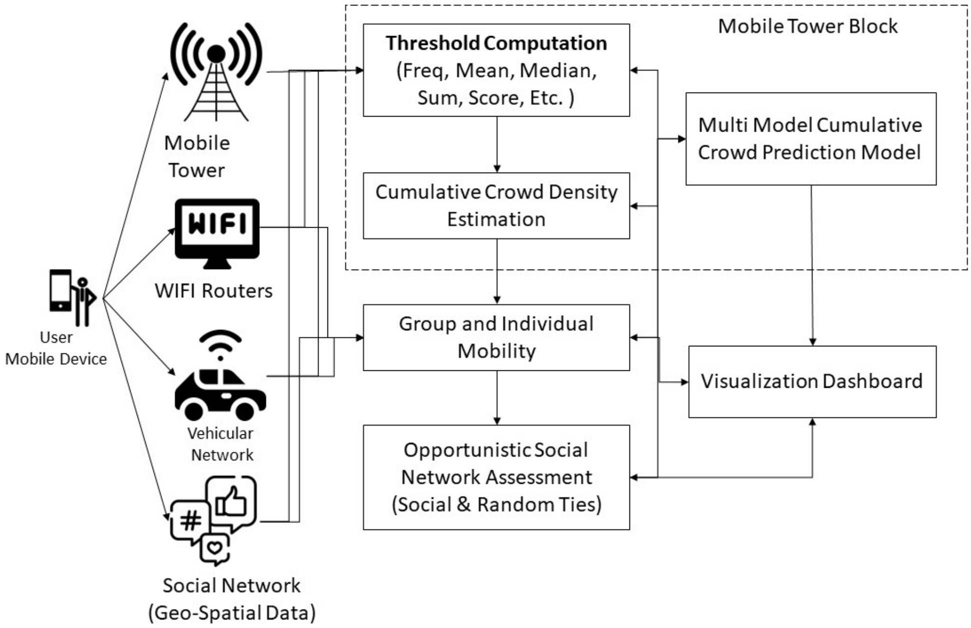


Figure 4-1: Proposed Framework

A network of multiple interacting components in terms of assessing the crowd approach is depicted on the diagram. It collects data from mobile phones, Wi-Fi routers, vehicle networks, and social networks for comprehensive crowd analysis. Baseline and benchmark threshold computation enable the system to analyse. The analysis of cumulative crowd density enables one to determine mobility behaviour in groups and on personal levels. The analysis also includes a study of social networks to detect social and chaotic communication relations between people. At the same time, a multi-model cumulative crowd prediction model is based on a mobile tower block to improve its reliability. These analytical processes generate outputs in a visualization dashboard, enabling users to understand the presented information and conclusions better. Each module will be discussed in detail, showing their algorithms and methodologies for data extraction, processing, and presentation. A critical analysis of this framework and its practical application, including its effectiveness in understanding and management of crowd dynamics at various levels, will be carried out on a detailed basis.

4.2 Data Collection and Sources

The analytical approach starts by looking at crowd density, which gradually leads to other factors influencing crowd steering, such as the directional behaviour of individuals and groups, movement patterns among the crowds, and eventually social and random ties. This hierarchal approach is necessary for viewing city development processes on scales ranging from global to local.

Crowd density analysis is the subject matter of discussion here. During this phase, the concentrations of people are being measured in particular urban regions at various times. Such analysis is crucial to locating areas with a high concentration of people and comprehending trends associated with aggregation within cities. This phase relies heavily on data drawn from MOBILE tower records indicating the population demography as it varies with the network coverage. Therefore, the research focuses on crowd steering analysis after that. In this case, we strive to understand how crowds move and dissipate in urban areas, such as examining the movement of people in terms of where they usually go and finding the most common paths and bottlenecks. Complementing vehicular trip data and wi-fi network information with Mobile data sheds light on the intricacies of crowd movement and paints a broader picture of urban mobility.

This section explores mobility patterns at both individual and group levels. Such fine-grained analysis is very important to understand the subtle intricacies of urban mobility. Analysing specific paths and clusters reveals how people roam about the city, talk to one another, and come together into certain groups. In this phase, social media provides an important source, especially data from Twitter and Instagram, offering a valuable contextualized dimension to behavioural studies.

The proposed research examines how social and random ties are crafted in these mobility patterns. This refers to determining how frequently individuals or groups interact in different locations and periods. This study seeks to identify social ties, suggestive of sustained and relevant links, versus random ties, which encompass casual and random meetings. Such analysis provides the basis for studying the social texture of urban mobility and the interaction of people in city settings. The study follows a layered approach that begins by analysing the crowd density before narrowing it down to behavioural and psychological assessments in an urban setting. The study integrates diversified data sources that provide a holistic and deep insight into how people move around, interact socially, and are connected through technology when living in a city.

4.3 Synthetic Data Generation

The creation of synthetic data in this research serves a crucial purpose: to validate and benchmark against a first-time, unpublished real-world MOBILE dataset from Nashik City, India. Given this dataset's novelty and unique characteristics, synthetic data is a vital tool for comparison and validation. This approach is especially important considering the absence of publicly available data to serve as a benchmark.

4.3.1 Generation Process of Cumulative Count Synthetic Data

The creation of synthetic data in this research serves a crucial purpose: validation and cross-checking using newly generated, primary MOBILE measurements acquired from a location in Nashik City, India, for the very time. Synthetic data is an important benchmark in assessing this new and distinct dataset. To that end, it is particularly significant because there are no appropriate data for comparison in general. The conceptual diagram figure 4-2 illustrates the stages of data generation.

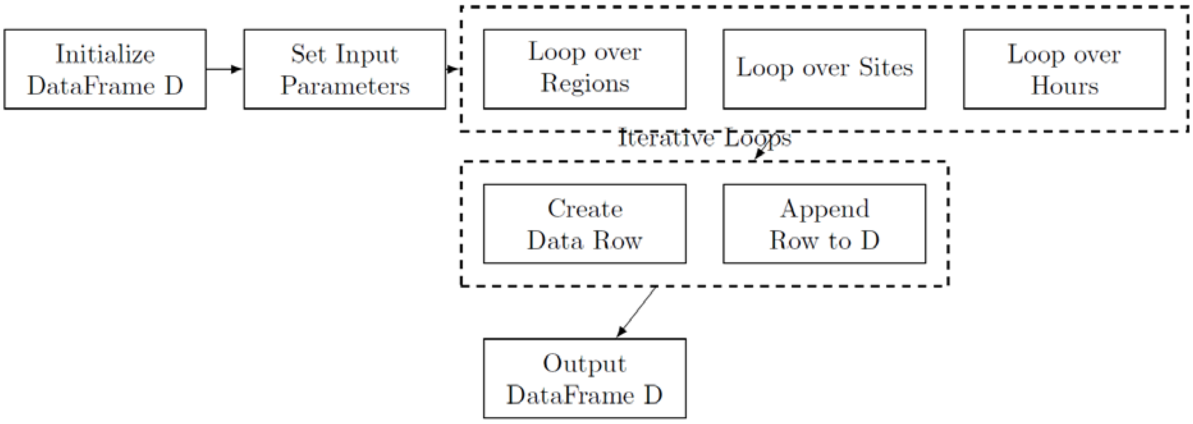


Figure 4-2: Conceptual Diagram of Synthetic Data for Cumulative Crowd Count

1. **Parameter Definition and Simulation Rules**:

Parameters such as the number of regions, sites per region, and the time intervals for data collection are defined, mirroring those in the real-world data. Simulation rules are established to ensure that the synthetic data replicates the real-world data's structure and behaviour while maintaining individual privacy.

2. **Data Generation**:

Various MOBILE tower sites in Nashik are used to generate data through Python scripts. As mentioned, the script generates data points including region, sub-region, Pincode, siteids, latitude, longitude, time, etc. These characteristics have been introduced in such a manner as to correspond with the actual-world table's entries for purposes of cross-comparison and analysis.

- **Data Characteristic Analysis:**

The initial analysis involves a deep dive into the real-world data to understand its key attributes.

Let D_{real} represent the real-world dataset. The attributes such as region (R), site ID (S), and time (T) are analyzed for their statistical properties like mean (μ), standard deviation (σ), and range ($[\min, \max]$).

- **Parameter and Rule Definition:**

Parameters are set based on the analysis of D_{real}. These include the number of regions. (N_r), sites per region (N_s), and the duration for data collection (N_t).

Define a function $F_{param}(N_r, N_s, N_t)$ That sets up the framework for synthetic data generation.

3 **Synthetic Data Generation Algorithm:**

Employing Python for implementation, the synthetic data generation is guided by a function G_{synth} that creates data points based on the defined parameters.

$$G_{synth}\left(F_{param}\right) \rightarrow D_{synth}$$

Eq: (4-1)

where D_{synth} is the synthetic dataset.

4 **Validation and Refinement:**

The synthetic dataset D_{synth} is statistically compared with D_{real} to validate its accuracy.

Refinement function $R_{fine}\left(D_{synth}, D_{real}\right)$ iteratively adjusts D_{synth} to enhance its congruence with D_{real}.

- **Nomenclature and Data Representation:**

The synthetic dataset, denoted as D_{synth} , comprises several attributes, each representing a key aspect of the MOBILE data presented in Table 4-1:

Table 4-1: Nomenclature

Regions (R): Represented as R_i Where i ranges from 1 to N_r, indicating different geographical areas.
Site IDs (S): Denoted as S_j For each site within a region, where j ranges from 1 to N_s.
Time (T): Time stamps are represented as T_k, ranging over N_t Intervals.
Latitude (λ) and Longitude (ϕ): Each site is assigned a geolocation coordinate. (λ_j, ϕ_j).
Reference Count (C_{ref}) and Current Count (C_{cur}) : Representing the baseline and observed connectivity at each site and time interval.

The structured tabular representation helps systematically organize the synthetic data, ensuring that each aspect of the data is clearly defined and accessible for analysis. Detailed Data presented in Table 4-2 Format of D_{synth}

Table 4-2: Synthetic Dataset Nomenclatures

Symbol	**Description**
R_i	Region identifier, where $i = 1,2,\dots,N_r$
S_j	Site ID within a region, where $j = 1,2,\dots,N_s$
T_k	Timestamp, where $k = 1,2,\dots,N_t$ (e.g., hourly intervals)
λ_j	Latitude coordinate of site S_j
ϕ_j	Longitude coordination of site S_j
$C_{ref}(R_i, S_j, T_k)$	Reference count at region R_i, site S_j, and time T_k
$C_{cur}(R_i, S_j, T_k)$	The current count at the region R_i, site S_j, and time T_k

Each row in the table 4-3 represents a specific element or attribute in the synthetic dataset, described using mathematical symbols.

Table 4-3: Synthetic Dataset Overview

Region	**Site ID**	**Time**	**Latitude**	**Longitude**	**Reference Count**	**Current Count**
R	S	T	λ	ϕ	C_{ref}	C_{cur}
R_1	S_1	T_1	λ_1	ϕ_1	$C_{ref}(R_1, S_1, T_1)$	$C_{cur}(R_1, S_1, T_1)$
R_1	S_2	T_2	λ_2	ϕ_2	$C_{ref}(R_1, S_2, T_2)$	$C_{cur}(R_1, S_2, T_2)$
…	…	…	…	…	…	…

This nomenclature and data representation format provides a structured and mathematically grounded approach to synthetic data generation in the context of MOBILE network analysis.

- **Synthetic Cumulative Crowd Data Algorithm**

Creating synthetic data of MOBILE tower logs is a fundamental prerequisite for various analytical and modelling purposes. The following Algorithm 4-1 presents a methodical approach to generating such data, encapsulating a range of attributes from geographical coordinates to network connectivity details. This algorithmic design is crucial for simulating realistic urban telecommunications scenarios in a controlled, privacy-compliant manner.

Algorithm 4-1: Generation of Synthetic MOBILE Tower Log Data

Input Parameters:
$N_{regions}$: Number of regions, denoted as num_regions in the code.
N_{sites} : Number of sites (MOBILE towers) per region, denoted as num_sites_per_region.
N_{hours} : Number of hours to simulate data for, denoted as num_hours.
P_{start} : The starting pincode for the regions is denoted as start_pincode.

Output:
A DataFrame D containing the synthetic MOBILE tower log data.

Algorithm Steps:
Initialization:

- Create an empty DataFrame D with columns:

```
'Region,' 'SubRegion,' 'Pincode,' 'SiteIDs,' 'Latitude,' 'Longitude,' 'Time,' 'Ref_Count',
'Current_Count'

Data Generation:
For each region r in {1,2, ... , N_regions } :
   For each site s in {1,2, ... , N_sites } :
      For each hour h in {0,1, ... , N_hours − 1} :
           Create a data row with the following values:
            'Region' = 'Region ' +r
            'SubRegion' = 'SubRegion' +s
             'Pincode' = P_start + r
             'SiteIDs' = 'SiteID' +s
             'Latitude' = Random value within [19.5,20.5]
             'Longitude' = Random value within [73.5,74.5]
             'Time' = Formatted hour h
             'Ref_Count' = Random integer within [1000,10000]
             'Current_Count' = Random integer within [1000,10000]
Append row to DataFrame D.
Return:
Return the DataFrame D containing the generated synthetic data.

End of Algorithm
```

The algorithm discussed creates proper, reliable, accurate MOBILE tower log data. It is an effective tool for researchers in simulating diverse scenarios across different regions and varying times to analyse MOBILE network behaviour. This algorithm is very flexible and profound; hence, it provides a good basis for research on urban planning, network optimization, and telecom strategy conforming to privacy law, respecting privacy and user rights.

4.3.2 Generation of Individual Mobility Scenario

A synthetic dataset is provided under the reference. D_{synth}, which replicates an individual's movement over various MOBILE towers over 24 hours. In this process records for numerous people who visit different MOBILE towers during their various time intervals in the day. The simulation represents urban street mobility and links to MOBILE towers.

The conceptual flow diagram figure 4-4 is presents a data processing algorithm with simulation models. First, the parameters are initialized and then a data list is created. After that, places and areas are determined resulting in a process of repetition for data input. This loop involves picking up one site, creation of new data record, before updating an existing site. This appends the new data to the present list. Once the loop has been concluded, the collected data is compiled in the form of a DataFrame that serves to be an output of the entire process. This is a stepwise protocol of how one may simulate/track data across different spaces and times possibly adopted in computational model for diverse applications including crowd simulation and network traffic analysis.

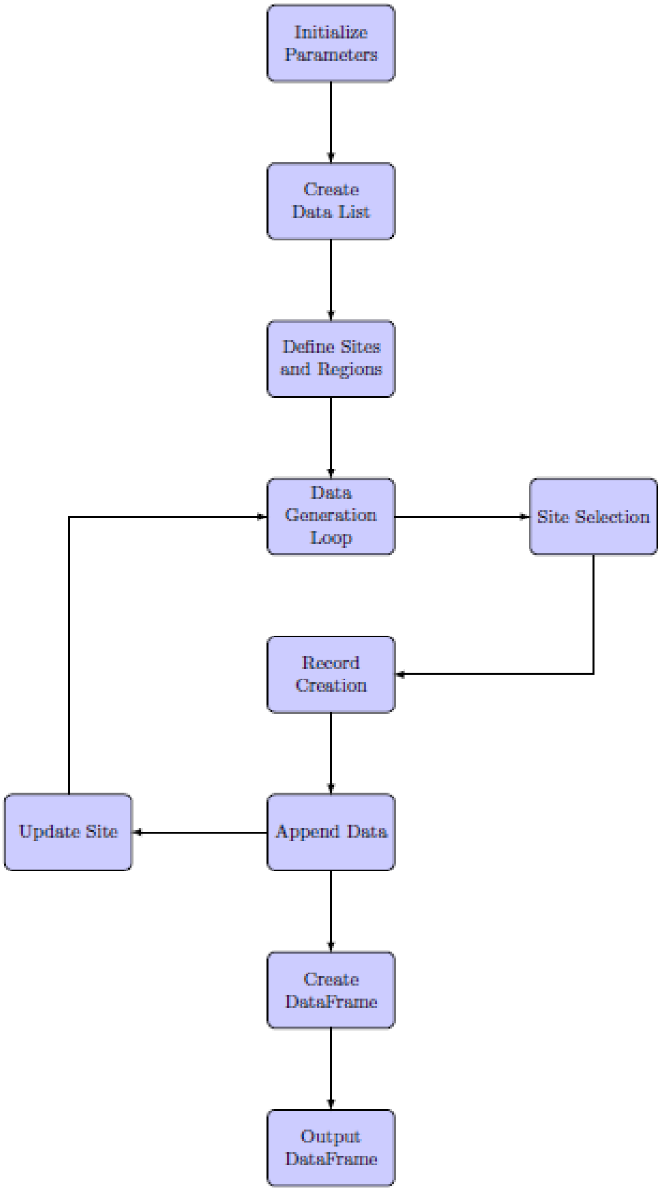


Figure 4-3: Synthetic Individual Mobility Scenario Dataset

A. Mathematical Formulation:

Individuals:

Let $I = \{i_1, i_2, \ldots, i_n\}$ Represent the set of individuals, where n is the total number of individuals in the simulation.

MOBILE Towers (Site IDs):

Let $S = \{s_1, s_2, \ldots, s_m\}$ Denote the set of MOBILE towers, where m is the total number of sites.

Time Intervals:

Time is discretized into hourly intervals, $T = \{t_1, t_2, \ldots, t_{24}\}$, representing each hour of the day.

Location Coordinates:

Each MOBILE tower $s_j \in S$ is associated with a geographic coordinate, represented as (lat_i, lon_i).

Data Record:

Each data record for an individual i at time t is denoted as:

$$D(i, t) = \{R, SubR, P, S_i, \text{lat } (\lambda_i), \text{lon } (\phi_i), t, C_{ref}, D_{next}, TS, TD, V, V_{avg}, T_{avg}, ST, TT, ID\}$$

Eq: (4-2)

The table 4-4 present the sample of dataset with nomenclatures and example values for an overview of individual mobility data from Mobile Towers.

Table 4-4: Nomenclature Presentation of Synthetic Dataset

Dataset Attribute	**Symbolic Representation**	**Example Symbolic Values**
Region	R_i	$R_1, R_2, \ldots$
SubRegion	$SubR_i$	$SubR_1, SubR_2, \ldots$
Pincode	P	$P_1, P_2, \ldots$
Site IDs	S_i	$S_1, S_2, \ldots$
Latitude	λ_i	$\lambda_1, \lambda_2, \ldots$
Longitude	ϕ_i	$\phi_1, \phi_2, \ldots$
Time	T_i	$T_1, T_2, \ldots$
Reference Count	C_{ref_i}	$C_{ref_1}, C_{ref_2}, \ldots$
Destination	D_{next_i}	$D_{next_1}, D_{next_2}, \ldots$
Travel Spots	TS_i	$TS_1, TS_2, \ldots$
Total Distance	TD_i	$TD_1, TD_2, \ldots$

Vehicle	V_i	$TT_1, TT_2, \ldots$
Average Velocity	V_{avg_i}	$V_1, V_2, \ldots$
Average Travel Time	T_{avg_i}	$V_{avg_1}, V_{avg_2}, \ldots$
Start Time	ST_i	$T_{avg_1}, T_{avg_2}, \ldots$
Traverse Time	TT_i	$ST_1, ST_2, \ldots$
ID	ID_i	$T_2, \ldots$

B. Synthetic Individual Mobility Data Generation Algorithm

The need to correctly estimate the movements and contacts of individuals within the city context is essential of mobility studies and urban mobility. A method for constructing a complete synthetic dataset using different events at different MOBILE sites, referring to the general situation with an ordinary street movement, is presented below in Algorithm 4-2.

Algorithm 4-2: Generation of Detailed Synthetic Data for Individual Movements

Input Parameters:
N_{ind} : Number of individuals, represented as num_individuals.
N_{sites} : Number of MOBILE tower sites, num_sites.
N_{hours} : Number of hours for simulation, num_hours.
P_{start} : Starting pincode, start_pincode.

Output:
A DataFrame D containing detailed synthetic data of individual movements.

Algorithm Steps:

Initialization:
Create an empty list of data for storing data records.
Define site IDs $S = \{s_1, s_2, \ldots, s_{N_{sites}}\}$.
Define regions $R = \{r_1, r_2, \ldots, r_{N_{sites}}\}$ And subregions SubR.

Data Generation Loop:
For each individual i in the set $\{1,2, \ldots, N_{ind}\}$:

Initial Site Selection:
Set the initial site $s_{current}$ by randomly selecting from the set of sites $S = \{s_1, s_2, \ldots, s_{N_{sites}}\}$.

Hourly Iteration:
For each hour h in the set $\{0,1, \ldots, N_{hours} - 1\}$:

Next Site Selection:
Select the next site s_{next} randomly from S, ensuring $s_{next} \neq s_{current}$.

Record Creation:

Construct a record row with the following attributes:
Region (R): Randomly chosen from the predefined set of regions.
SubRegion (SubR): Randomly chosen corresponding to R.
Pincode (P): Derived as P_{start} + region index.
Current Site (S_{cur}) : Set as $s_{current}$.
Latitude (λ): Generated randomly within a specified range.
Longitude (ϕ): Generated randomly within a specified range.
Time (T): Set as the current hour h in the format 'HH:MM: SS.'
Reference Count (C_{ref}) : Random integer within a specified range.
Destination (D_{next}) : Set as s_{next} .
Travel Spots (TS): Random integer within a specified range.
Total Distance (TD): Random integer within a specified range.
Vehicle (V): Randomly chosen from a predefined set of vehicle types.
Average Velocity (V_{avg}) : Random integer within a specified range.
Travel Time (T_{avg}) : Computed based on TD and V_{avg} .
Start Time (ST): Randomly chosen from the set of all possible start times.
Traverse Time (TT): Computed based on T_{avg} And ST.
ID (ID): Unique identifier constructed for each row.

Data Appending:
Append row to the data collection data.
Site Update:
Update $s_{current}$ to s_{next} for the next iteration.
End of Loop

Create DataFrame:
Convert data list to a DataFrame D.
Return:
Return the DataFrame D.
End of Algorithm

The algorithm was carefully developed to manufacture an artificial dataset imitating city movement, MOBILE antennas, etc. Its role is crucial in urban studies and telecommunications research, providing data on individuals' mobility and allowing diverse studies. The depth and flexibility make the algorithm applicable to many studies, including those that involve virtual datasets lacking in the realness of specific data.

C. Validation and Refinement

To confirm that the synthesis data can imitate crucial structure and distribution, the real-world dataset is rigorously compared with it. Synthetic data is adjusted as needed to conform with the spatiotemporal characteristics of actual data.

4.3.3 Utility of Synthetic Data in Crowd Density Estimation

- Benchmarking and Validation: A synthetic dataset acts as a reference point, which allows for analysis and comparison of the credibility of real-world data. It is also crucial to note that in the real world, the data provided has never been previously published or externally validated.
- Model Testing and Simulation: These models can be tested using the synthetic dataset before being applied to the real data. This stage enables one to identify possible challenges that may be evident when estimating more refined crowd density estimates.
- Risk Mitigation: Synthetics can handle fewer risks in the initial stage of the model-building process when proper handling protocols have not yet been defined.
- Scenario Analysis: This synthetic data also helps create and understand scenarios that would be impossible to get using actual data to enrich our understanding of crowd dynamics in other situations.

This research thus confirms that the synthetic data created using the same Nashik MOBILE environment as the original data has shown that the analysis is more robust because it will have the power to detect different crowd density trends and movement patterns in the city's environment.

4.4 MOBILE Data Collection

Using real-life MOBILE (Global System for Mobile Communications) is crucial in a holistic investigation into the telecommunication dynamics in urban areas. The third part of this chapter explores how to obtain direct MOBILE data by extracting it from tower logs. Briefly, this data is nothing but a little picture representing a wider dataset that gives an irreplaceable understanding of how urban citizens communicate and how they load the network and migrate from one point to another. The validity and precision of this information are unparalleled in any world-class study on telecommunications.

Detailed MOBILE data from the tower logs covering mobile communication activity in different city areas. The dataset provides comprehensive details, such as dates, site IDs (for different MOBILE towers), signal strength indicators, and connectivity counts. Key aspects of the data collection include:

The Figure 4-4 shows the MOBILE network, the location of the towers, and data flow is crucial. The diagram will visually represent the data collection sources, demonstrating a visual link between the raw data and reality. The scheme below illustrates an intricate communications network for evaluating population density using mobile- phone data analysis. The framework is called a 'Telecommunications network' and comprises several appropriately located MOBILE towers, which are provided with stations and controllers. Icons serve as symbols for these towers and correspond to selected data points representing a human population within a specific locational area.

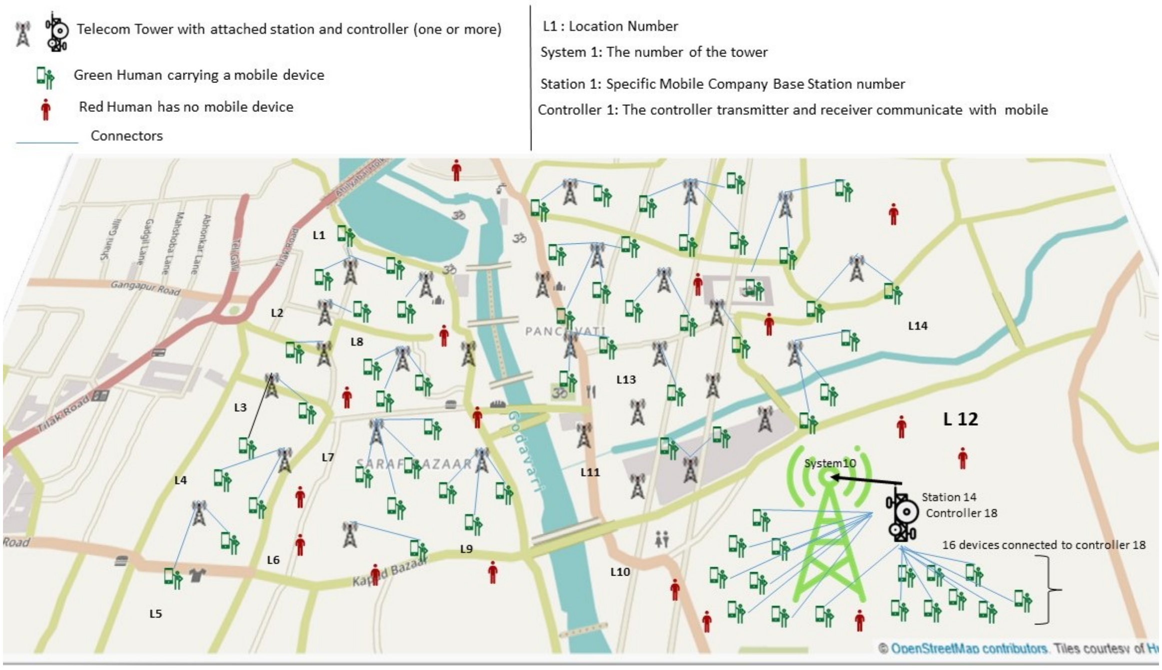


Figure 4-4: MOBILE Data Collection Conceptual Diagram

The green man indicates that the person within the network's coverage area carries a mobile phone, and the red man means he does not have one. The green figures represent the possible data sources from which they can collect data. For example, they can gather information from mobile phones that are usually carried around by a person and actively connect to the MOBILE towers. Red human icons illustrate the figures of persons who do not carry mobile phones

Key components of the diagram are labelled for clarity:

- L1: They refer to particular locations for this network’s grids.
- System 1 simply means how many MOBILE towers each network can host.
- Station 1: It has the base station identification number that is typical for the MNO network.
- Controller 1: They represent the controllers connected to each MOBILE tower, whose primary function is sending and receiving messages through the mobile phones.

The complicated design for this diagram indicates the relationship of each Mobile tower and network from where data is sent from a single mobile to the core system controller. The main function of this controller is to gather information regarding connections made by every tower for use in measuring crowd density indicators within the stipulated area under review. It includes GPS, a feature in every base station, making data local and time-specific. GPS-based crowd movement data and call detail records provide much insight into how crowds move and their density.

By showing urban density analysis of telecommunication network that utilizes MOBILE technology as a basis of this study and generates useful results. With such a network, researchers can quantitatively measure the crowd's density and identify mobility trends and strategies suitable for urban planning and management. The compact diagram summarizes the network's technical complexities, showing how data flows and analytical procedures for urban telecommunication studies.

4.4.1 MOBILE Data Details

Data on communication patterns was collected from a wide range of urban samples at certain points in time, thus allowing for a temporal perspective for the dataset. The MOBILE dataset presented in table 4-5 provides an overall picture covering many locations that show what is happening in the networks and users. Markets The dataset taken from the actual MOBILE tower log is essential to comprehend the movements in mobile communication in certain cities and regions. The list of attributes that come with its definition is vital to telecoms research, especially in the design and implementation of optimal networks and in modelling customer behaviour.

Table 4-5: Mathematical Presentation of MOBILE Dataset

Column Header	Symbolic Notation	Example Data
Local Area Code (LAC)	L	$L_1 - 156, L_2 - 156$
Telecom Provider	T	T_1 − "AIRCEL, "T_2 − "AIRCEL"
Site Identifier	S_{ID}	S_{ID1} − "ACTV-MAHNASNAS0109", S_{ID2} − "ACTV-MAHNASNAS 0150"
Site Name	S_{Name}	S_{Name1} − " JATRA", S_{Name2} − " ADGAON"
Latitude	λ	$\lambda_1 - 20.0262, \lambda_2 - 20.0335$
Longitude	ϕ	$\phi_1 - 73.8424, \phi_2 - 73.8637$
Region	R	R_1 − "REGION8, R_2 − "REGION9"
Taluka	Ta	Ta_1 − " NASIK" Ta_2 − " NIPHAD"
Sub Route	SR	SR_1 represents "ADGAON_RASBIHARI" SR_2 also represents "ADGAON_RASBIHARI"
Highway Route	HR	HR_1 represents "DHULE_ROUTE" HR_2 also represents "DHULE_ROUTE."
Reference Count	C_{ref}	$C_{ref1} - 699, C_{ref2} - 822$
Start Time	ST	$ST_1 - 1442167200, ST_2 - 1442167200$
Date and Time	DT	$DT_1 - {}^{n}13 - 09 - 201523{:}30", DT_2 - {}^{1}13 - 09 - 201523{:}30"$
Current Count	C_{cur}	$C_{ctu1} - 477, C_{cur2} - 797$

Mathematical nomenclature is used here to portray this elaborate description of the MOBILE dataset to illustrate how accurate real-world telecommunications datasets can be used in scholarly activities. This organization of the datasets into spatial, chronological, and network-specific elements greatly benefits the theory builders dealing with mobile communications and urban studies.

Data Integrity and Privacy: Ethical research dictates that data quality should be maintained while users' privacy and data integrity are also assured. Hence, the research is carried out with synthetic data, and the real-world MOBILE data from India is taken in a cumulative format and not at an individual level.

Collecting MOBILE data from tower logs becomes critical in disclosing the complex aspects of urban telecommunication. This data set will serve as an analytical base and offer opportunities for predictive modelling and network optimization. Integrating it into a research thesis for telecommunication and urban studies gives credibility to the research by making it applicable to real-world conditions and enhances its standards.

4.5 Data Preprocessing Techniques

4.5.1 Data Distribution Assessment:

A statistical exploration is conducted to understand the nature of the dataset. This can be mathematically represented by examining the M-o-Ments of the dataset:

Mean

$(\mu): \mu = \frac{1}{N}\sum_{i=1}^{N} x_i$

Eq: (4-3)

Variance

$$(\sigma^2): \sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$$

Eq: (4-4)

Skewness and Kurtosis are also considered to understand the asymmetry and tailedness of the distribution.

4.5.2 Spatio-Temporal Analysis

Outliers in space and time are identified using methods that consider both spatial location and temporal occurrence:

- Spatial Analysis: Identify spatial outliers by calculating the spatial median and interquartile range (IQR) for latitude (λ) and longitude (ϕ) coordinates.
- Temporal Analysis: Detect temporal outliers by assessing the time series data points (T) for significant deviations from a moving average or other temporal trendlines.

4.5.3 Outlier Identification

Outliers are identified based on their deviation from the expected distribution without normalizing the data, as normalization could distort the real-time context. Mathematically, outliers can be detected when a data point x satisfies the condition: $|x - \mu| > k \cdot \sigma$, where k is a threshold value based on the desired sensitivity of the outlier detection.

4.5.4 Non-Normalization Justification

The proposed algorithm operates on data without normalization. (x_{raw}) To preserve the real-time nature:

x_{raw} : The raw data point is collected from the source.
x_{norm} : This research does not use A normalized data point to avoid loss of real-time sensitivity.

4.5.5 Algorithmic Adaptation to Raw Data

The algorithm is mathematically adapted to handle raw data, ensuring that decisions are made based on the most current and contextually relevant information:
The decision function $D(x_{raw})$ is thus designed to operate on x_{raw} instead of x_{norm}.

The rationale is that the research emphasizes the raw state of data, which preserves the authenticity of real-time responsiveness. The mathematical formulation behind every preprocessing stage demonstrates a commitment to empirical precision and is tuned to the data's real-time dynamics. This approach reinforces the ability of the research to provide meaningful and directive insights in a reactive environment.

4.6 Evolution of Threshold Algorithm

4.6.1 Threshold Algorithm 1: Threshold Computation

The Threshold Algorithm 4-3 proposed aims to understand crowd flows within a city during the day, taking into account data obtained from the MOBILE towers. Dynamic calculation of hourly thresholds based on counties explains city population shifts over time. There is a need to develop the infrastructure of data generation. Therefore, it entails setting up a series of MOBILE towers in the town that shall supply cumulative count data at certain periods. ‘n’ denotes the total number of locations or MOBILE towers, while ‘Ccount(n)’ refers to the crowd counts. The algorithm continuously stores the total count from any MOBILE tower every hour. The threshold is calculated using these counts to determine which should be used for that particular hour. To compute the threshold for a specific hour ‘h,’ the cumulative crowd count in other hours starting from 1st hour to h^n determined, and the median of this calculation will be considered for h. This process is repeated every hour until there are 24 thresholds, and everyone corresponds to the median of all crowds counted so far.

A. Diagrammatic Presentation

The algorithm’s workflow figure 4-5 involves initialization, which includes setting up the number of MOBILE towers and data structure to store the accumulated crowds. The 'Data Collection' phase is a more complex state involving a series of nested steps: go through every hour of the day and proceed with each MOBILE tower to make updates and store the crowd counts. First, this step entails important information that is essential for more analysis. The algorithm shifts to its “calculate threshold” state upon data gathering. It then calculates the medians of cumulative crowd counts for every hour, thus setting thresholds. It is an important step that helps to comprehend variations in crowd density at different times. Finally, the algorithm arrives at the ‘final output’ phase, upon which it proceeds to produce a 24-element array of thresholds, one value per hour. The result of the algorithm is that this array hints at an hour’s crowd flow at MOBILE towers. The directionality of the diagram is vertical in alignment with the algorithm's sequential nature, making it easier to comprehend the processes from initialization to final output.

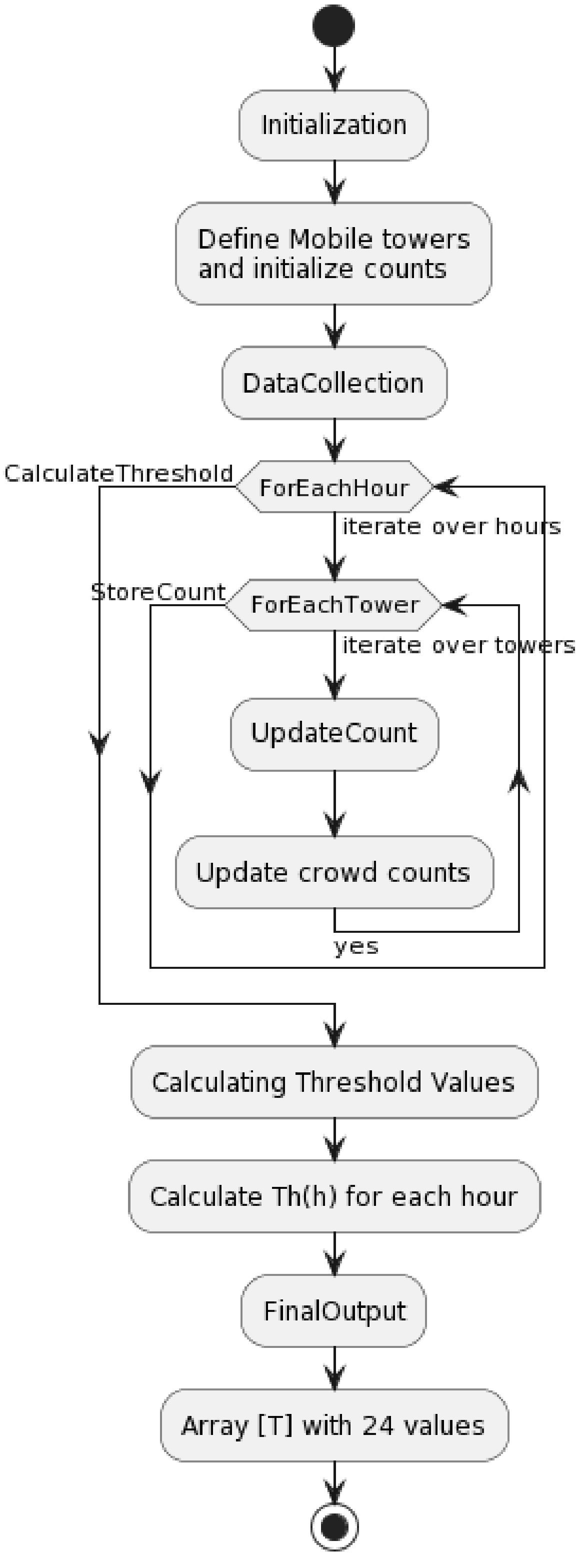


Figure 4-5: Workflow Diagram Threshold Algorithm

B. Mathematical Representation

The mathematical formulation of the algorithm can be expressed as follows:

Let $C = [C_1, C_2, \ldots, C_n]$ Represent the array of cumulative crowd counts, where C_i Is the count at hour i, and n is the total number of hours (24).

The threshold value for hour h is calculated using the median function:

$$T(h) = \text{median}\ (C[1], C[2], \ldots, C[h])$$

Eq: (4-5)

where h ranges from 1 to 24.

The Algorithm 4-3 follows up on MOBILE tower crowding within 24 hours. It first instantiates data structures for holding cumulative counts by a tower, subsequently records hourly data, computes median-based threshold values, produces arrays with thresholds hour-wise, and hence gives an image of hourly crowd behaviour patterns.

Algorithm 4-3: Threshold Computation

Step 1: Initialization

Let n be the total number of MOBILE tower locations.
Define MobileTowers (n) to represent the n MOBILE towers.
Initialize $\text{Ccount} = [0,0, \ldots, 0]$ as an array of length n to store cumulative crowd counts from each tower.
Define T as an array of length 24 to store threshold values for each hour.

Step 2: Data Collection for Each Hour

For each hour h from 1 to 24 :
    For each tower i from 1 to n :
        Update Ccount [i] with the current crowd count from MOBILE tower i at hour h.
        Store the updated Ccount values.

Step 3: Calculating the Threshold Values

For each hour h from 1 to 24 :
    Let C_h be a subset of Ccount containing crowd counts from hour 1 to hour h.
    Calculate $T(h)$ as:

$$T(h) = \text{median}\ (C_h)$$

    Store $T(h)$ in the array T.

Step 4: Final Output

    Array [T] will have a total of 24 threshold values

Algorithm 4-3 establishes an hourly threshold value set derived from the accumulated crowd counts fetched from the MOBILE towers. The threshold of crowd density is evaluated every hour on these 24 points, and the average of all is considered the median value. It is the groundwork for the larger picture presented in algorithm 4, which goes into a detailed discussion on crowd patterns. Building on Algorithm 3; transitioning to Algorithm 4. Algorithm 3 was crucial in highlighting hourly crowd dynamics, but Algorithm 4 will go a step further as it seeks to incorporate daily and weekly variations. Using a 2D array C_" count" [d][h]" helps us appreciate crowd counts variations on an hourly basis as well as a daily level basis. In this regard, computing both "T_{daily}" [d][h] and "T_{weekly}" [h] thresholds enable a multilayered approach to crowd dynamics analysis and response on multiple time scales. Such an integrative strategy makes the tracking, predicting, and response system of crowd flows within cities more rigorous.

4.6.2 Threshold Algorithm 2: Daily & Weekly Threshold Array

The algorithm for computation of daily or weekly threshold and data from the MOBILE tower is used by algorithms. The updated algorithm for daily or weekly threshold calculation analyses city crowd activities. To determine daily or weekly thresholds, the algorithm "The Updated Algorithm for Daily or Week " is structured to compute daily and weekly average thresholds and hourly trends. This algorithm determines dynamic median-based thresholds every hour that provide the basis for understanding movement's spatial distribution with temporal development.

A. Mathematical Representation

Daily Threshold Calculation

$$T_{daily}\,[d][h] = \text{median}\,(C_{count}\,[d][1], C_{count}\,[d][2], \dots, C_{count}\,[d][h])$$
$$\text{where } d = 1,2, \dots ,7 \text{ and } h = 1,2, \dots ,24.$$

Eq: (4-6)

Weekly Threshold Calculation

$$T_{weekly}\,[h] = \text{median}\,(C_{count}\,[1][h], C_{count}\,[2][h], \dots, C_{count}\,[7][h])$$

Eq: (4-7)

where $h = 1,2, \dots ,24$.

Example Data Presentation of expected outcome of the algorithm computation is presented in table 4-6 for 24 hours and table 4-7 for weekly thresholds.

Table 4-6: Daily Threshold Values Preview

Hour	T_{daily} (1h)	T_{daily} (2h)	T_{daily} (3h)	...	T_{daily} (24h)
1	25	-	-	...	-
2	-	30	-	...	-
3	-	-	35	...	-
...	...	...	...	...	...
24	-	-	-	...	55

Table 4-7: Weekly Threshold Values Preview

Hour	T_{weekly} (1h)	T_{weekly} (2h)	T_{weekly} (3h)	...	T_{weekly} (24h)
1	26	-	-	...	-
2	-	31	-	...	-
3	-	-	36	...	-
...	...	...	...	...	...
24	-	-	-	...	57

B. Diagrammatic Presentation

The “Algorithm for Daily and Weekly Threshold Calculations” with flow diagram in figure 4-6 is a powerful technology that studies city crowds with information from MOBILE towers. Cumulative crowd counts are summarized into appropriate hourly thresholds for every day cumulatively per week. The algorithm 4-4 is unique because it produces highly detailed data daily and larger-scale weekly trends, which are crucial aspects of urban planning, distributing resources accordingly, and revealing patterns in population movements. Such thorough analysis is necessary in modern urban settings due to the dual-layered approach it employs to study crowd dynamics.

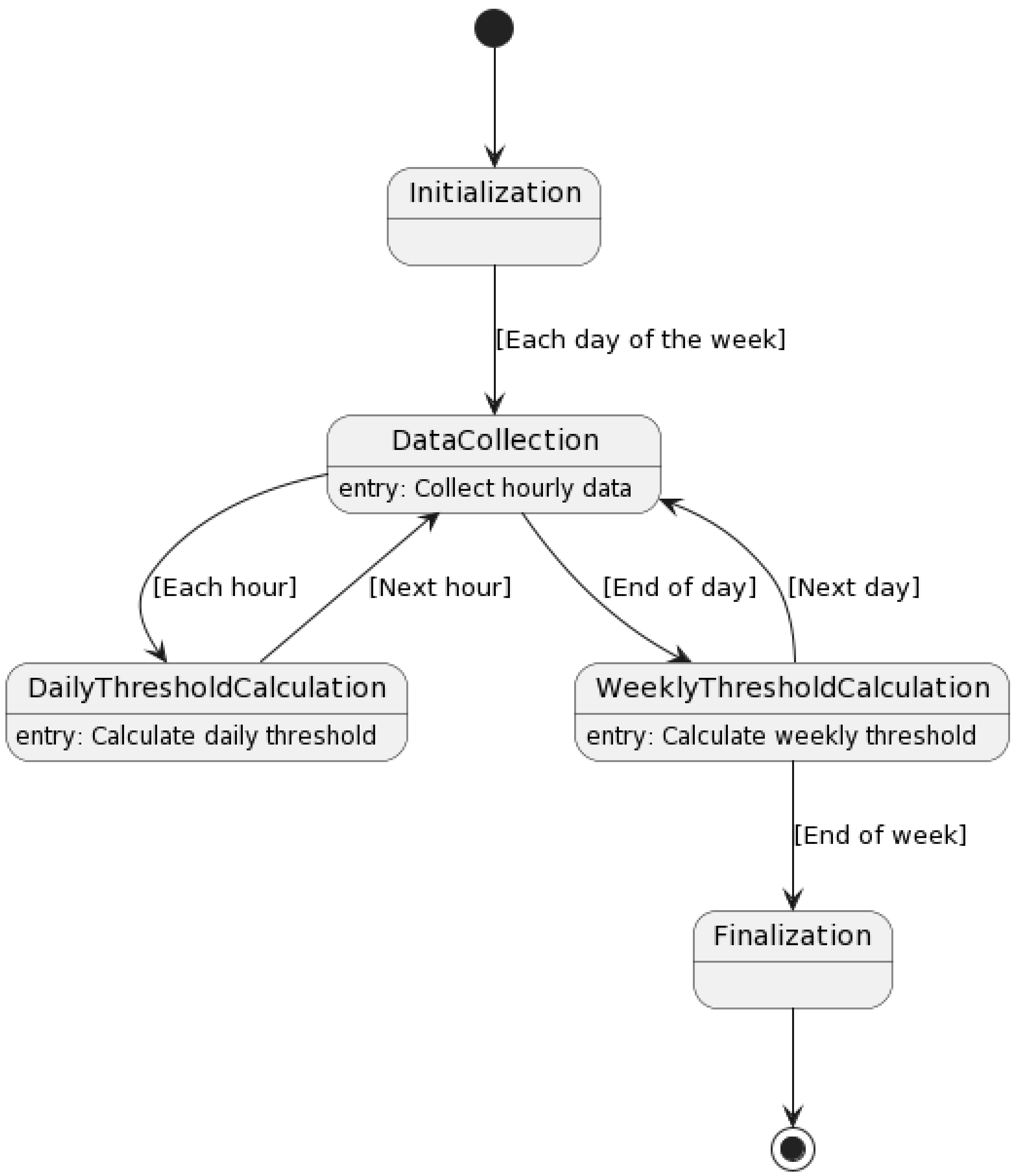


Figure 4-6: Flow Diagram for Daily and weekly Threshold Computation

A systematic representation of the algorithm 4-4 gives full insight into how it will analyse crowd behaviour daily and over one week.

Algorithm 4-4: Threshold Computation For Daily & Weekly

Step 1: Initialization
Let n be the total number of MOBILE tower locations.
Define MobileTowers (n) to represent the n MOBILE towers.
Initialize a 2D array. $C_{count}[d][h]$ to store cumulative crowd counts, where d is the day index (1 to 7 for a week), and h is the hour index (1 to 24 for a day).
Define a 2D array. $T_{daily}[d][h]$ to store daily threshold values.
Define an array $T_{weekly}[h]$ to store weekly threshold values.

Step 2: Data Collection for Each Hour and Each Day
For each day d in $\{1,2,\ldots,7\}$:
For each hour h in $\{1,2,\ldots,24\}$:
Collect $C_{count}[d][h]$, the cumulative crowd count at each MOBILE tower for the current hour and day.

Step 3: Calculating Daily Threshold Values
For each day d in $\{1,2,\ldots,7\}$:
For each hour h in $\{1,2,\ldots,24\}$:
Calculate the threshold for the current day and hour as:

$$T_{daily}[d][h] = \text{median}\,(C_{count}[d][1], C_{count}[d][2], \ldots, C_{count}[d][h])$$

Step 4: Calculating Weekly Threshold Values
For each hour h in $\{1,2,\ldots,24\}$:
Create an array weekly_counts [h] to store the C_{count} for hour h across all days of the week.
For each day d in $\{1,2,\ldots,7\}$:
Append $C_{count}[d][h]$ to weekly_counts [h].
Calculate the weekly threshold for hour h as:

$$T_{weekly}[h] = \text{median}\,(\text{weekly\_counts}\,[h])$$

Step 5: Storing and Updating Threshold Values
Store each calculated $T_{daily}[d][h]$ in the 2D array T_{daily} .
Store each calculated $T_{weekly}[h]$ in the array T_{weekly} .

At the end of a week. T_{daily}, the daily threshold levels for every hour will be included in this file revealing detailed information on daily crowd movements. T_{weekly}. Weekly median hour-by-hour thresholds for a weekly view of the weekly pattern. The algorithm has a step-by-step mathematical structure that allows it to be presented as a thorough tool used to analyse a crowd's behaviour daily or weekly.

Algorithm 4-4 offers a thorough framework for studying crowd behaviour within daily and weekly periods. The method for determining these threshold values, "T_{daily}" [d][h] and "T_{weekly}" [h], has a refined procedure for dealing with the details of temporal crowd patterns in the MOBILE tower location. Its power is based on distinguishing different crowd densities, creating good prerequisites for more complicated studies. This leads to the next algorithm (Algorithm 4-5) that adds a ‘Median-of-Medians’ statistical layer to improve our comprehension of crowd movements. To further enhance the analysis, this technique examines the average of the crowd's daily norms and the median of such averages over multiple days. It gives a more uniform and reliable threshold, which factors in daily fluctuations but removes outliers. The evolution of such a strategy in analytical technique for a better crowd behaviour monitoring and prediction system is complex urban environments.

4.6.3 Threshold Algorithm 3: Median-of-Median Threshold Computation

The algorithm 4-5 has been proposed to determine the hour-by-hour medians and one combined median-of-median for all 24 hours. It then computes a median threshold from the cumulative hourly crowds-counts the flow diagram is presented in figure 4-7. It gradually calculates the median of each value up to an hour, producing an evolving image of the threshold changes over the day. The example representation is shown in table 4-8 for a better understanding of expected threshold modelling. This is the most important aspect of this research as the focus of the research is to present a stable method of crowd density estimation with logs and spatiotemporal meta data from Mobile Tower. Median of Median approach is expected to maintain consistence in the threshold values and when classifications are applied not location will be left out as outlier. Let’s say at 00:00 hours some x location is in outlier but the same location at 05:00 hours may not be in outliers. In order to keep these unforeseen situations a steady approach is proposed.

Table 4-8: Median-of-Median Threshold Computation

Hour	Hourly Median (C_{count})	Median	Median-of-Median
$T_{hourly}(1h)$	T_{1h}	$T_{m\_1} = \{T_{1h}\}$	$T_{M-o-M\_1} = \{T_{1h}\}$
$T_{hourly}(2h)$	T_{2h}	$T_{m\_2} = \{T_{1h}, T_{2h}\}$	$T_{M-o-M\_2} = \{T_{1h}, T_{m\_2}\}$
$T_{hourly}(3h)$	T_{3h}	$T_{m\_3} = \{T_{1h}, T_{2h}, T_{3h}\}$	$T_{M-o-M\_3} = \{T_{1h}, T_{m\_2}, T_{m\_3}\}$
$T_{hourly}(4h)$	T_{4h}	$T_{m\_4} = \{T_{1h}, T_{2h}, T_{3h}, T_{4h}\}$	$T_{M-o-M\_4} = \{T_{1h}, T_{m\_2}, T_{m\_3}, T_{m\_4}\}$
…	…	…	…
$T_{hourly}(24h)$	T_{24h}	$T_{m\_24} = \{T_{1h}, T_{2h}, T_{3h}, T_{4h} \dots T_{24}\}$	$T_{M-o-M\_24} = \{T_{1h}, T_{m\_2}, T_{m\_3}, T_{m\_4} \dots, T_{24}\}$

A. Mathematical Representation

1 Daily Median Thresholds:

For each day d (where $d = 1,2,\dots,7$), the daily median threshold, denoted as T_{m_d}, is calculated from the hourly median values of that day.

The expression for the daily median threshold on day d is:

$$T_{m_d} = \text{median}\left(\left\{T_{hourly\ 1}, T_{hourly\ 2}, \dots, T_{hourly\ 24}\right\}\right)$$

Eq: (4-8)

Here, $T_{hourly\ h}$ Represents the median crowd count at hour h of day d.

2 Median-of-Median Thresholds (Weekly):

The Median-of-Median (M-o-M) threshold for each day d is the median of all daily median thresholds up to that day.

The expression for the M-o-M threshold on day d is:

$$T_{M-o-M_d} = \text{median}\left(\left\{T_{m_1}, T_{m_2}, \dots, T_{m_d}\right\}\right)$$

Eq: (4-9)

The median-of-median (M-o-M) is a vital instrument for urban analysis. Daily and weekly dynamics within the crowd are unveiled by refining this hourly data derived from one of thousands of MOBILE towers. These snapshots are then integrated into a week that presents vital information for urban planning and research.

B. Diagrammatic Presentation

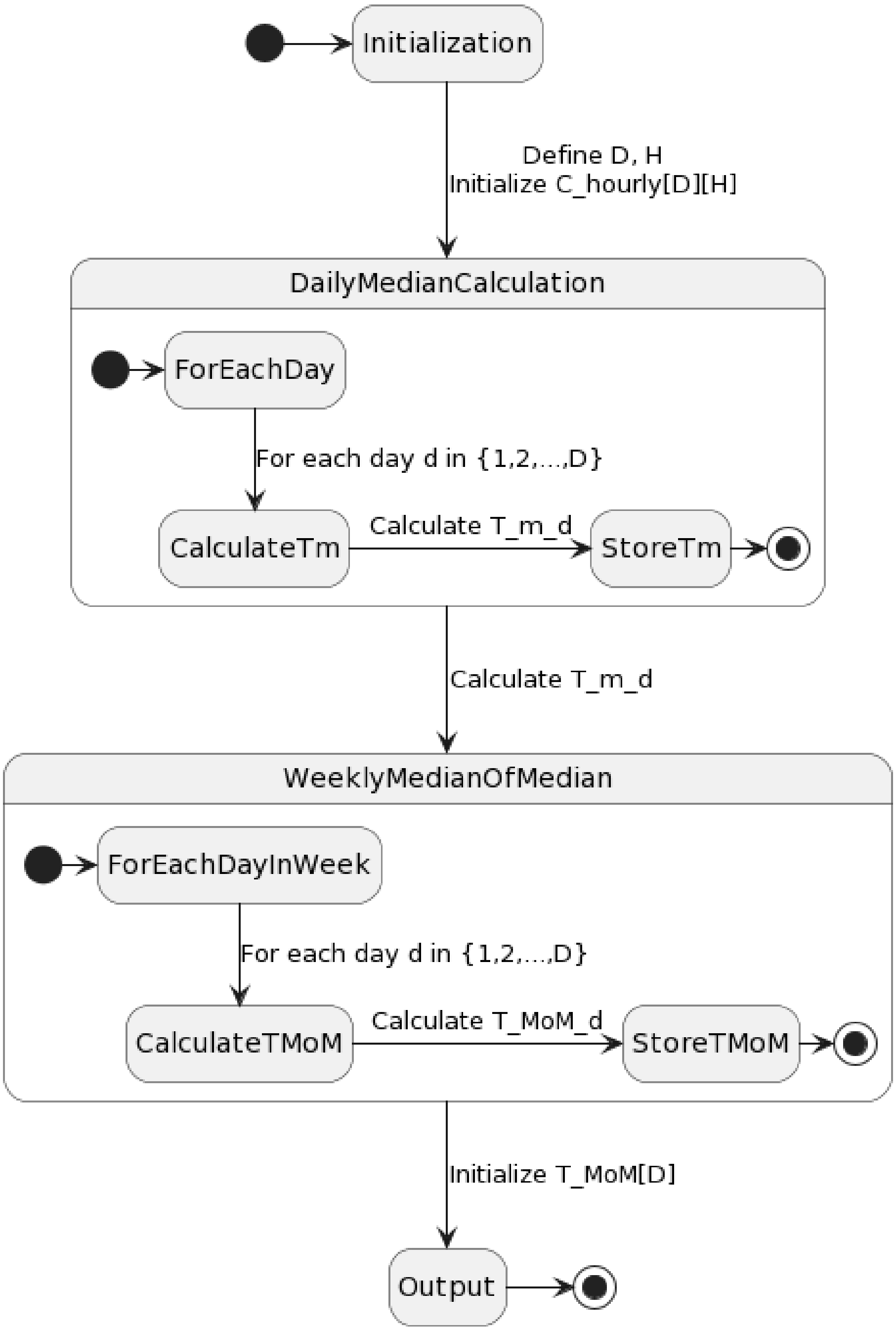


Figure 4-7: Flow Diagram for Median-of-Median Threshold Algorithm

Algorithm 4-5 present the sequential steps for computing the Median of Median threshold algorithm.

Algorithm 4-5: Median-Of-Median Threshold Computation

Step 1: Initialization

Define the number of days (D) and hours (H) in the observation period. Typically, $D = 7$ (days) and $H = 24$ (hours).

Initialize a 2D array. $C_{hourly}[D][H]$ to store the hourly crowd counts for each day and hour.

Step 2: Daily Median Calculation:

For each day d in $\{1,2,\ldots,D\}$:

T_{m_d} It is calculated as the median of hourly counts for day d.

$$T_{m_d} = \text{median}\left(\{C_{hourly}[d][1], C_{hourly}[d][2], \ldots, C_{hourly}[d][H]\}\right)$$

Step 3: Weekly Median-of-Median Calculation:

Initialize an array $T_{M-o-M}[D]$ to store the median-of-median thresholds.

For each day d in $\{1,2,\ldots,D\}$:

Calculate T_{M-o-M_d} as the median of the array $T_m[1:d]$.

$$T_{M-o-M_d} = \text{median}\left(\{T_{m_1}, T_{m_2}, \ldots, T_{m_d}\}\right)$$

Step 4 : Output:

The algorithm outputs the array T_{M-o-M}, which contains the median-of-median thresholds for each day of the week.

End of Algorithm

The M-o-M algorithm 4-5 is simple yet profound, and at the same time, that is what makes it easy to model high volume of data that will be collected from thousands of mobile towers in an urban infrastructure. A weekly and monthly threshold in crowd pattern analysis unveils urban rhythms. It is more than just data. It is how cities unfold themselves in hours and days, revealing their sophisticated rhythms.

4.7 Validation Median Vs. Average

In crowd density estimation using MOBILE tower data (**C_{count}**), the choice between using the median or the average (mean) of threshold levels has significant implications. Both statistically and logically, there are reasons why the median is often preferred over the average for such data. Here's a comparative analysis:

4.7.1 Mathematical and Statistical Perspective

1. **Outlier Sensitivity**:

Average: The average (mean) is sensitive to outliers or extreme values. In the context of **Ccount**, an unusually high count (perhaps due to a special event) can skew the average, making it unrepresentative of typical crowd conditions.

Assume we have a set of daily crowd counts for a week from a MOBILE tower: ' [100, $110,105,1500,115,120,130]'$. The value ' $1500'$ It is an outlier, likely due to a special event.

Average Calculation:

$$\text{Average } = \frac{100 + 110 + 105 + 1500 + 115 + 120 + 130}{7} \approx 311.4$$

Median: The median is more robust against outliers. It represents the middle value in a sorted list of numbers and is less affected by extreme values. This quality makes the median a more reliable indicator of 'typical' crowd density, especially in environments where outliers are common.

Median Calculation:

Sorted Data: ' $[100,105,110,115,120,130,1500]$ '

Median (middle value): '115'

The median (' 115°) is more representative of typical data than the average ('311.4'), skewed by the outlier.

2. **Skewed Data**:
    - **Average**: When the data distribution is skewed (not symmetric), the average can be misleading. For instance, the average might suggest higher crowd density than commonly experienced in areas with typically low crowd counts but occasional spikes.

- **Median**: The median is not affected by how the data is skewed. It effectively splits the dataset into two halves, offering a more accurate reflection of the central tendency in skewed distributions.

Consider a scenario with weekly 'Ccount ' values skewed by low regular counts and occasional high peaks: ' [80,85,90,95,1000,105,110] '.

The ' 1000 ' count would disproportionately influence the average, whereas the median would still reflect the more typical range of data.

4.7.2 Diagrammatic Comparison of Median vs Mean

Figure 4-8 illustrates how an outlier affects calculating the mean and median in a crowd count for one week. On the fourth day, there is a remarkable increase in the number of participants, unlike on the other days. This outlier causes the mean to shift significantly upward, as it now becomes 311.43 instead of the original medium reading of 115.0, which is almost untouched by this outlier. This emphasizes mean’s weakness towards extreme values in comparison with median as a central tendency indicator when outliers are present.

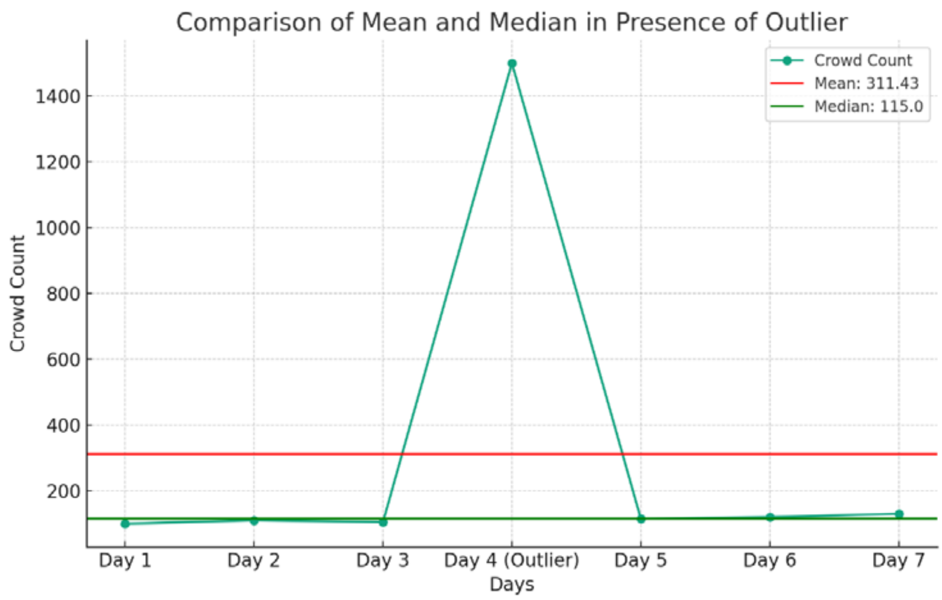


Figure 4-8: Mean Vs. Median Conceptual Comparison

4.7.3 Logical and Practical Perspective

1. **Representative Value**: In practical scenarios, city planners and managers are often more concerned with typical, day-to-day crowd levels rather than occasional extremes. The median provides a more accurate reflection of this 'typical' crowd density.

2. **Stability Over Time**: The median tends to be more stable over time, especially in the face of sporadic spikes or drops in crowd density. This stability is crucial for making long-term plans and assessments.

3. **Predictive Analysis**: When using historical data to predict future conditions, a measure less sensitive to extreme variations (like the median) can provide more reliable forecasts. This is particularly relevant for consistent crowd management and resource allocation.

Mathematically and logically, the median of threshold levels offers a more robust and representative measure for crowd density estimation using MOBILE tower data. It provides a more reliable baseline for understanding typical crowd conditions, is less affected by data irregularities, and offers greater stability for predictive analysis and planning. While the average can still be useful for understanding overall trends, its susceptibility to outliers and skewed distributions can be limiting in the nuanced context of urban crowd dynamics.

4.8 Evolution of Proposed DUCSIM Crowd Density Algorithm

4.8.1 Proposed Algorithm 1: Quartile Classification for Crowd Density

A more detailed approach is the Comprehensive Algorithm (Quartile Classifications) for Crowd Density Assessment as an instrument of Urban Crowd Management. Quartile classification for analysing population dynamics daily and weekly using the MOBILE cell tower data. Moreover, this algorithm 4-6 utilizes the median threshold, classifies quartiles' patterns, and enables prediction and planning. The strategy is especially applicable to urban planners and authorities to achieve an effective distribution of resources as well as secure public safety.

A. Mathematical Presentation

To present the mathematical aspects of the Comprehensive Algorithm for Crowd Density Analysis with Quartile Classifications in a more detailed format, we'll delve deeper into the equations and statistical concepts used:

1 Initialization and Data Structure:

Let n be the total number of MOBILE towers.

Define $C_{count}[d][h]$ as a 2D array to store cumulative crowd counts from each MOBILE tower, where d is the day (1 to 7) and h is the hour (1 to 24).

2 Daily and Weekly Threshold Calculations:

Daily Threshold: For each day d and hour h,

$$T_{daily}[d][h] = \text{median}\left(\{C_{count}[d][1], C_{count}[d][2], \ldots, C_{count}[d][h]\}\right)$$

Eq: (4-10)

Weekly Threshold: For each hour h,

$$T_{weekly}[h] = \text{median}\left(\{C_{count}[1][h], C_{count}[2][h], \ldots, C_{count}[7][h]\}\right)$$

Eq: (4-11)

3 Quartile Calculation and Classification:

For each array of counts (either daily or weekly), calculate quartiles $Q1, Q2$ (median), $Q3$, and $Q4$ to classify crowd density. Quartiles are calculated using standard statistical methods, where:

$Q1$ is the median of the lower half of the data (excluding the median if an odd number of data points).

$Q2$ is the median of the data set.

$Q3$ is the median of the upper half of the data.

Crowd density classification for a given count x is done as follows:

If $x < Q1$, classify as 'Low'.

If $Q1 \leq x < Q2$, classify as 'Medium.'

If $Q2 \leq x < Q3$, classify as 'High.'

If $x \geq Q3$, classify as 'Very High'.

The meticulous mathematical format constitutes a sound basis for the statistically substantiated algorithm. The technique enables the measurement of crowd density and prediction, which is critical in predictive urban management and planning. Including quartile analysis deepens comprehension of crowd dynamics and thus makes the approach more practical.

B. Diagrammatic Presentation

A process flow of crowd management through data analytics is demonstrated in the Figure 4-9 commencing from variable initialization and array creation. It adopts a systematic procedure based on data retrieval for the day and hour, threshold determination per day and week, as well as categorizing data as quartiles. It utilizes historical data for predictive mapping and crowd way finding; this methodology is proposed to enable intelligent management and direction of groups or masses within the society. The given flow is indicative of a systematic plan for making data-oriented decisions regarding crowd control cases.

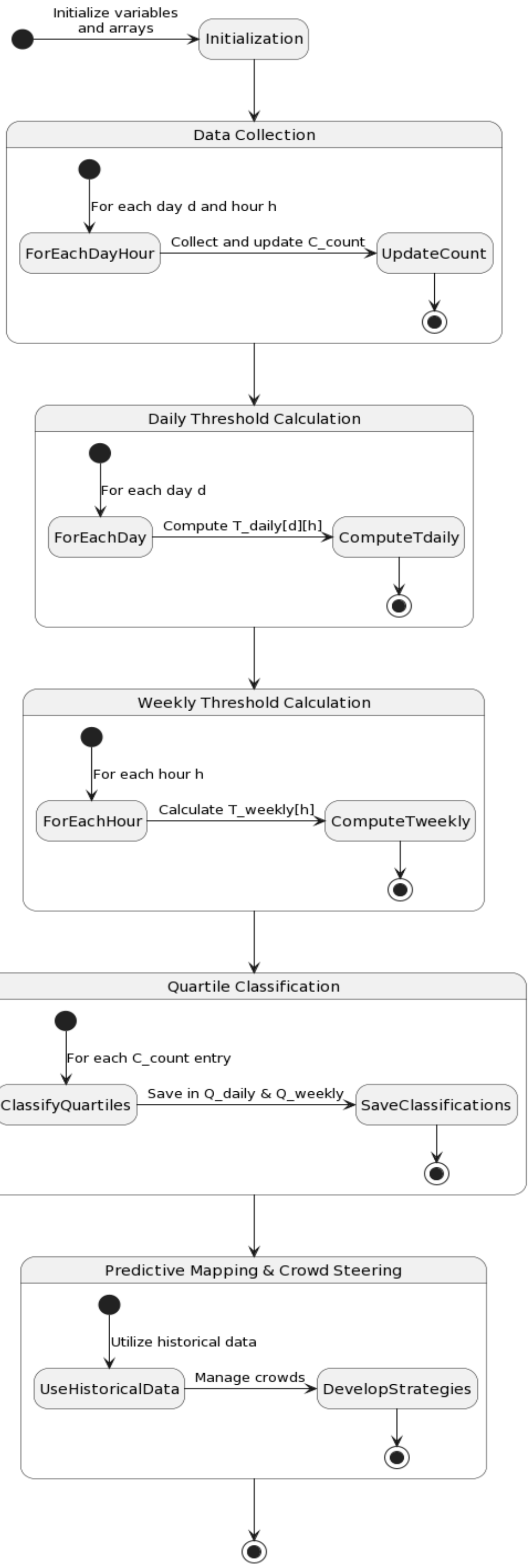


Figure 4-9: Flow Diagram of Comprehensive Algorithm for Crowd Density Analysis

Algorithm 4-6: Multivariate Crowd Density Analytical Framework (CD_AF)

Step 1: Initialization
Initialize the number of MOBILE towers, n.
Set up arrays:
C_{count} [d][h]: 2D array to store crowd counts from each MOBILE tower for each hour h and day d.
T_{daily} [d][h] and T_{weekly} [h]: Arrays for daily and weekly threshold values.
Q_{daily} [d][h] and Q_{weekly} [h]: Arrays for storing quartile classifications.

Step 2: Data Collection
For each day d and each hour h :
Collect and update the crowd count C_{count} [d][h].

Step 3: Calculating Daily Threshold Values
For each day d and each hour h :
Compute the daily threshold T_{daily} [d][h] as the median of C_{count} [d][1] to C_{count} [d][h].

Step 4: Calculating Weekly Threshold Values
For each hour h :
Aggregate crowd counts for hour h across all days.
Calculate the weekly threshold. T_{weekly} [h] as the median of these aggregated counts.

Step 5: Quartile Classification for Crowd Density
For each C_{count} entry (hourly and daily):
Classify into quartiles based on the distribution of crowd counts.
Use quartiles Q1, Q2, Q3, Q4 to categorize crowd density as Low, Medium, High, and Very High.
Apply this classification relative to T_{daily} or T_{weekly} to determine the density level.

Step 6: Storing Quartile Classifications
Save the quartile classifications in Q_{daily} and Q_{weekly} .

Step 7: Mapping Crowd Density for Prediction
Utilize historical C_{count} data and quartile classifications to establish a predictive mapping for future crowd densities.

Step 8: Crowd Steering and Management
Use predictions and classifications to develop strategies for managing crowds, especially during peak times or events.

The algorithm 4-6 uses statistical analysis and data-driven insights to understand and manage town crowd behaviour. The method utilizes median and quartile calculations to systematically categorize and forecast crowd density, which is essential for effective town planning and crowd control. A basic structure for an overall approach to crowd density analysis is algorithm 4-6. It commences with the initialization stage involving allocating memory space for crowd count arrays/data structures, threshold values, and quartile classifications. It continues with data collecting, determination of the daily and weekly thresholds, and categorizing crowd densities into quartiles. These quartile classifications accumulate, and historical data is used to predict future crowd density and an ordered manner of seeing and labelling people as they come together.

Algorithm 4-7 proposed in next section is a natural extension of Algorithm 4-6, shifting from reactive analysis to active crowd control. Algorithm 4-6 looks into the collection and analysis of data, whereas Algorithm 4-7 uses historical data along with quartile groupings to arrive at optimal measures for guiding and managing crowds. The tool presents step-by-step actions, i.e., resource allocation and control of crowds in different situations, such as during rush hours or unique occasions. Essentially, Algorithm 4-7 builds on Algorithm 6 to develop and utilize the knowledge gained to purposefully direct the crowd's behaviour and improve crowd-focused procedures, thereby increasing the effectiveness and efficiency of crowd management.

4.8.2 Proposed Algorithm 2: Historical Threshold-Based Crowd Density Classification

Algorithm 4-7 is a next-generation algorithm that builds on what was established by algorithm 4-6. This is an advanced process of crowd steering and controlling. Algorithm 7 helps improve the decision-making process by utilizing historical crowd data, quartile classifications, and predictive modelling to achieve timeliness and relevance in the decision-making process. It facilitates resource allocation and crowd control on-site at peak hours and when many people are on-site, for instance, during major events. The above algorithm links data analysis with on-ground crowd-control measures in diverse situations, improving crowd management's safety, orderliness, and coordination across any environment. This characteristic makes it an excellent instrument for guaranteeing the crowd's security and contentment.

A. Median Calculation

For each hour t, compute the median of 'Ref_Count' values. Let RefCounts be the $_t$ Set of 'Ref_Count' values for hour t. The median M_t It is defined as:

$$M_t = \text{median}\left(\text{RefCounts}_{\ t} \right)$$

Eq: (4-12)

- Initial Classification

Classify each 'Current_Count ' $C_{t,i}$ at hour t as 'Above Median' or 'Below Median' based on comparison with M_t :

$$\text{Initial Class}_{\ t,i} = \begin{cases} \text{"Above Median"} & \text{if } C_{t,i} \geq M_t \\ \text{"Below Median"} & \text{if } C_{t,i} < M_t \end{cases}$$

Eq: (4-13)

- Quartile-Based Further Classification

For each classification ('Above Median' and 'Below Median'), determine the quartiles within that subset for hour t. Denote the first and third quartiles of 'Current_Count ' within each classification as $Q1_{t,\ class}$ and $Q3_{t,\ class}$ respectively. Then, classify each $C_{t,i}$ As:

$$\text{Final Class}_{\ t,i} = \begin{cases} \text{"Low"} & \text{if } C_{t,i} < Q1_{t,\ class} \\ \text{"Medium"} & \text{if } Q1_{t,\ class} \leq C_{t,i} < Q3_{t,\ class} \\ \text{"High"} & \text{if } C_{t,i} \geq Q3_{t,\ class} \end{cases}$$

Eq: (4-14)

Where,

Final Class $t_{t,i}$ Is the final classification of 'Current_Count' at hour t and instance i.

$Q1_{t,\ class}$ and $Q3_{t,\ class}$ are the quartile thresholds calculated for each subset of data ('Above Median' or 'Below Median') for each hour.

B. Diagrammatic Presentation

The flow diagram in figure 4-10 shows how hourly data records are analysed. Starting from the initialization stage, it analyses an hour’s record with the help of a median calculation for the corresponding hour. An initial classification follows to find out whether a particular entry value is higher or lower than the mean. According to these categories are established, quartiles are calculated using the records higher or lower than the median, resulting in a finer classification of data. This data analysis sequence ends up in the finalization step which is the last one in the whole process.

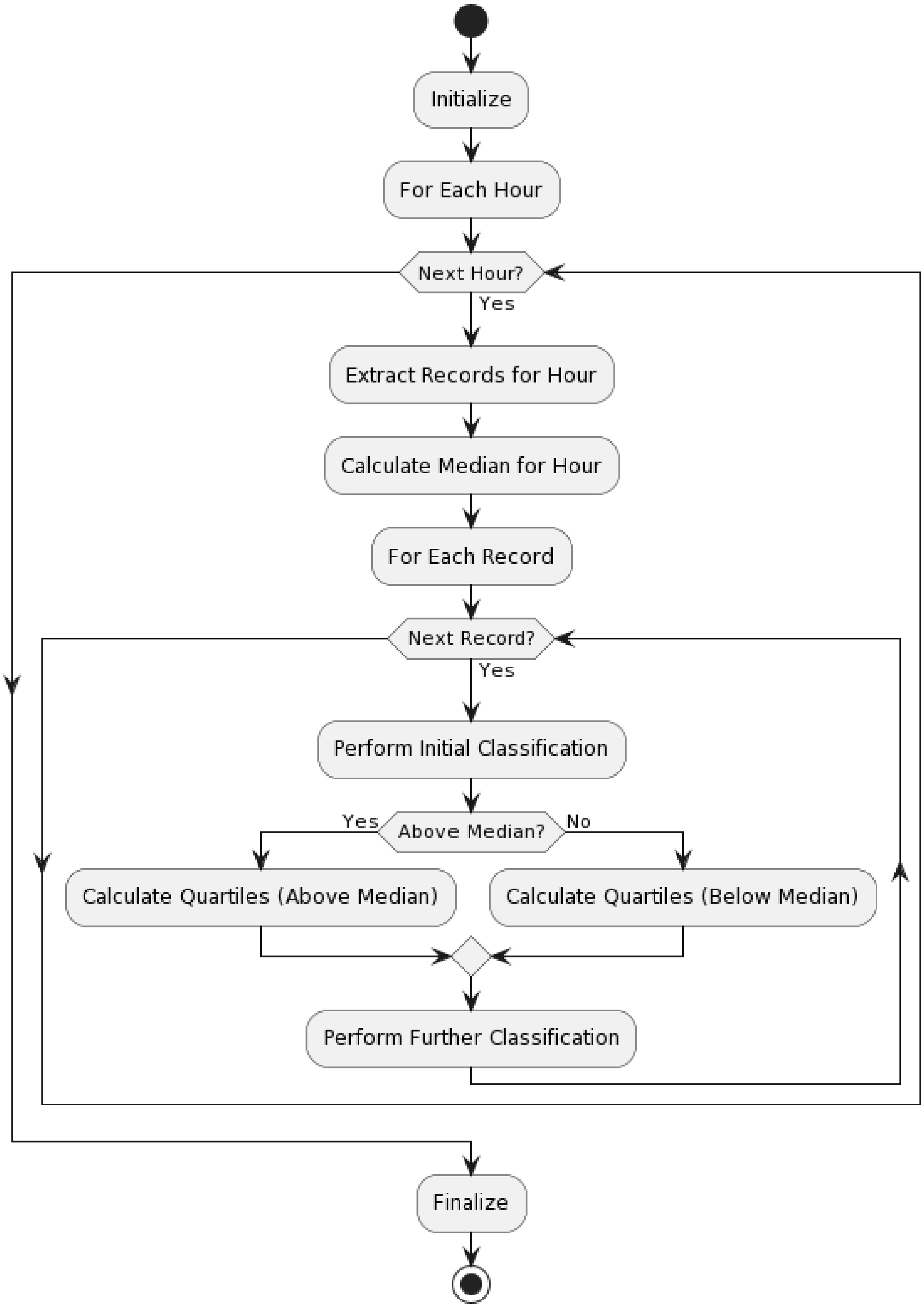


Figure 4-10: Flow of Historical Threshold-Based Crowd Density Classification

Algorithm 4-7: Proactive Crowd Management Paradigm (PCMP)

Inputs:

Data: A set of records, each containing 'Time,' 'Ref_Count,' and 'Current_Count.'
Hours: A set of distinct hours in the dataset.

Process:

For Each Hour:
$t \in$ Hours
Extract all records for hour t from Data.
Calculate the median of 'Ref_Count' for hour t, denoted as M_t.
Initial Classification Function:
ClassifyInitial $(C_{t,i}, M_t)$ → "Above Median" or "Below Median"
Quartile-Based Further Classification Function:
ClassifyQuartiles $(C_{t,i}, Q1_{t,\text{ class}}, Q3_{t,\text{ class}})$ → "Low", "Medium", "High"

Main Algorithm:

For each record i at hour t :
Perform initial classification:
InitialClass $_{t,i}$ = ClassifyInitial $(C_{t,i}, M_t)$
Calculate quartiles $Q1_{t,\text{ class}}$ and $Q3_{t,\text{ class}}$ for the classified group.
Perform further classification:
FinalClass $_{t,i}$ = ClassifyQuartiles $(C_{t,i}, Q1_{t,\text{ class}}, Q3_{t,\text{ class}})$

Outputs:

A classification for each record in Data, indicating 'Low,' 'Medium,' or 'High' crowd density within the 'Above Median' or 'Below Median' groups.

Algorithm 4-7 is a breakthrough in the theory of crowd management. It follows algorithm 4-6 based on full-density analyses of a crowd. Algorithm 4-7 extends this analysis by including historical data, QUARTILES, and predictive modelling that can be used in real-time to guide and control dense crowd scenes. It provides decision-makers with the right equipment for the allocation of resources and crowd management to ensure effective response during peak times/special events.

The next is Algorithm 4-8 is an upgraded version of algorithm 4-7 and is meant for categorization and control of crowd density. On the other hand, algorithm 4-8 is a straightforward approach that classifies individuals based on threshold levels. It puts the dataset to hours, calculates the hourly median thresholds, classifies the data points to “above threshold” or “below threshold” for each hour, and performs the quartile classification on the whole dataset. Algorithm 4-8 enables fast decision support based on data, making crowd control simpler to enhance public safety and satisfaction.

4.8.3 Proposed Algorithm 3: Crowd Density Classification with Median Threshold

The Crowd Density Classification and Threshold Computation Algorithm analyses and categorizes data based on crowd counts obtained from MOBILE tower records. Its primary function is to calculate hourly median thresholds of crowd density and subsequently classify each record based on these thresholds. The algorithm utilizes quartile-based classification to categorize crowd density into distinct levels. It is particularly useful for urban planning, event management, and scenarios where understanding crowd patterns is crucial for decision-making.

A. Mathematical Presentation

- Initialization (Grouping Data by Hour):

Let $\mathcal{D}$ represent the dataset, where each record $r \in \mathcal{D}$ has the attribute 'Hour.' (h_r) and 'Current_Count' (c_r).

Define $\mathcal{D}_{\text{grouped}}$ as a set of subsets of $\mathcal{D}$, where each subset corresponds to a unique hour:

$$\mathcal{D}_{\text{grouped}} = \{\mathcal{D}_h \mid \mathcal{D}_h = \{r \in \mathcal{D} \mid h_r = h\}, h \in \text{ Hours }\}$$

Eq: (4-15)

- Compute Hourly Median Thresholds:

For each group $\mathcal{D}_h$ in $\mathcal{D}_{\text{grouped}}$:

Calculate the median threshold. T_h, which is the median of 'Current_Count' values in $\mathcal{D}_h$:

$$T_h = \text{median}\left(\{c_r \mid r \in \mathcal{D}_h\}\right)$$

Eq: (4-16)

T_h is the threshold value for hour h.

- Categorization of Counts (Above/Below Threshold):

For the categorization based on threshold comparison, the algorithm processes each record r in the dataset $\mathcal{D}$. The key steps are as follows:

Threshold Comparison for Each Record:

Let c_r Denote the 'Current_Count' for a record r.

Let h_r Be the 'Hour' attribute of the record r.

Each record r is associated with a threshold. T_{h_r}, which is the median

'Current_Count' for the hour h_r It is computed previously.

For each record r, the algorithm compares. c_r with T_{h_r} :

$$\text{Category (r)} = \begin{cases} \text{"Above Threshold"} & \text{if } c_r > T_{h_r} \\ \text{"Below Threshold"} & \text{otherwise} \end{cases}$$

Eq: (4-17)

This step categorizes each record as 'Above Threshold' or 'Below Threshold' based on whether its count exceeds the median for its respective hour.

- Quartile Classification:

This module of the algorithm involves classifying each record into quartiles based on the overall distribution of 'Current_Count' in the dataset:

1 Determination of Quartile Thresholds:

Calculate the quartile thresholds $Q1, Q2, Q3$ for the 'Current_Count' values across all records in $\mathcal{D}$.

These thresholds are computed using the quartile function:

$$Q1, Q2, Q3 = \text{quartiles}\ (\{c_r \mid r \in \mathcal{D}\})$$

Eq: (4-18)

Q1 is the lower quartile, Q2 is the median, and Q3 is the upper quartile of the 'Current_Count' distribution.

2 Classification into Quartiles:

Each record r is then classified into a quartile category based on c_r :

$$\text{QuartileClass (r)} = \begin{cases} \text{"Low"} & \text{if } c_r \leq Q1 \\ \text{"Medium"} & \text{if } Q1 < c_r \leq Q2 \\ \text{"High"} & \text{if } Q2 < c_r \leq Q3 \\ \text{"Very High"} & \text{if } c_r > Q3 \end{cases}$$

Eq: (4-19)

This classification assigns each record a label ('Low,' 'Medium,' 'High,' 'Very High') based on how its 'Current_Count' compares to the overall distribution of counts in the dataset.

B. Diagrammatic Presentation

Figure 4-11 diagram explains a data analysis process which uses a dataset formed up by hours for calculating the median count per hour after each record is labelled either high over the median threshold or not (quartiles). The procedure ends with exports of the categorised and classified results together with calculated thresholds are presented in results.

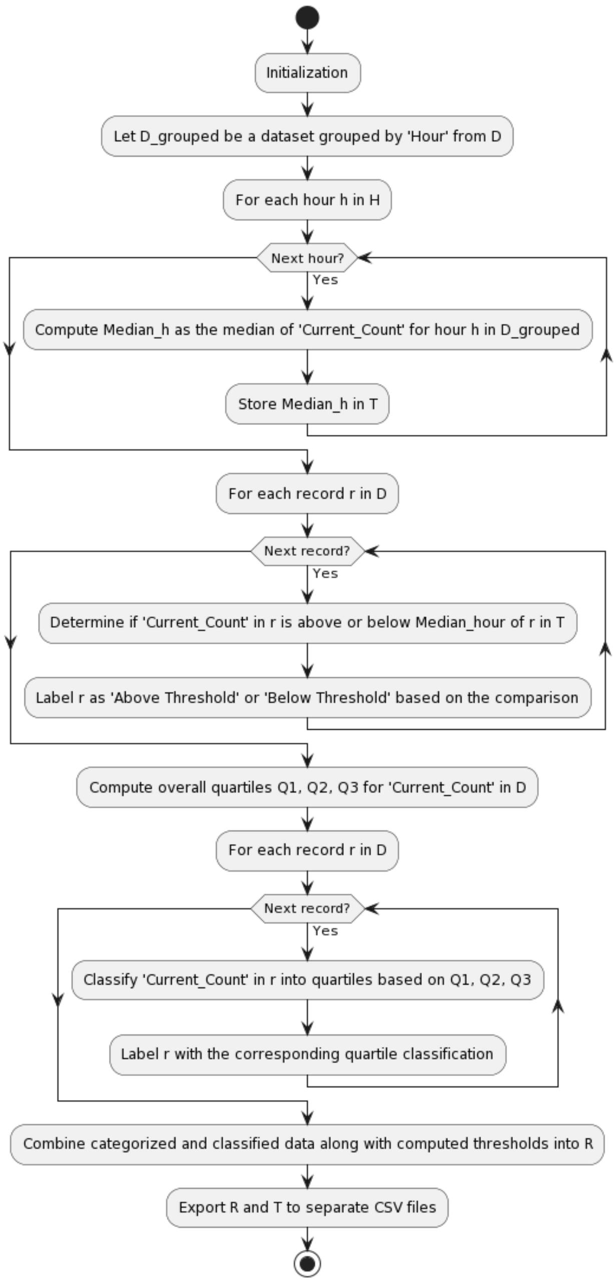


Figure 4-11: Crowd Density Classification with Median Threshold

Algorithm 4-8: Streamlined Crowd Density Taxonomy (SCDT) (Median Threshold)

Inputs:

$\mathcal{D}$: Dataset containing records with fields 'Hour,' 'Current_Count,' and other relevant information.

H : Set of unique hours in $\mathcal{D}$.

Outputs:

$\mathcal{R}$ Resultant dataset with classifications and computed thresholds.

Procedure:

1 Initialization:

Let $\mathcal{D}_{\text{grouped}}$ be a dataset grouped by 'Hour' from $\mathcal{D}$.

2 Compute Hourly Median Thresholds:

For each hour $h \in H$:

Compute Median $_h$ as the median 'Current_Count' for hour h in $\mathcal{D}$.

Store Median $_h$ in a collection $\mathcal{T}$.

3 Categorization of Counts:

For each record r in $\mathcal{D}$:

Determine if 'Current_Count' in r is above or below the Median hour of r in $\mathcal{T}$.

Label r as 'Above Threshold' or 'Below Threshold' based on the comparison.

4 Quartile Classification:

Compute overall quartiles Q1, Q2, Q3 for 'Current_Count' in $\mathcal{D}$.

For each record r in $\mathcal{D}$:

Classify 'Current_Count' in r into quartiles based on Q1, Q2, Q3.

Label r with the corresponding quartile classification.

5 Compilation of Results:

Combine the categorized and classified data along with computed thresholds into $\mathcal{R}$.

6 Export to CSV:

Export $\mathcal{R}$ and $\mathcal{T}$ to separate CSV files.

End of Algorithm

This algorithm 4-8 efficiently processes MOBILE tower data to provide insights into crowd density patterns. Classifying data into quartiles and computing hourly median thresholds offers a structured approach to understanding and predicting crowd behaviours. The algorithm's output can enhance crowd management strategies, improving safety, resource allocation, and overall crowd control measures. Its application is valuable in various contexts where crowd dynamics are critical. Threshold-based categorization of simple crowd density classifications dominated in Algorithm 4-8.

On the contrary, Algorithm 4-9 goes far ahead by looking at personal mobility patterns and social interaction dynamics in crowds. It applies a 24-hour movement profile on each person based on mobilities via MOBILE tower connections. With these profiles, one can cluster people into groups and subsequently estimate the strength of the ties among them.

Algorithm 4-9 utilizes time and place analysis to determine how long a group stayed together and how often it reunited. Lastly, it utilizes predictive modelling to forecast upcoming mobility and social interactions. The shift reflects a shift from simple threshold-oriented classification to deeper insight into the mechanisms of crowd action.

4.8.4 Proposed Algorithm 4: Social Mobility of Crowd – Individual & Group

This algorithm 4-9 delves into the intricate study of individual mobility and social connections within an urban setting, leveraging MOBILE tower data to track personal trajectories and social interactions. Mapping each person's hourly connection to specific MOBILE towers constructs a detailed tapestry of movement patterns. The algorithm 4-9 further discerns social ties and random associations among individuals, offering a nuanced perspective on how personal paths intersect in the bustling urban environment.

A. Mathematical Presentation of the Algorithm

The algorithm for analysing individual mobility patterns and social/random ties using MOBILE tower data integrates several mathematical and statistical concepts. Here's a more detailed breakdown:

1 Data Structure and Collection:

Let n be the total number of individuals tracked.

Let $T_{id}[i][h]$ denote the MOBILE tower ID that individual i (where $i = 1,2, \dots, n$) is connected to at hour h (where $h = 1,2, \dots, 24$).

Each MOBILE tower is mapped to a specific location in the city.

2 Mobility Pattern Analysis:

Construct a mobility matrix $M_{profile}$ of size $n \times 24$, where each entry $M_{profile}[i][h]$ represents the location of individual i at hour h.

The mobility profile for each individual, $M_{profile}[i]$ is a sequence showing the MOBILE tower connections over the day.

3 Group Identification and Relationship Inference:

For each hour h, identify groups G_h of individuals who are connected to the same MOBILE tower. This forms a collection of sets $\{G_h^1, G_h^2, ...\}$ For each hour.

Define a social tie strength metric. S_{tie} [G] for each group G, which quantifies the frequency and consistency of the group's members being at the same location.

Increase S_{tie} [G] for groups frequently appearing together across multiple hours/days.

Classify relationships within each group G :

If S_{tie} [G] exceeds a predefined threshold, classified as 'social.'

Otherwise, classify as 'random'.

4 Time and Location Analysis for Relationships:

Analyse the duration D_G and frequency F_G Of encounters within each group G to further validate the strength of inferred social ties.

5 Integration with Crowd Density Analysis:

Combine individual movement data with crowd density metrics to understand the impact of individual and group behaviour on overall crowd dynamics.

The algorithm leverages MOBILE tower data to analyse individual mobility patterns and social ties in urban environments. It tracks individual locations over time, identifies groups based on shared locations, and classifies relationships as 'social' or 'random.'

B. Diagrammatic Presentation

Figure 4-12 illustrates the flow how the data was analyzed. A set of samples classified according to time, is taken to estimate the median frequency for each hour after which they are labelled above or below this median threshold to finally categorize them under quartiles. Finally, the categorized and classified data as well as calculated thresholds are exported into separate CSV files to be used in other applications or analyses.

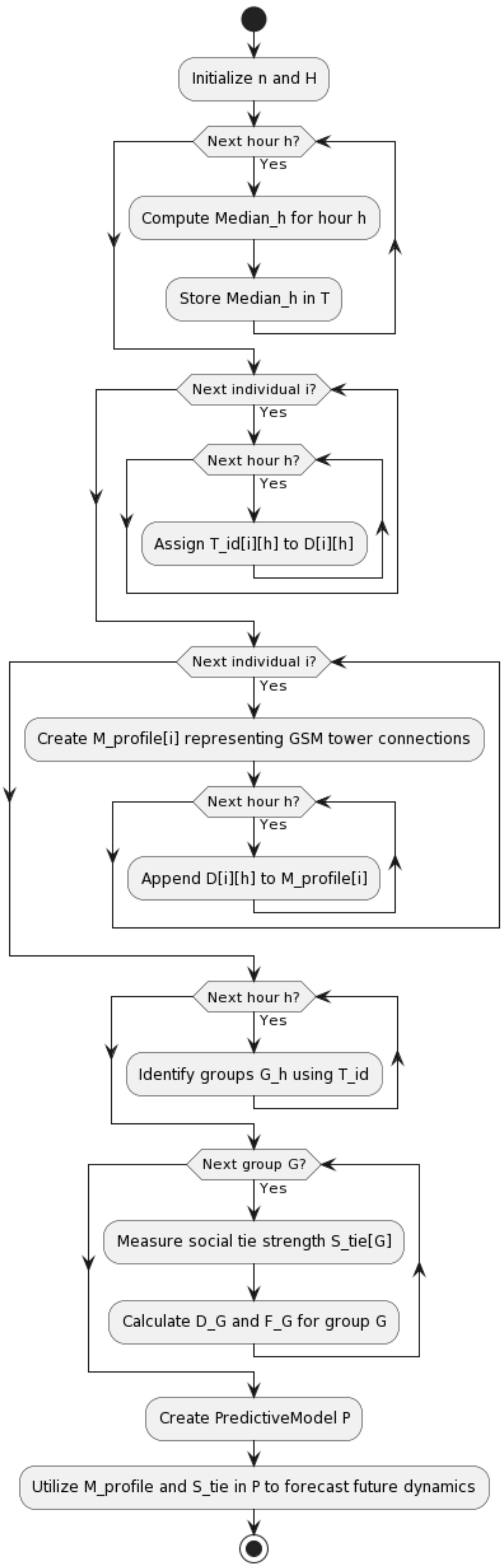


Figure 4-12: Social Mobility of Crowd – Individual & Group

Algorithm 4-9: Individual and Social Dynamics Integration (ISDI)

Step 1: Input
Let n be the number of individuals.
Let H be the number of hours in the analysis period, typically 24.

$T_{id}[i][h]$: MOBILE tower ID that individual i (where $i = 1,2, \dots, n$) is connected to at hour h (where $h = 1,2, \dots, H$).

Compute Hourly Median Thresholds:
For each hour $h \in H$:
Compute Median $_h$ as the median of 'Current_Count' for hour h in $\mathcal{D}$.
Store Median $_h$ in a collection $\mathcal{T}$.

Step 2: Process
Formulate a data matrix D of size $n \times H$ where each entry $D[i][h] = T_{id}[i][h]$.

Mobility Pattern Analysis
Define $M_{profile}[i]$ as a vector representing the sequence of MOBILE tower connections for each individual i over H hours.
$M_{profile}[i] = (D[i][1], D[i][2], \dots, D[i][H])$.

Group Identification:
For each hour h, identify groups G_h using:
$G_h = \bigcup_{i=1}^{n} \{i \mid D[i][h] = k\}$, for each tower ID k.

Relationship Classification:
Define $S_{tie}[G]$ as a function measuring the social tie strength within group G, calculated as:
$S_{tic}[G] = \sum_{h=1}^{H} \omega_h \times Ind(G, h)$, where ω_h is a weighting factor for hour h, and $Ind(G, h)$ is an indicator function that is 1 if G meets at hour h, 0 otherwise.

$$Ind(G, h) = \begin{cases} 1 & \text{if group G meets at hour h,} \\ 0 & \text{otherwise.} \end{cases}$$

Time and Location Analysis:
Analyse D_G and F_G where D_G is the total duration and F_G The frequency of encounters for group G.

Predictive Modelling:
Utilize $M_{profile}$ and S_{tic} in a predictive model, $\mathcal{P}$ to forecast future dynamics.
$\mathcal{P}(M_{profile}, S_{tic}) \rightarrow$ Forecast of future movements and interactions.

Step 3: Output
Mobility profiles $M_{Profile.}$.
Group dynamics and social tie strengths G_h and $S_{tie}[G]$.
Predictive models for future mobility and social interaction patterns.

This detailed algorithm presentation underscores the algorithm's capacity to unravel complex urban dynamics by tracking individual movements and inferring social ties. By synthesizing granular MOBILE data into meaningful patterns and relationships, the algorithm offers profound insights into individual mobility and social interactions within urban environments. This approach is invaluable for urban planning and social research, providing a data-driven foundation for understanding and shaping the fabric of city life.

Algorithm 4-9 went a long way to help understand individual movement habits within crowds and how people interacted socially. On the other hand, such details are integrated into microscopic crowd density analysis in Algorithm 4-10. By taking advantage of mobile phone data from MOBILE towers and counting and tracking individuals, it is possible to analyse how crowds move in cities. Predictive modelling, F (C, M), is introduced in Algorithm 4-10 to enable the prediction of future crowd dynamics and their impact on society. This transition reflects the evolution of looking at the crowd behaviour in isolation but towards an integrative and predictive approach, which helps improve urban crowd management practices.

4.8.5 Proposed Algorithm 5: Integrated Model for Urban Crowd Dynamics and Social Interaction Analysis

The Algorithm 4-10 presented here is an amalgamation of Algorithms 4-8 and 4-9 , which integrate as one framework computing the crowd density and the crowd mobility in both group and individual manners. The Algorithm 4-10 is further presented below in detail. The "Dynamic Urban Crowd and Social Interaction Model" (DUCSIM) stands at the forefront of urban analytics, offering a dual-layered approach to deciphering the complexities of city life. Harnessing MOBILE tower data, DUCSIM delves into the macroscopic world of crowd density analysis alongside the microscopic realm of individual mobility and social interactions. This innovative model illuminates the intricate dance between individual behaviours and collective crowd dynamics, painting a comprehensive picture of urban social fabric.

A. Mathematical Presentation of DUCSIM

- **Part One: Macroscopic Crowd Density Analysis**

- Location and MOBILE Tower Representation:

Define $L = \{L_1, L_2, \dots, L_n\}$ As the set of city locations.

Define $T = \{T_1, T_2, \dots, T_n\}$ As the set of corresponding MOBILE towers.

- Crowd Count and Threshold Calculation:

Let $C(t, h, d)$ denote the crowd count at tower t, hour h, and day d.

Daily Threshold Calculation:

$$T_{daily}(d, h) = \text{median}(\{C(t, 1, d), C(t, 2, d), \dots, C(t, h, d)\}) \text{ for all } t \in T$$

Eq: (4-20)

Weekly Threshold Calculation:

$$T_{weekly}(h) = \text{median}(\{C(t, h, 1), C(t, h, 2), \dots, C(t, h, 7)\}) \text{ for all } t \in T$$

Eq: (4-21)

- Quartile Classification:

Define quartiles $Q1, Q2, Q3, Q4$ based on the distribution of $C(t, h, d)$.

Classification of $C(t, h, d)$ is determined by its relation to these quartiles.

- **Part Two: Microscopic Individual Mobility and Social Interaction Analysis**

- Individual Tracking and Mobility Patterns:

Define a mapping function M: $M: P \times H \times D \to T$, where P is the set of individuals, H is the set of hours, D is the set of days, and T is the set of towers.

$M(p, h, d)$ maps individual p to a MOBILE tower t at hour h on day d.

- Social and Random Relationship Inference:

Define a similarity measure S: $S: P \times P \times H \times D \to \mathbb{R}$.

$S(p_1, p_2, h, d)$ quantifies the closeness of individuals based on shared locations and time. Aggregate S over time to deduce social ties and random encounters.

- Integration with Crowd Density Analysis:

Correlate individual mobility data $M(p, h, d)$ with macroscopic crowd density data $C(t, h, d)$.

- Predictive Modelling and Applications

Develop predictive models based on historical data: $\mathcal{F}(C, M) \rightarrow$ Predictions, where $\mathcal{F}$ is the forecasting function.

B. Diagrammatic Presentation

A basic DUCSIM algorithm is shown in figure 4-13 the flow diagram that starts from initialization to different analysis on macroscopic crowd density, microscopic individual movement and interaction. These analyses are combined to form predictive models that predict future crowd behaviour and social interactions, which then produce outputs like crowd density metrics, mobility profiles for individual people, and predictive models.

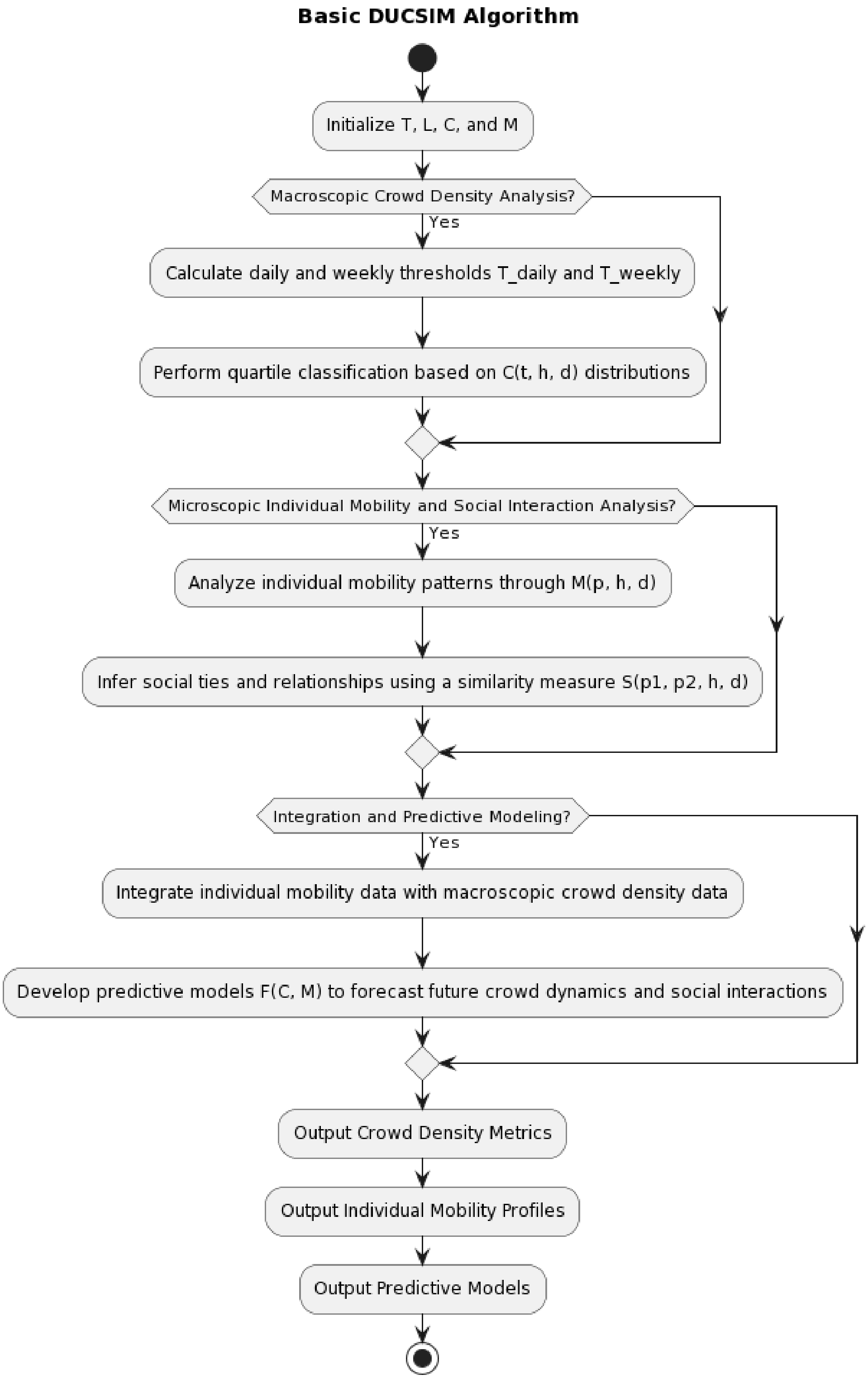


Figure 4-13: Basic DUCSIM Algorithm

Algorithm 4-10: Dynamic Urban Crowd and Social Interaction Model (DUCSIM)

Input

1 MOBILE Tower Data:

$T = \{T_1, T_2, \ldots, T_n\}$: Set of MOBILE towers in the city.

$L = \{L_1, L_2, \ldots, L_n\}$: Corresponding set of city locations.

2 Crowd Counts Data:

$C(t, h, d)$: Crowd count at tower t, hour h, and day d.

3 Individual Tracking Data:

$M(p, h, d)$: Mapping of individuals p to MOBILE towers t at specific times.

Process

1 Macroscopic Crowd Density Analysis:

Calculate daily and weekly thresholds:

$T_{daily}\ (d, h) = \text{median}\ \{C(t, 1, d), \ldots, C(t, h, d)\}$ for each t.

$T_{weekly}\ (h) = \text{median}\ \{C(t, h, 1), \ldots, C(t, h, 7)\}$ for each t.

Perform quartile classification based on $C(t, h, d)$ distributions.

2 Microscopic Individual Mobility and Social Interaction Analysis:

Analyse individual mobility patterns through $M(p, h, d)$.

Infer social ties and relationships using a similarity measure $S(p_1, p_2, h, d)$.

3 Integration and Predictive Modelling:

Integrate individual mobility data with macroscopic crowd density data.

Develop predictive models $\mathcal{F}(C, M)$ to forecast future crowd dynamics and social interactions.

Output

1 Crowd Density Metrics:

Daily and weekly threshold values: T_{daily} and $T_{Weekly.}$.

Quartile classifications for crowd densities.

2 Individual Mobility Profiles:

Movement patterns and social interaction data for each individual.

3 Predictive Models:

Models capable of predicting future urban crowd dynamics and social behaviour patterns.

DUCSIM Algorithm 4-10 effectively synthesizes detailed urban data into actionable insights, offering a multi-dimensional view of urban dynamics. This algorithm is particularly valuable for understanding the interplay between large-scale crowd movements and the nuanced patterns of individual behaviour, making it a powerful tool for urban planning and sociological studies. DUCSIM emerges as a groundbreaking tool in urban planning and sociological research, adept at forecasting and interpreting the multifaceted nature of urban crowds. Its predictive capabilities, rooted in a blend of detailed crowd density data and nuanced social interaction patterns, provide invaluable insights for city planners, emergency responders, and policymakers. The model's ability to intertwine macroscopic trends with microscopic human interactions marks a significant leap in our understanding of urban ecosystems, paving the way for more responsive and intelligent city management strategies.

Algorithm 4-10 marked improvement in crowd management as it combined a microscopic approach of individual mobility patterns with macroscopic crowd density analysis and allowed the prediction of behavioural movement. Next comes Algorithm 4-11 which becomes a significant advancement when considering cumulative crowd movement analysis, density forecasting, and refined prediction modelling. However, this enhanced approach goes beyond that to better understand crowd dynamics, track movement over time, and forecast future density variations, inferring evolving social association and enabling proactive steering of crowds. Algorithm 4-11 stresses model revision and an ongoing feedback loop approach to continuous improvement with a wide range of outputs for urban crowd control and planning. These achievements allow decision-makers to lead crowds in movement, allocate resources to better effect, and provide improved security.

4.8.6 Proposed Algorithm 6: DUCSIM Algorithm with Self-Learning Module

The Algorithm 4-11 is Enhanced Dynamic Urban Crowd and Social Interaction Model (DUCSIM) represents a groundbreaking advancement in urban analytics. Merging sophisticated mathematical algorithms with real-world MOBILE data offers a multifaceted approach to understanding urban crowd dynamics and the intricacies of social interactions. This innovative model is a testament to the power of integrating macroscopic and microscopic analyses to unravel the complex tapestry of urban crowd mobility.

A. Mathematical Presentation of Enhanced DUCSIM

The Enhanced DUCSIM Algorithm 4-11 is a comprehensive model combining complex mathematical techniques to analyse urban crowd dynamics and individual social interactions. This granular presentation focuses on detailed mathematical equations and formulations.

- **Mathematical Definitions and Data Structure**

City Locations and MOBILE Towers:

Define locations as $L = \{L_1, L_2, \ldots, L_n\}$.

MOBILE towers are represented as $T = \{T_1, T_2, \ldots, T_n\}$.

Crowd Count Function:

Let $C_{raw} : T \times H \times D \rightarrow \mathbb{N}$ be the function where $C_{raw}(T_i, h, d)$ gives the crowd count at the tower T_i, hour h, and day d.

- **Threshold Calculations**

Daily Threshold:

$$T_{daily}(d, h) = \text{median}\left(\{C_{raw}(T_i, 1, d), \ldots, C_{raw}(T_i, h, d)\}\right) \forall T_i \in T$$

Eq: (4-22)

Weekly Threshold:

$$T_{weekly}(h) = \text{median}\left(\{C_{raw}(T_i, h, 1), \ldots, C_{raw}(T_i, h, 7)\}\right) \forall T_i \in T$$

Eq: (4-23)

- **Cumulative Crowd Mobility Analysis**

Crowd Movement Dynamics Function:

Define $CM: T \times T \times H \times D \rightarrow \mathbb{R}$ where $CM(T_i, T_j, h, d)$ quantifies the crowd movement from T_i to T_j During hour h on day d.

Mobility Matrix: $M_{h,d} = [m_{ij}]$ where $m_{ij} = CM(T_i, T_j, h, d)$.

Density Estimation and Movement Prediction:

Density Estimation:

$$D_{est}(T_i, h, d) = \sum_j m_{ij} - \sum_j m_{ji}$$

Eq: (4-24)

Movement Prediction Model:

$F_{mobility}\left(M_{h,d}\right) \rightarrow M_{h+1,d}$

Eq: (4-25)

.

- **Individual Mobility and Social Interaction Analysis**

Individual Mapping Function:

Define $M: P \times H \times D \rightarrow T$ to map individuals to MOBILE towers.

Social Interaction Analyses:

Social Interaction Matrix : $S_{h,d} = \left[s_{pq}\right]$

where s_{pq} Measures interaction strength between p and q.

Update Function: $S_{updated} = F_{social}\left(S_{h,d}, M, h, d\right)$.

- **Predictive Modelling**

Historical Data Utilization:

Use historical crowd data C_{raw} and threshold values T_{daily}, T_{weekly} for predictive modelling.

Predictive Function:

$$P_{future}(h+1, d) = F_{predict}\left(T_{current}, F_{pattern}(H)\right)$$

Eq: (4-26)

Directional Prediction Model:

$$D_{future}(h+1, d) = F_{direction}(CM_{current}, CM_{historical})$$

Eq: (4-27)

- **Model Refinement and Feedback Loop:**

1. Feedback Mechanism:

Define a feedback function $F_{feedback}$ that evaluates the accuracy of predictions.

The function compares predicted outcomes with actual observed data.

2. Prediction Accuracy Measurement:

Let $P_{predicted}(h, d)$ be the predicted crowd dynamics (density, movement patterns) for hour h and day d.

Let $P_{actual}(h, d)$ represent the actual observed data for the same.

Define an accuracy metric $\mathrm{Acc}\,(\mathrm{h}, \mathrm{d})$ to measure how closely. $P_{\text{predicted}}$ aligns with P_{actual} :

$$\mathrm{Acc}\,(\mathrm{h}, \mathrm{d}) = \mathrm{function}\left(P_{\text{predicted}}\,(\mathrm{h}, \mathrm{d}), P_{\text{actual}}\,(\mathrm{h}, \mathrm{d})\right)$$

Eq: (4-28)

This function could be based on statistical measures like mean squared error, correlation coefficients, or other relevant metrics.

3 Model Parameter Adjustment:

Based on $\mathrm{Acc}\,(\mathrm{h}, \mathrm{d})$, adjust the model's internal parameters to improve future predictions.

Define a parameter update function. F_{update} that modifies the model parameters based on the feedback:

New Parameters $= F_{\text{update}}$ (Current Parameters, $\mathrm{Acc}\,(\mathrm{h}, \mathrm{d})$)

This update function involves recalibrating thresholds, adjusting weights in predictive algorithms, or modifying the criteria used in social tie analysis.

- Iterative Process:

Implement this as an iterative process, where after each prediction and subsequent observation, F_{feedback} and F_{update} are applied to refine the model.

The Enhanced DUCSIM algorithm leverages complex mathematical formulations to provide a multi-faceted analysis of urban crowd dynamics and individual social interactions. Its depth in both macroscopic and microscopic analysis positions it as a potent tool for urban planning and sociological research, enabling comprehensive understanding and prediction of urban crowd behaviour.

B. Diagrammatic Presentations

Figure 4-14 presents the flow chart signifies an organized procedure of launching points as well as Mobile stations, performs macro-based crowd density assessment, accumulated crowd movement and moves on to with respect to microscopic mobility and inter-personal interactions. This combines various analyses including those based on historical data and predictive modelling to make improvements in different models and anticipate crowd movements resulting in outputting crowd density metrics, mobility profiles, and predictive insights.

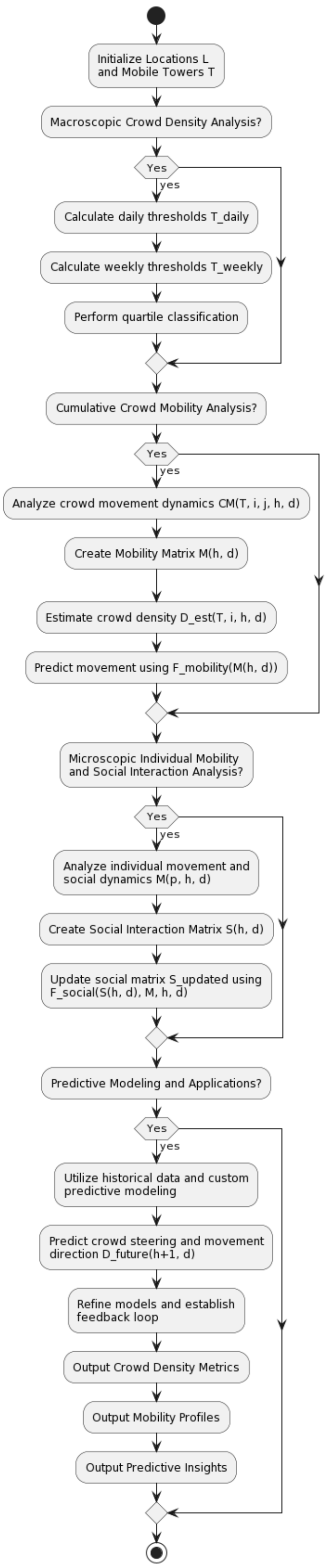


Figure 4-14: Enhanced DUCSIM with Cumulative Crowd Mobility Analysis

The Enhanced DUCSIM Algorithm 4-11 is designed for advanced urban crowd dynamics and social interaction analysis, integrates intricate mathematical methodologies to analyse crowd patterns and individual behaviours using MOBILE tower data.

Algorithm 4-11: Comprehensive Mobility and Social Interaction Model with Enhanced DUCSIM

Input

1. Locations and MOBILE Towers:
   Locations : $L = \{L_1, L_2, \ldots, L_n\}$.
   MOBILE Towers: $T = \{T_1, T_2, \ldots, T_n\}$.
2. Crowd Counts Data:
   Raw Crowd Count: $C_{raw}(T_i, h, d)$ for each tower T_i, hour h, and day d.

Process

1. Macroscopic Crowd Density Analysis:
   Daily Threshold Calculation:
   $T_{daily}(d, h) = \text{median}\{C_{raw}(T_i, 1, d), \ldots, C_{raw}(T_i, h, d)\} \forall T_i \in T$

   Weekly Threshold Calculation:
   $T_{weekly}(h) = \text{median}\{C_{raw}(T_i, h, 1), \ldots, C_{raw}(T_i, h, 7)\} \forall T_i \in T$

   Quartile Classification based on $C_{raw}(T_i, h, d)$ distributions.

2. Cumulative Crowd Mobility Analysis:
   Crowd Movement Dynamics: $CM(T_i, T_j, h, d)$ indicating crowd movement from T_i to T_j During hour h on day d.
   Mobility Matrix: $M_{h,d} = [m_{ij}]$ where $m_{ij} = CM(T_i, T_j, h, d)$.
   Density Estimation and Movement Prediction:
   Density Estimation : $D_{est}(T_i, h, d) = \sum_j m_{ij} - \sum_j m_{ji}$.
   Movement Prediction Model: $F_{mobility}(M_{h,d}) \rightarrow M_{h+1,d}$.

3. Microscopic Individual Mobility and Social Interaction Analysis:
   Individual Movement and Social Dynamics: $M(p, h, d) \rightarrow T_i$.
   Social Interaction Matrix: $S_{h,d} = [s_{pq}]$ where s_{pq} Measures interaction strength between individuals p and q.
   Inference of Social and Random Relationships:
   Social Matrix Update: $S_{updated} = F_{social}(S_{h,d}, M, h, d)$.

4. Predictive Modelling and Applications:
   Historical Data Utilization and Custom Predictive Modelling.
   Crowd Steering and Movement Direction Prediction: $D_{future}(h + 1, d) = F_{direction}(CM_{current}, CM_{historical})$.
   Model Refinement and Feedback Loop.

Output

1. **Crowd Density Metrics**: Calculated daily and weekly thresholds and quartile classifications.
2. **Mobility Profiles**: Individual movement patterns and social interaction data.
3. **Predictive Insights**: Forecasts of future crowd dynamics and social interactions.

The Enhanced DUCSIM algorithm, with its comprehensive mathematical foundation, offers a profound capability to analyse and predict complex urban crowd dynamics and social interactions. Meticulously processing MOBILE data provides actionable insights for effective urban planning and sociological research.

However, building on the complexity of Algorithm 4-11, Algorithm 4-12 constitutes another milestone for crowd analyses and control systems. Cumulative crowd mobility analysis, predictive modelling, and social interaction analysis were all built into algorithm 4-11 and offered a holistic view of urban crowd dynamics. Algorithm 4-12, on the other hand, provides for dynamic self-learning and adapting along with tailored predictive modelling that takes crowd management a step above exactness and adaptability. In this case, it uses emerging crowd data to update model parameters in real-time, thus increasing its adaptability to the dynamic environment. In addition, Algorithm 4-12 draws on past information to develop the predictor functions that will enhance future predictions of crowd behaviour and society relations. This transformation from Algorithm 4-11 to Algorithm 4-12 marks a shift from static (analysis) to dynamic self-improving algorithms, helping the decision-makers with up-to-date information (in the current situation in the city) and evolving crowd control techniques so that the town can be safer.

4.8.7 Proposed Algorithm 7: Dynamic Urban Crowd and Social Interaction Model (DUCSIM) With Median of Median Threshold

DUCSIM’s next Algorithm 4-12 that analyses and forecasts urban crowds' dynamic behaviour and social relationships. DUCSIM uses MOBILE tower data, timestamping, and crowd history to plot comprehensive movement and urban socio patterns. The proposed model is complex, considering macroscopic crowd density statistics, accumulated crowd mobility dynamics, and microscopic analysis of individual actions and social linkages. Dynamic self-learning and adaptive predictive modelling enhanced DUCSIM, allowing a comprehensive understanding of urban crowd behaviour necessary for planning event management and ensuring public safety strategies.

A. Mathematical Presentation

- Macroscopic Crowd Density Analysis

This stage entails the assessment of population distribution in various zones over time. It calculates the raw crowd count at each MOBILE tower and uses a Median-of-Medians (M-o-M) approach to establish daily and weekly crowd density thresholds, and are presented as:

Raw Crowd Count: $C_{raw}(T_i, h, d)$ for each tower T_i, hour h, and day d.

Daily Threshold:

$$T_{daily}(d, h) = M - o - M\left(\{C_{raw}(T_i, 1, d), \ldots, C_{raw}(T_i, h, d)\}\right)$$

Eq: (4-29)

Weekly Threshold:

$$T_{weekly}(h) = M - o - M\left(\{C_{raw}(T_i, h, 1), \ldots, C_{raw}(T_i, h, 7)\}\right)$$

Eq: (4-30)

- Cumulative Crowd Mobility Analysis

This module is about understanding crowd dynamics for the movements from one place to another at various times. It builds a mobility matrix that estimates the net flow of people from one BTS to another, defining a function for tracking the movement.

Crowd Movement Function: $CM\left(T_i, T_j, h, d\right)$.

Mobility Matrix:

$$M_{(h,d)} = \left[m_{ij}\right]$$

Eq: (4-31)

where m_{ij} represents the movement from T_i to T_j.

Estimated Density: $D_{est}(T_i, h, d)$.

- Microscopic Individual Mobility and Social Interaction Analysis

The next stage involves movement analysis as well as interaction in social matters. It monitors people’s moves between towers, thereby developing an interaction matrix constantly updated according to changing social relationships.

Individual Movements: $M(p, h, d) \rightarrow T_i$.

Social Interaction Matrix :

$$S_{(h,d)} = [s_{pq}]$$

Eq: (4-32)

Updated Social Matrix: $S_{Updated.}$

- **Dynamic Self-Learning and Adaptation**

The model adjusts in this stage by incorporating the newly obtained information into its parameters for more accurate projections.

Update Function:

$F_{update}\left(C_{raw}, M_{(h,d)}, S_{(h,d)}\right) \rightarrow$ New Parameters.

Eq: (4-33)

5 Custom Predictive Modelling

In this last step, the historical record is fused into a special predicting function to aid in projecting the future crowd dynamics and people interaction.

Predictive Function:

$P_{future}(h + 1, d) = F_{predict}\left(T_{current}, F_{pattern}(H)\right)$

Eq: (4-34)

Collectively, these modules form a complete urban crowd dynamics analysis and prediction system.

B. Diagrammatic Presentation

Figure 4-15 illustrates of an entire procedure of determining crowd dynamics analysis and prediction is from Mobile tower data initiation, timeline, as well as previous crowds data. Beginning with macroscopic density analysis of a group of people to microscopically analyzing their movements and interactions. The algorithm 4-12 learns from prior counts and makes customized predictive models. This system produces the prediction of future crowd dynamics and social interactions.

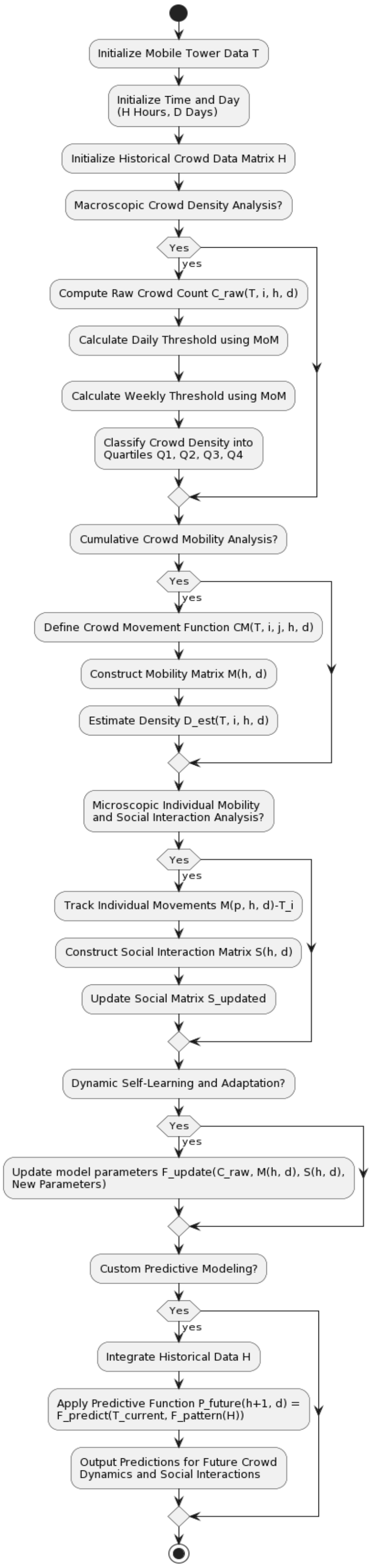


Figure 4-15: Adaptive Learning and Customized Predictive Analytics With DUCSIM-TM-o-M

Algorithm 4-12: Adaptive Learning and Customized Predictive Analytics With DUCSIM-$T_{M\text{-}o\text{-}M}$

Input

MOBILE Tower Data: $\mathcal{T} = \{T_1, T_2, \dots, T_n\}$
Time and Day: $h \in$ Hours, $d \in$ Days
Historical Crowd Data Matrix: **H**

Process

1 Macroscopic Crowd Density Analysis:
Compute Raw Crowd Count $C_{raw}(T_i, h, d)$ for each tower T_i, hour h, and day d.
Calculate Daily Threshold using M-o-M:
$T_{daily}(d, h) = M - o - M(\{C_{raw}(T_i, 1, d), \dots, C_{raw}(T_i, h, d)\})$.
Calculate Weekly Threshold using M-o-M:
$T_{weekly}(h) = M - o - M(\{C_{raw}(T_i, h, 1), \dots, C_{raw}(T_i, h, 7)\})$.
Classify Crowd Density into Quartiles $Q1, Q2, Q3, Q4$ based on C_{raw}.

2 Cumulative Crowd Mobility Analysis:
Define Crowd Movement Function $CM(T_i, T_j, h, d)$.
Construct Mobility Matrix $\mathbf{M}_{h,d} = [m_{ij}]$ where m_{ij} Is the movement from T_i to T_j.
Estimate Density $D_{est}(T_i, h, d)$ as the net flow of crowds.

3 Microscopic Individual Mobility and Social Interaction Analysis:
Track Individual Movements $M(p, h, d) \to T_i$.
Construct a Social Interaction Matrix $\mathbf{S}_{h,d} = [s_{pq}]$.
Update the Social Matrix to reflect evolving ties. $\mathbf{S}_{updated}$.

4 Dynamic Self-Learning and Adaptation:
Update model parameters: $F_{update}(C_{raw}, \mathbf{M}_{h,d}, \mathbf{S}_{h,d}) \to$ New Parameters.

5 Custom Predictive Modelling:
Integrate Historical Data **H**.
Apply Predictive Function: $P_{future}(h + 1, d) = F_{predict}(T_{current}, F_{pattern}(\mathbf{H}))$.

Output

Predictions for Future Crowd Dynamics and Social Interactions

DUCSIM's advanced Algorithm 4-12 is an important development in terms of urban dynamics and the modelling of social interactions. The complexity of data analysis, the movement tracker, and the predictive approach give a deeper understanding of crowd behaviour and social networking in urban areas. This ability to adjust to real-time data enhances the model's performance and improves prediction capability, offering a dependable model for crowd control in urban areas. The application of smart technology might facilitate city planning, reduce crowd movement, and thus make the towns more responsible.

4.9 Evolution stages of DUCSIM Algorithm

To present the evolution of the Enhanced DUCSIM algorithm in a more comprehensive and structured manner, a table format can effectively highlight the progression and key additions at each stage. This format facilitates a clear understanding of how the algorithm has been refined and expanded.

4.9.1 Algorithm 4-6: Multivariate Crowd Density Analytical Framework (CD_AF)

Formulation: Algorithm 4-6 established the groundwork for a comprehensive crowd density analysis, CD_AF. It incorporated both temporal threshold calculations. $\left(T_{daily}, T_{weekly}\right)$ And quartile classification mechanisms (Q1, Q2, Q3, Q4), thus offering a multidimensional perspective of crowd dynamics. The mathematical representation can be expressed as:

$$CD_AF = f\left(T_{daily}, T_{weekly}, Q\right)$$

Eq: (4-35)

where Q represents quartile-based categorization.

4.9.2 Algorithm 4-7: Proactive Crowd Management Paradigm (PCMP)

Advancement: Extending Algorithm 4-6 to Algorithm 4-7 signified a paradigm shift to proactive crowd management, labeled PCMP. This algorithm integrated historical data (H) and quartile classifications (Q) into a real-time decision-making matrix. The functional representation is:

$$PCMP = g(H, Q) \rightarrow \text{Real-Time Decisions}$$

Eq: (4-36)

where g is the decision-making function based on historical and quartile data.

4.9.3 Algorithm 4-8: Streamlined Crowd Density Taxonomy (SCDT) (Median Threshold)

Simplification: Algorithm 4-8 refined the crowd density classification process, focusing on hourly median thresholds (T_{median}). This simplification, termed SCDT, facilitated rapid categorization of crowd densities, essential for immediate applications. The formulaic expression is:

$$\mathrm{SCDT} = \mathrm{h}(\mathrm{T}_{\mathrm{median}}) \rightarrow \text{Density Categories}$$

Eq: (4-37)

where h represents the classification function based on median thresholds.

4.9.4 Algorithm 4-9: Individual and Social Dynamics Integration (ISDI)

Extension: Algorithm 4-9 introduced the integration of individual mobility patterns (M), group identification (G), and social tie strength (S_tie) metrics. This extension, ISDI, enhanced predictive modelling capabilities for future crowd behaviour, thereby improving the granularity of analysis. The mathematical formulation is:

$$\mathrm{ISDI} = \mathrm{P}(\mathrm{M}, \mathrm{G}, \mathrm{S_t ie}) \rightarrow \text{Behavioural Forecast}$$

Eq: (4-38)

4.9.5 Algorithm 4-10: Dynamic Urban Crowd and Social Interaction Model (DUCSIM)

Integration: Algorithm 4-10 seamlessly blended individual mobility patterns (M) with macroscopic crowd density analysis (C_raw, T_median), forming the DUCSIM framework. This integration facilitated more comprehensive and precise urban crowd dynamic predictions. The representational equation is:

$$\mathrm{DUCSIM} = \mathrm{P}(\mathrm{M}, \mathrm{C_{raw}}, \mathrm{T_{median}}) \rightarrow \text{Urban Forecast}$$

Eq: (4-39)

4.9.6 Algorithm 4-11: Comprehensive Mobility and Social Interaction Model with Enhanced DUCSIM (CMSIM-EDUCSIM)

Enhancement: Algorithm Xl elevated the analysis framework by incorporating cumulative crowd mobility (CM) and dynamic social interaction (S_dyn). Termed CMSIM, this advanced predictive modelling approach offered proactive crowd management solutions and dynamic insights. Its functional representation is:

$$\mathrm{CMSIM} = \mathrm{P}\left(\mathrm{C_{raw}}, \mathrm{CM}, \mathrm{S_{dyn}}\right) \rightarrow \text{Dynamic Forecast}$$

Eq: (4-40)

4.9.7 Algorithm 4-12: Adaptive Learning and Customized Predictive Analytics (ALCPA) With DUCSIM-TM-o-M

Evolution: Representing a significant leap, Algorithm 4-12 introduced a dynamic, self-learning mechanism (A) with real-time parameter updates (U) under the banner of

ALCPA. This paradigm shift enabled the adaptation of models in response to evolving data, thus enhancing accuracy. The mathematical expression is:

$$\text{ALCPA} = P_c \text{ custom } (C_r\text{aw}, A, U, H) \rightarrow \text{Real-Time Insights}$$

Eq: (4-41)

Sequential development presented in the table 4-9 indicates a transformation process that starts with basic crowd density investigation and leads to sophisticated crowd management systems. These improvements will address the requirement for precise measures of urban crowd control, highlighting the need for immediate adjustments, personalized analysis of people's behaviours, and forecasting how best to handle the urban population dynamics.

Table 4-9: Summary of Evolution Stages of DUCSIM Algorithm

Stage	Focus Area	Enhancements and Key Additions
I. Introduction	**Objective Setting**	- Expanded scope for analyzing urban dynamics. - Adaptation to various data sources: WiFi, vehicular networks, and ride-sharing data.
II. Data Integration and Preprocessing	**Data Handling**	- Integration of diverse urban data sources. - Standardization of data formats. - Alignment of spatial and temporal data for consistency.
III. Macroscopic Urban Dynamics Analysis	**Spatial Analysis**	- Definition of spatial nodes based on data sources. - Activity count and threshold computation using the median-of-medians. - Quartile-based classification for varying urban density areas.
IV. Cumulative Mobility Analysis	**Mobility Flow Dynamics**	- The mobility flow function between nodes is introduced. - Construction of flow matrix for movement and density estimation. - Implementation of predictive algorithms for mobility patterns.
V. Microscopic Individual and Group Behaviour Analysis	**Individual & Group Dynamics**	- Analysis of individual/group movement patterns (specific to ride-sharing data). - In-depth study of social interactions and mobility behaviours.

This table showcases the step-by-step enhancements made to the DUCSIM algorithm. Each stage marks a significant development, from broadening its objective scope to the intricate analysis of individual and group behaviour within urban spaces. The evolution reflects the model's increasing sophistication in handling diverse data sources, enhancing spatial analysis capabilities, and deepening the understanding of urban mobility and social dynamics. This progression highlights the algorithm's adaptability and growth and underscores its potential as a comprehensive tool for urban planners and researchers.

4.10 Proposed Algorithm 6: Real-Time Crowd Density and Mobility Prediction (Opportunistic Environment)

Different from the previous crowd analysis algorithms, this algorithm departs from some specific points. The new and improved version of this algorithm goes against the flow of its predecessors with more emphasis on real-time activity counts for particular nodes during specific times and not on historical crowd data derived from cell towers. Such a shift reflects the need for proper response in realistic situations that characterize evolving crowds. The “threshold” function (f_thresh) compares the current activity count to M-o-Ms Medians (T_M-o-M (g,t)). The algorithm has an additional thresholding mechanism that allows it to adjust its predictions by assessing the real-time level of activities against predefined levels. This type of flexibility plays an important role, especially when it comes to crowd dynamics that are constantly changing and, therefore, require quick adaptation on the part of the algorithm.

The algorithm creates a prediction model ($D_{pred\ (g,\ t+1)}$) built on several attributes such as the mean of historical values ($\overline{D_{hist(g)}}$), current values of activity count ($D_{real\ (g,t)}$) and some The predictive model incorporates weighted coefficients (w_{act}, w_{hist}, and w_{thresh}), enabling tuning of the model’s sensitivity to different circumstances and creating customized predictions for a particular case. The suggested algorithm is different because it targets the live data using a threshold function and the weighted predictive approach. The innovation enables it to project dynamic and adaptable crowd density, thus providing vital assistance in traffic management systems, emergency evacuation plans, or organizing events. Unlike the historical empirical approaches, this innovation provides the option of more accurate crowd control in dynamic circumstances.

The novel approach to predict urban crowd density and movement patterns. This model integrates real-time activity data with historical crowd density records, utilizing a Median-of-Median threshold. The methodology hinges on correlating current activities with past trends to forecast future crowd dynamics, providing a valuable urban planning and management tool. This section delineates the construction of an innovative model designed to forecast urban crowd densities. The model ingeniously synthesizes real-time activity measures with historical density patterns, employing the Median-of-Medians (M-o-M) threshold as a pivotal determinant.

[1]. Mathematical Presentations

Defining Key Variables:

Represent the real-time activity count at node g and time t as $A_{raw}(g, t)$.

Historical crowd density at node g and time t is denoted by $D_{hist}(g, t)$.

The M-o-M threshold for node g at time t is expressed as $T_{M-o-M}(g, t)$.

Future crowd density predictions are indicated by $D_{pred}(g, t + 1)$.

Threshold Function

The analytical cornerstone of our model is the binary threshold function, designated as f_{thresh}. This function is crucial for interpreting the current activity data concerning established thresholds. The function is defined mathematically as follows:

$$f_{thresh}(A_{raw}(g, t), T_{M\text{-}o\text{-}M}(g, t)) = \begin{cases} 1 & \text{if } A_{raw}(g, t) > T_{M-o-M}(g, t) \\ 0 & \text{otherwise} \end{cases}$$

Eq: (4-42)

$A_{raw}(g, t)$ represents the observed activity count at a given node g and time t, while $T_{M-o-M}(g, t)$ stands for the Median-of-Medians threshold applicable to that node and time. The function outputs a value of 1 if the current activity count surpasses the threshold, indicating a higher crowd density state. Conversely, a value of 0 is returned if the activity count falls below the threshold, signifying a lower-density state.

This binary threshold function is pivotal in the proposed model, serving as a fundamental determinant in the subsequent predictive analysis. It allows for a nuanced differentiation of crowd density states based on real-time urban activity data, thus enhancing the precision of our predictive modelling.

- Model Formulation:

  The predictive model is articulated as follows:

$$D_{pred}(g, t+1) = w_{act} \cdot A_{raw}(g, t) + w_{hist} \cdot \overline{D_{hist}(g)} + w_{thresh}$$
$$f_{thresh}(A_{raw}(g, t), T_{M\text{-}o\text{-}M}(g, t))$$

Eq: (4-43)

In this equation, w_{act}, w_{hist}, and w_{thresh} represent the weights assigned to current activity, the average of historical density, and the threshold function, respectively.

[2]. Diagrammatic Presentation

This figure 4-16 indicates the procedure for forecasting crowd density involving the commencement of input initiation and then threshold assessment. Where the data crosses a specific threshold, a predictor model uses raw data, weighted data, and historical data and then predicts crowd density; however, where the data does not exceed a specific threshold, crowd density is assumed as zero. Then the result of either the developed model or setting it to null results is served as the predicted crowd density.

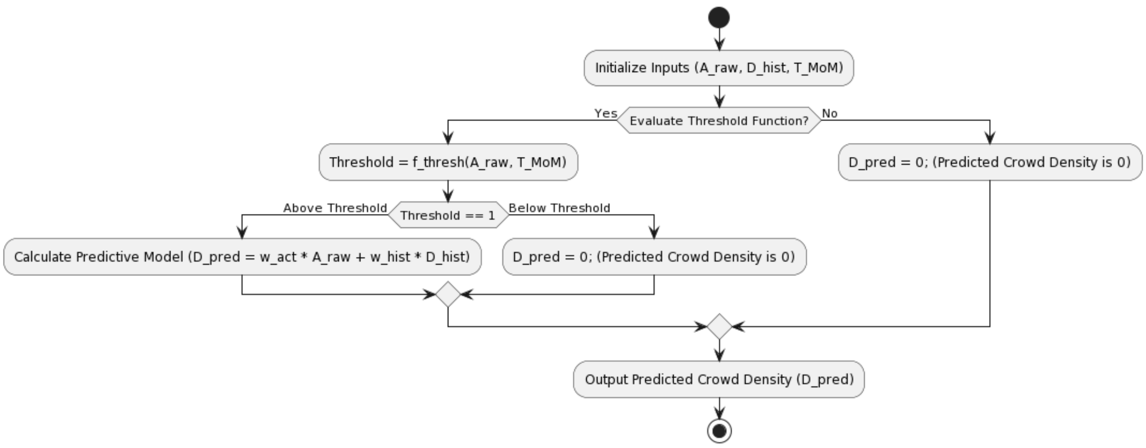


Figure 4-16: Crowd Density and Movement Prediction

Data Procurement: Aggregate data for $A_{raw}(g, t)$, $T_{M-o-M}(g, t)$, and $D_{hist}(g, t)$. Threshold Determination: Implement f_{thresh} to compare current activity with the M-o-M threshold. Model Execution: Employ the formulated equation to estimate $D_{pred}(g, t+1)$. Accuracy Evaluation: Examine the model's precision, adjusting weights as necessary.

Algorithm 4-13: Advanced DUCSIM Integrative Predictive Model (ADIPM)

Inputs

1 Real-Time Activity Count:

$A_{raw}(g,t)$: Activity count at node g and time t.

2 Historical Crowd Density:

$D_{hist}(g,t)$: Historical density at node g and time t.

3 Median-of-Medians (M-o-M) Threshold:

$T_{M-o-M}(g,t)$: Threshold value for node g at time t.

Process

1 Threshold Function (f_{thresh}) :

Defined as $f_{thresh}(A_{raw}(g,t), T_{M-o-M}(g,t)) = \begin{cases} 1 & \text{if } A_{raw}(g,t) > T_{M-o-M}(g,t) \\ 0 & \text{otherwise} \end{cases}.$

2 Model Formulation:

The predictive model is expressed as:

$D_{pred}(g,t+1) = w_{act} \cdot A_{raw}(g,t) + w_{hist} \cdot \overline{D_{hist}(g)} + w_{thresh} \cdot f_{thresh}(A_{raw}(g,t), T_{M-o-M}(g,t)),$

where $w_{act}, w_{hist}, w_{thresh}$ are weights assigned to current activity, average historical density, and the threshold function, respectively.

Output

Future Crowd Density Prediction:

$D_{pred}(g,t+1)$: The predicted crowd density at node g for the next interval $(t+1)$.

The predictive model, which is integrated with the Algorithm, has results discussed along with the results of Algorithm 14, which has demonstrated promising results in forecasting urban crowd density, highlighting its potential in smart city applications. The model offers a nuanced understanding of crowd behaviour by effectively blending historical data with current urban dynamics. This innovation paves the way for more informed urban decision-making, potentially enhancing the efficiency of city management and emergency response strategies. The developed predictive model is a crucial advancement in understanding and anticipating the complexities of urban crowd movements. Integrating current activity data with historical trends and threshold assessments offers insightful foresight into urban crowd behaviours, which is instrumental for strategic urban planning and management. The model's refinement and empirical validation will further its applicability across urban settings.

4.11 Proposed Algorithm 7: Self-Learning Crowd Density Estimation Algorithm with Median-of-Median Threshold (Integrated Version)

The Algorithm 4-14 effectively sorts and evaluates information to forecast future crowd densities depending on previous patterns and actual numbers. It uses sophisticated statistical techniques such as M-o-M thresholding, count binning, and quartiles. The algorithm can adopt different data trends, therefore making forecasts. Therefore, such a tool will be helpful in urban planning and governance.

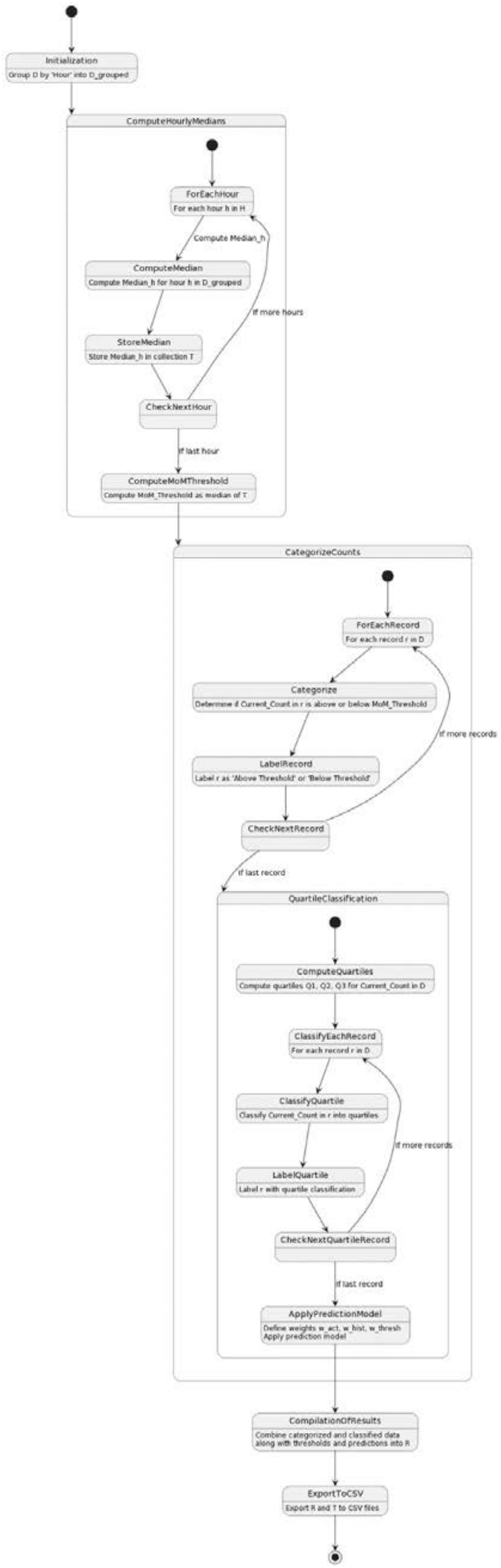


Figure 4-17: Flow of Self-Learning Crowd Density Estimation Algorithm with Median-of-Median Threshold (Integrated Version)

Algorithm 4-14: Self-Learning Crowd Density Estimation Algorithm with Median-of-Median Threshold (Integrated Version)

Inputs:

D: Dataset containing records with fields 'Hour,' 'Current_Count,' and other relevant information.

H: Set of unique hours in D.

Outputs:

R: Resultant dataset with classifications, computed thresholds, and predictions.

Procedure:

1 **Initialization:**

Let $D_{grouped}$ be a dataset grouped by 'Hour' from D.

2 **Compute Hourly Median Thresholds and Median-of-Medians (M-o-M) Thresholds:**

For each hour $h \in H$:

Compute Median_h as the median of 'Current_Count' for hour h in $D_{grouped}$.

Store Median h in a collection T.

Compute M-o-M_Threshold as the median of the hourly medians in T.

3 **Categorization of Counts Based on M-o-M Threshold:**

For each record r in D :

Determine if 'Current_Count' in r is above or below the M-o-M_Threshold.

Label r as 'Above Threshold' or 'Below Threshold' based on the comparison.

4 **Quartile Classification:**

Compute overall quartiles $Q1, Q2, Q3$ for 'Current_Count' in D.

For each record r in D :

Classify 'Current_Count' in r into quartiles based on $Q1, Q2, Q3$.

Label r with the corresponding quartile classification.

5 **Prediction Model:**

Define weights $w_{act}, w_{hist}, w_{thresh}$ for current activity, historical density, and threshold function.

For each record r in D, apply the prediction model: $D_{pred}(r) = w_{act} \cdot \text{Current\_Count}(r) + w_{hist} \cdot \text{Reference\_Count}(r) + w_{thresh} \cdot f_{thresh}(\text{Current\_Count}(r), \text{M-o-M\_Threshold})$ where $f_{thresh}(x, y) = 1$ if $x > y$ else 0 .

6 **Compilation of Results:**

Combine the categorized and classified data along with computed thresholds and predictions into R.

7 **Export to CSV:**

Export R and T to separate CSV files.

End of Algorithm

The Algorithm 4-14 combines complicated statistical calculations with a predictive model and provides accurate crowd-density forecasting. It provides an organized procedure for processing the hourly crowds through categorizing counts and using a definite prediction model, making it possible to develop a holistic approach to forecasting crowd behaviour. It is adaptable and accurate, thus making it a sturdy framework that can be used for crowd analysis with future crowd density estimation in different cases.

4.12 Algorithm 8: Heterogeneous Opportunistic DUCSIM with Enhanced Crowd Density Analysis

The Enhanced DUCSIM Algorithm 4-15 is a sophisticated tool designed to analyze and predict the complexities of urban dynamics. It harnesses diverse data sources such as WiFi networks, vehicular traffic, ride-sharing activities, and social media interactions, encapsulating the pulse of urban life. By integrating these data streams, represented mathematically as $D(g, t)$, the algorithm creates a multi-dimensional snapshot of urban activity. It delves into macroscopic analyses of crowd density and movement patterns and microscopic examinations of social interactions, synthesizing these insights into predictive models. Enhanced DUCSIM stands at the forefront of urban analysis, offering a comprehensive lens to view and interpret the intricate web of urban dynamics. The figure 4-18 illustrates the flow of the algorithm execution.

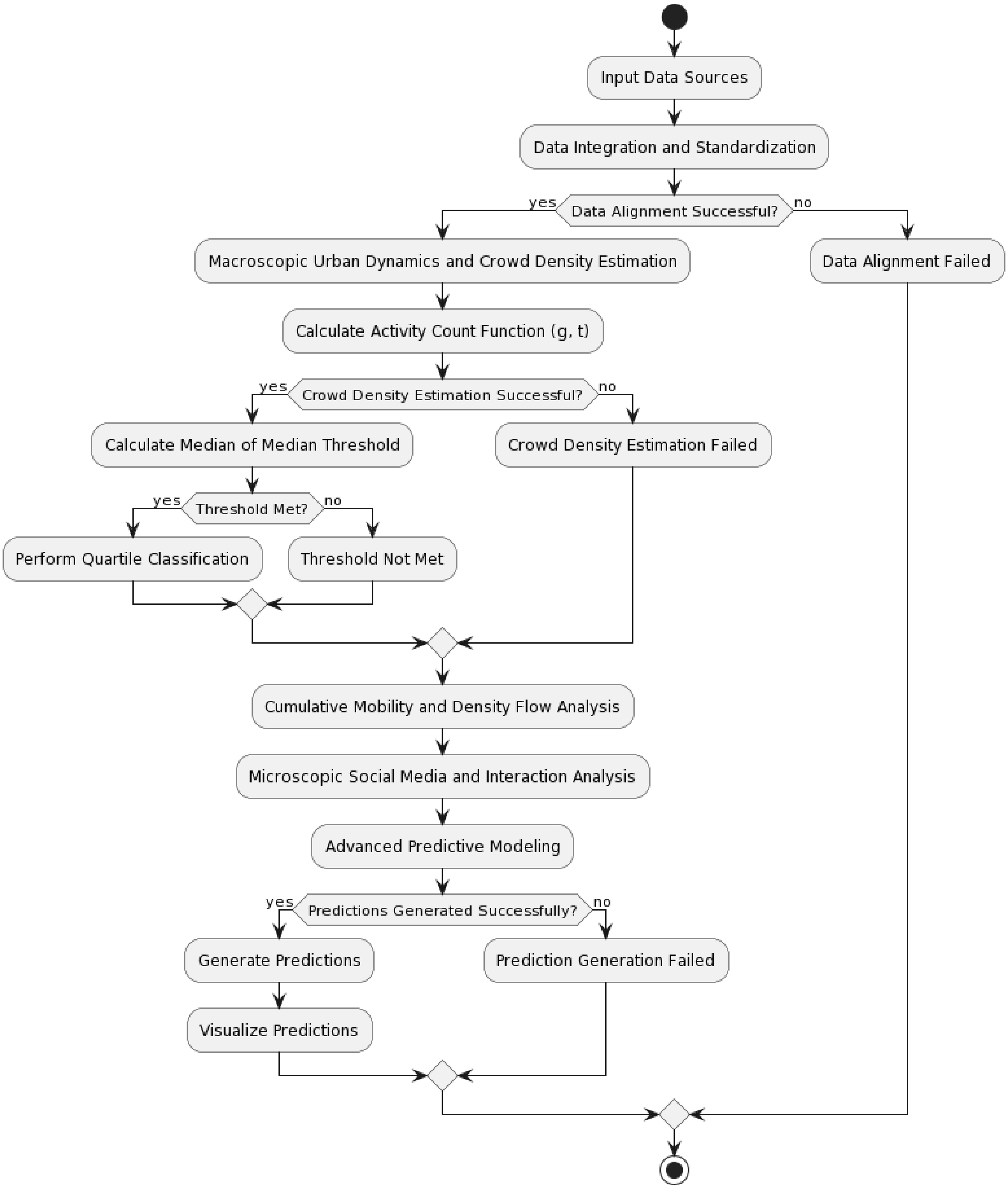


Figure 4-18: Flow of Heterogeneous Opportunistic DUCSIM with Enhanced Crowd Density Analysis

Algorithm 4-15: Heterogeneous Crowd Density Analysis - DUCSIM

INPUT
Data Sources

1 WiFi Networks:

Let $W(g, t)$ represent the data from WiFi networks, indicating user connections at location g at time t.
$W(g, t)$ could be a count or a more complex measurement based on WiFi network usage.

2 Vehicular Networks:

Let $V(g, t)$ denote traffic and vehicular flow data at location g at time t.
$V(g, t)$ might include metrics like vehicle count, average speed, or flow rate.

3 Ride-Sharing Platforms:

Define $R(g, t)$ as the data representing ride-sharing trips starting or ending at location g at time t.
$R(g, t)$ could be a count of trips or other relevant measurements like trip duration.

4 Social Media (Geo-Tagged Tweets):

Use $T(g, t)$ to represent the number of geotagged tweets from location g at time t.
Direct count of tweets.

5 Sentiment Analysis and Keyword Density:

Let $S(g, t)$ represent the aggregate sentiment score from tweets at location g at time t.
Define $K(g, t, \text{keyword})$ as the density of specific keywords in tweets at location g at time t.

Integrating Data Sources:
The data from these sources can be integrated to create a comprehensive urban dynamics dataset:
Integrated Data Representation:
$D(g, t) = \{W(g, t), V(g, t), R(g, t), T(g, t), S(g, t), K(g, t, \text{keyword})\}$.
$D(g, t)$ represents a multi-dimensional data point for location g at time t, incorporating various aspects of urban dynamics.

PROCESS:

1 Data Integration and Standardization

Integration Function:
Integrate (W, V, R, T, S, K) where W, V, R, T, S, K represent data from WiFi, vehicular networks, ride-sharing, tweets, sentiment, and keyword density, respectively.
Alignment of Data:
Align spatial (g) and temporal (t) data across all sources.

2 Macroscopic Urban Dynamics and Crowd Density Estimation

Spatial Nodes Definition:
Define nodes $N = \{n_1, n_2, \dots, n_m\}$ Based on integrated data locations.
Activity Count Function:
$A_{raw}(g, t)$: Compute total activity for each node g at time t.
Crowd Density Estimation:

$$T_{M-o-M}(g, t) = \text{median}\{A_{raw}(g, t_1), \dots, A_{raw}(g, t_n)\}.$$

Quartile Classification:
Determine quartiles $Q1, Q2, Q3, Q4$ based on $A_{raw}(g, t)$ distribution.

3 Cumulative Mobility and Density Flow Analysis

Flow Functions:
$CM(g_1, g_2, t)$ = Function representing movement and density flow from g_1 to g_2 at time t.
Flow Matrix Construction:
$M_t = [m_{ij}]$ where $m_{ij} = CM(g_i, g_j, t)$ for nodes g_i, g_j.

4 Microscopic Social Media and Interaction Analysis

Behaviour Analysis Function:
Analyze behaviours using ride-sharing and social media data.
Interaction and Sentiment Matrices:
Construct matrices I and S for social interactions and sentiment.

5 Advanced Predictive Modelling

Predictive Function:
$P_{urban}(g, t+1) = F_{predict}(M_t, T_{M\text{-}o\text{-}M}, T_{sentiment}, K_{density})$.
This function predicts urban dynamics for the next time interval.

OUTPUT

Predictions of Urban Dynamics

1 Crowd Density Prediction:

$D_{predicted}(g, t)$: Represents the predicted crowd density at location g and time t.
Calculated using the predictive function from the algorithm:
$D_{predicted}(g, t) = P_{urban}(g, t)$.

2 Mobility Pattern Prediction:

$M_{predicted}(g_1, g_2, t)$: Denotes the predicted mobility flow from the location g_1 to g_2 at time t.
Derived from the mobility and density flow analysis within the algorithm.

3 Social Interaction Prediction:

$S_{predicted}(g, t)$: Indicates predicted social interaction patterns at location g and time t.
Based on the analysis of interaction and sentiment matrices.

VISUALIZATION

Define a set of functions V that transform the predictions into visualizable data.
For: $V_{density}(D_{predicted})$, $V_{mobility}(M_{predicted})$, and $V_{social}(S_{predicted})$ It could be functions to visualize respective predictions.

The Enhanced DUCSIM algorithm culminates in a powerful predictive model capable of forecasting urban dynamics with a novel level of precision and depth. It offers predictions on crowd density by transforming complex urban data into actionable insights. $(D_{predicted}(g, t))$, mobility patterns $(M_{predicted}(g_1, g_2, t))$, and social interactions

($S_{predicted}(g, t)$). These predictions, visualized through a series of functions V, provide invaluable guidance for urban planning and sociological studies. Enhanced DUCSIM, with its advanced analytical capabilities and comprehensive approach, represents a significant stride in understanding and shaping the future of urban landscapes. It's not just an algorithm; it's a blueprint for smarter, more responsive city management.

4.13 Comparison & Validation of DUCSIM Algorithm

Comparing and validating the Enhanced DUCSIM algorithm with the three prominent crowd models presented in table 4-10 - Social Force Models, Continuum Models, and Agent-Based Models - involves assessing each model's methodology, applicability, and effectiveness in various scenarios. This comparison is crucial in a scientific context to understand the strengths and limitations of DUCSIM concerning established models.

Table 4-10: Comparision & Validation of DUCSIM Algorithm

Feature/Model	Enhanced DUCSIM	Social Force Models	Continuum Models	Agent-Based Models
Core Concept	Integrates macroscopic crowd density analysis with microscopic individual and social interaction analysis.	It focuses on individual behaviours influenced by psychological and physical 'social forces.'	Treats crowds as continuous flows, akin to fluids, with less emphasis on individual behaviour.	Simulates crowds through the interactions of autonomous agents with decision-making capabilities.
Data Utilization	It uses real-world MOBILE tower data for tracking individual movements and interactions.	Does not typically incorporate real-world tracking data.	Generally abstracts from real-world data, focusing on crowd flow patterns.	It can use various data types but emphasizes the rules and behaviours assigned to agents.
Analytical Approach	Employs sophisticated mathematical and statistical methods for dynamic analysis and prediction.	Utilizes mathematical models to simulate forces affecting individual movement.	Applies principles of fluid dynamics to model crowd movements.	It uses complex algorithms to simulate individual behaviours and interactions.
Strengths	- Comprehensive urban dynamics analysis. - Real-	- Effective in simulating immediate	- Efficient for large-scale, high-density crowd	- Highly adaptable to various scenarios

	world data integration. - Detailed individual and social pattern analysis.	behavioural responses - Good for evacuation and emergency scenarios.	analysis. - Useful where individual behaviours are less distinguishable.	- Capable of simulating diverse behaviours and interactions.
Limitations	- Requires extensive data processing. - Potentially less focused on immediate psychological responses.	- Less focused on macroscopic crowd dynamics. - May not account for the broader urban context.	- Lacks individual granularity. - Not ideal for scenarios requiring detailed behavioural analysis.	- Can be computationally intensive. - Model accuracy depends on the rules and behaviours assigned.
Ideal Applications	- Urban planning and sociological studies. -Comprehensive crowd dynamics analysis.	- Emergency response planning. - Situations requiring behavioural analysis under stress.	- Large events with high-density crowds. - Situations where crowd behaviour resembles fluid flow.	- Studies requiring detailed behavioural simulations. - Scenarios with complex individual interactions.

This comparative table illustrates that while each model has specific strengths and applications, the Enhanced DUCSIM offers a more holistic and data-driven approach, making it particularly suitable for urban analysis requiring macroscopic and microscopic insights. Social Force and Continuum Models provide valuable perspectives for specific crowd behaviours and fluid-like movements. In contrast, Agent-Based Models excel in scenarios demanding detailed behavioural simulations and individual interactions. The choice of model should align with the specific needs and goals of the study or application in question.

4.14 Conclusion

In the concluding section of the proposed methodology chapter, the evolutionary arc of the algorithm suite, spanning from the foundational DUCSIM to the advanced Enhanced DUCSIM Predictive Analysis Framework (EDPAF), is succinctly summarized. This progression encapsulates a strategic development trajectory, beginning with fundamental crowd density analysis and advancing towards sophisticated predictive modelling. The initial algorithms established the groundwork in crowd dynamics, focusing on basic counting and data collection methodologies. Subsequent iterations expanded these capabilities, incorporating elements such as MOBILE tower-based data, quartile classifications, and historical data integration, enhancing predictive accuracy.

A notable advancement is observed with Algorithm 4-6, which introduced a comprehensive approach to crowd density analysis, integrating daily and weekly thresholds and quartile classifications. This framework evolved further with Algorithm 4-7, transitioning towards active crowd management through real-time historical and quartile data decision-making. Continuing this trajectory, Algorithms 4-8 to 4-9 brought methodological enhancements. These included streamlined density classifications, integrating individual mobility patterns with macroscopic analyses, and incorporating cumulative crowd mobility and dynamic social interactions into the predictive models. The Enhanced DUCSIM Predictive Analysis Framework (EDPAF) represents the culmination of this evolutionary process. This advanced algorithm integrates the strengths of its predecessors and introduces dynamic, self-learning mechanisms with real-time parameter adaptations. It represents a significant stride in crowd management technologies, aiming for heightened precision and adaptability in urban crowd dynamics. The proposed algorithms demonstrate a comprehensive and progressive approach toward understanding and managing urban crowd dynamics. Each step in this evolution marks a technological enhancement and reflects a commitment to improving the efficacy and responsiveness of crowd management systems in complex urban settings.

CHAPTER 5: CUMULATIVE CROWD DENSITY RESULT

The chapter empirically examines cumulative crowd density using simulated and real-world data from MOBILE towers. This chapter critically examines characteristics and evaluates the results using various data resources. A multimodal approach is emphasized, which plays a key role in portraying the intricate features in the dynamics of an urban crowd. This starts with a comprehensive analysis of the data. In this stage or process, the researcher examines the distribution, variation, and inherent pattern in the data through their effects on the study results. As a matter of great importance in the assessment, eliminating outliers is considered paramount.] Statistics aside, in real-life crowd tracking, an outlier is not a statistical fluke but rather a very important scenario like a singularly dense gathering or peculiar flows. Concerning those outliers, a strategy is designed to handle them in such a way that brings no bias during the evaluation process to conserve the authenticity of the results. After the datum analysis, emphasis shifts to applying the multi-modal method in calculating the cumulative crowd density. This technique fuses different data sources and analytic forms, utilizing their combined power to comprehend how a crowd moves and behaves. Simulation data aids in validating and improving algorithms in controlled conditions, while actual urban crowd behaviour can be estimated with real-world MOBILE tower data used in India.

The results of the comprehensive analysis are carefully examined. This covers the effectiveness of the applied algorithms and approaches and what it means to do in light of the results obtained. It examines how these findings may improve urban designs, public security, and smart city programs. The chapter takes into account possible constraints and recommends possibilities for further investigation to provide a comprehensive review of the conducted study. This chapter seeks to provide an unambiguous illustration of the findings, providing insights into the complexities of crowd movements and behaviour, thus advancing the urban crowd management models.

5.1 Synthetic Data Details

Developing reliable and authentic datasets is an important factor in the developing domain of crowd density estimation. This analysis is based on a carefully designed synthetic data set, which imitates the nuances of Mobile tower logs and related metadata. This constitutes the starting point in designing and testing highly reliable automatic estimation systems that yield quality results. Crowd density estimation has varied uses, including urban planning, traffic control, emergency response, and event coordination. Reliable data is critical in helping people properly analyze and predict human congregation patterns.

On the other hand, using actual MOBILE data also involves many issues, such as privacy problems, and it is too complicated just based on pure information analysis. To bypass these barriers, synthetic data sets became an imperative instrument. It allows researchers to train their methodologies in a safe space. The table 5-1 below presents the dataset with index notations and a preview. The designed synthetic dataset reflects diverse aspects of actual world data, including regionalization for different areas, locations, sites, and timestamps, as well as crowd density estimates. Each record within the dataset is a composite of the following attributes: r_i, s_i, p_i, id_i, lat_i, long_i, t_i, ref_i, curr_i, and float_i. The characteristics have been merged to show the movement of crowds as if they were being tracked by a set of MOBILEs spread around different locations.

Table 5-1: Real World MOBILE Dataset Example Preview

Region	**SubRegion**	**Pincode**	**SiteID**	**Latitude**	**Longitude**	**Time**	**Ref Count**	**Current Count**	**Floating Count**
r_1	s_1	p_1	id_1	lat_1	$long_1$	t_1	ref_1	$curr_1$	$float_1$
r_2	s_2	p_2	id_2	lat_2	$long_2$	t_2	$ref\ f_2$	$curr_2$	$float_2$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
r_i	s_i	p_i	id_i	lat_i	$long_i$	t_i	ref_i	$curr_i$	$float_i$

This dataset is created to develop a scenario where the robustness and accuracy of crowd density estimation algorithms are validated. This makes it possible for researchers to introduce several scenes that cut across variations associated with time, distribution patterns, and even density changes – critical factors when conducting these estimates. Data is formulated systematically, similar to the records in MOBILE towers, while ensuring that no private or secret information concerning individuals is exposed, thus protecting their integrity and identity. The subsequent section presents an overall description of the dataset's structure, details on how it is formed, and what implications it has concerning modelling models intended to estimate crowds' density. Consequently, such a dataset is hoped to support the verification process toward hypothesis/algorithms with a view toward innovation and actual implementations.

5.2 Nature of Dataset

A dataset is integral and realistic in computational data analysis for the credibility of any following conclusions. The presented dataset was algorithmically constructed to embody the messiness and stochasticity characteristic of actual data, particularly regarding crowd density measurements based on MOBILE tower logs. Contrary to manually invented datasets with defined distributions, this synthetically crafted one is unconcerned with presupposed statistically fixed parameters. The absence of such restraints ensures some uncertainty and natural variance, making detecting the inherent pattern complex yet important. The nature of the dataset is computed by applying descriptive statistical analysis, and distribution functions are applied as mentioned below:

5.2.1 Descriptive Statistics:

The descriptive statis table presented in table 5-2 explains the dataset comprises 6000 records resembling MOBILE tower logs aiming at estimating crowd density. The report consists of ten distinct regimes as provided under the '**Region**' column and further subdivided into 50 regimes in general. The high degree of geographic segmentation permits comprehensive investigations into people's movement patterns within different places. Data for this study also cover numerous postal codes, spanning 422001 – 422050, providing an additional local perspective. The dataset also provides around 250 distinct SiteIDs, symbolizing different data-gathering locations like MOBILE towers. This is crucial because it will form a basis for spatial analysis in crowd density studies. The dataset also offers exact geographic coordinates, while the latitudes vary from 19.80932 to 20.21041, and the longitudes vary between 73.50958 and 73.96233. These coordinates are important in mapping the space and crowd movement/density dynamics. The temporal aspect of the dataset is captured through two columns: 'Time' and 'TimeNumeric.' These data are collected in twenty-four unique hourly segments of time. TimeNumeric shows time in a numeric shape. An average of 690 and a standard deviation of 415,3658 from between zero and thirteen hundred eighty. This range provides an overall time distribution of the data. The dataset also includes critical measures of crowd density: 'Ref_Count,' 'Current_Count,' and 'Floating_Count.' An average of 4511.023 and a standard deviation 1447.026 indicate ‘Ref_Count’ as the benchmarked crowd level. It has an average of 6739.734 and a standard deviation of 2570.006; this means that in a current count, it changes relatively much (real-time crowd density). Knowing about conventional and exceptional crowd densities in different places and times makes it vital.

Table 5-2: Descriptive Statistics For Synthetic Dataset

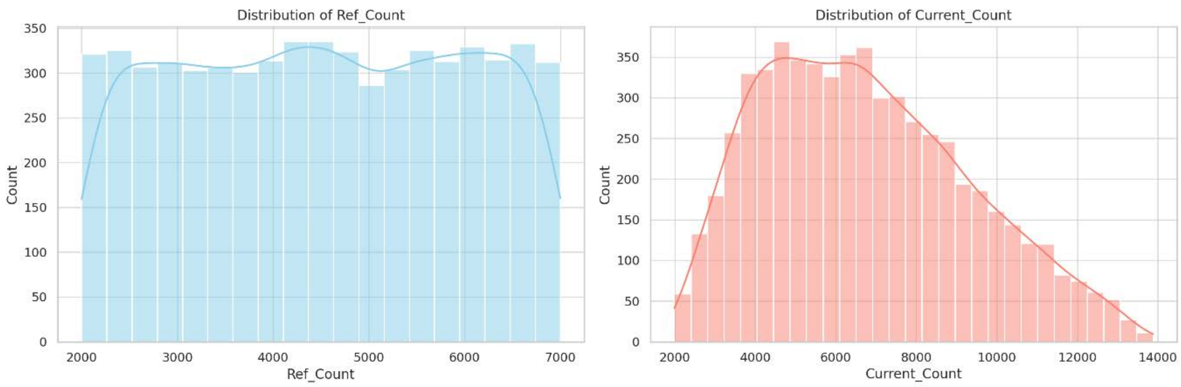
Dataset	**Data Type**	**Non-Null Count**	**Mean**	**Std Deviation**	**Min**	**Max**	**Unique Count**
Region	Object	6000	-	-	-	-	10
SubRegion	Object	6000	-	-	-	-	50
Pincode	Int64	6000	-	-	422001	422050	-
SiteIDs	Object	6000	-	-	-	-	250
Latitude	Float64	6000	-	-	19.80932	20.21041	-
Longitude	Float64	6000	-	-	73.50958	73.96233	-
Time	Object (Hr)	6000	-	-	-	-	24
TimeNumeric	Int64	6000	690	415.3658	0	1380	-
Ref_Count	Int64	6000	4511.023	1447.026	2000	6999	-
Current_Count	Int64	6000	6739.734	2570.006	1988	13881	-

The comprehensive nature of the dataset’s geographic, temporal, and density-related characteristics renders it an effective instrument for precise estimation in crowd density research. It applies a structured methodology for replicating empirical data, making it suitable for designing and validating prediction algorithms relevant to the crowd’s denseness investigations.

5.3 Data Distribution

To learn about the nature of the dataset, we can see the kernel density distribution in the picture below—the Histogram of the ‘Ref_Count’ variable in figure 5-1 (A). The distribution seems uniform when we look at the bar heights at different value ranges. This indicates a relatively low variation in the crowd densities' reference counts across the dataset, with none consistently higher than the rest.

Figure 5-1 (B) is illustrated in the "Current-Count" histogram on the right. It is right-skewed with a peak that implies a modal value on the lower end of the scale. To the right are the long tails showing occasional high current count occurrences. Skew may indicate periods with higher and lower crowd density than average.



(A) (B)

Figure 5-1: Data Distribution for Reference_Count & Current_Count

The figure 5-2 below is the scatter plot, whereby colour and symbols symbolize one of ten regions for each site's latitude and longitude. Clustering of the points implies a wide representation of the data collection locations within certain regions, whereas some others do not feature such clusters. Site diversification is essential in analysing the patterns of crowd density since it shows that the data set included various urban densities covering likely different environments like commercial, residential, or industrial.

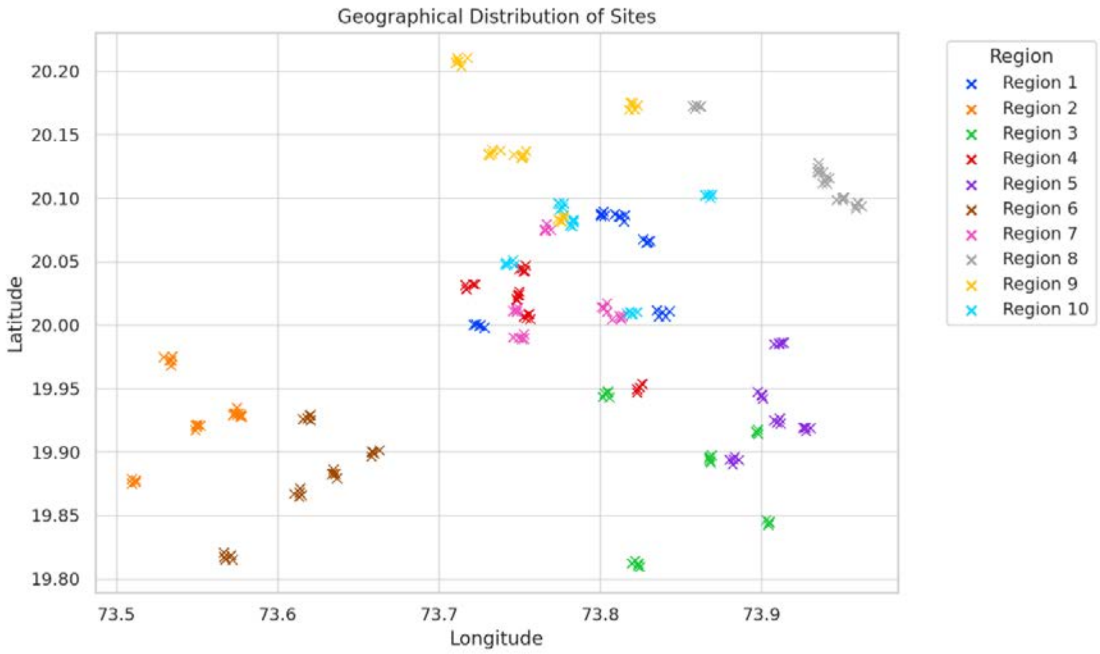
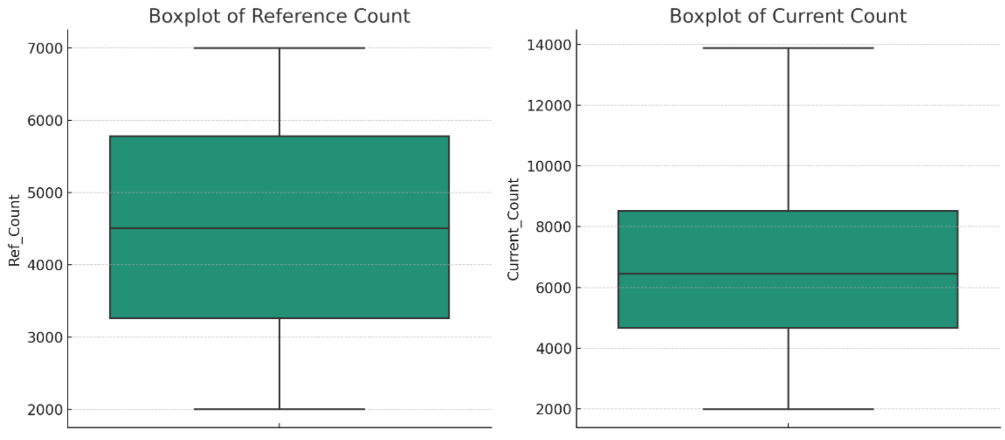


Figure 5-2: Geographical Distribution of MOBILE Tower Sites

These plots show that the design of the dataset was intended to include a broad range of people per square meter (density) as well as geographical locations, a condition that provides appropriate tests against crowd density modelling using typical or actual data.

5.4 Median Threshold: Reference Count & Current Count

Based on the box plot figure 5-3, (A) the median value of the reference count lies around 4500, the middle line of the respective box. In addition, the width of the box – the interval between 4000 and 5000, depicts the inner middle quarter of the numbers. These whiskers are about 2000 – 7000, denoting most non-outlier data. There seems to be no one point outside whiskers, implying a relative constancy of reference without excessive swings. The boxplot Figure 5-3 (B) for the Current Count shows a larger median towards the neighbourhood of 7000, which is also greater than the reference count. The reference count is narrower, with a range similar to about 6000 and 8000 in its middle 50%, revealing a larger distribution for the interquartile range. Whiskers reach around 2000 and above 13999, signifying an increased range of data and, hence, a higher variance in the current count. Such patterns may indicate a day having higher crowds, e.g., during certain hours or at specific events like festivals and parades.



(A) (B)

Figure 5-3: Median Threshold Distribution for Reference_Count & Current_Count

5.5 Time Series of Current Count

Figure 5-4 presents the Crowd density does change over the time series plot of the Current Count at certain times of the day. Accordingly, the graph reveals the count's constant rise and fall, implying rushes during the day at some points or occurrences. Some dips are apparent, especially at around 3,9,21, which may mirror the average nighttime timeslots. The apex of eight, ten, and sixteen suggest possible times of the day with the largest masses.

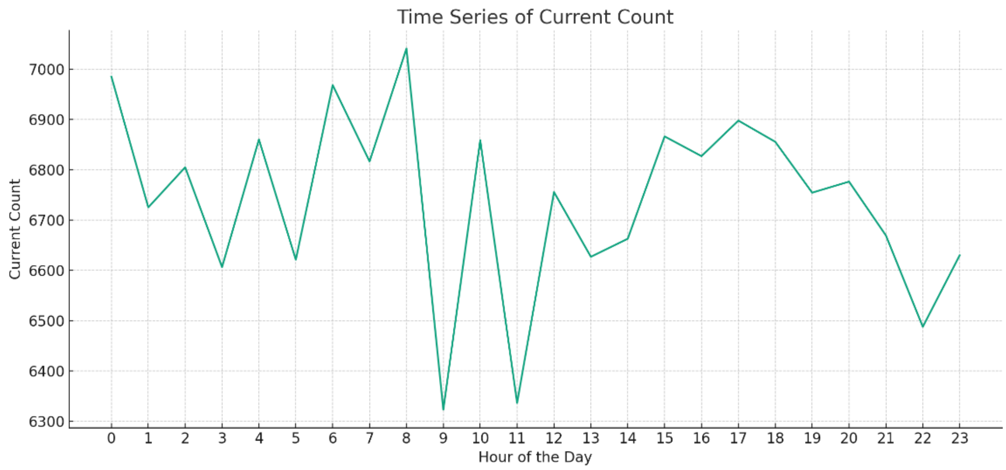


Figure 5-4: Temporal Distribution of Current_Count

The plots indicate that the dataset captures variations in crowd density both over time and across different metrics (Reference vs Current Count). The Reference Count remains relatively stable, whereas the Current Count has greater variability and higher peaks, which are crucial for understanding crowd dynamics. The temporal analysis indicates that the density fluctuates throughout the day, which could be valuable for planning crowd management and infrastructure usage.

5.6 Grouped Histogram

The histograms figure 5-5 below for 'Ref_Count' and 'Current_Count' suggest distinct distributions of crowd density measurements. 'Ref_Count' shows a relatively uniform distribution across its range, with counts evenly distributed, indicating a consistent baseline of crowd density measurements across various sites or times. The 'Current_Count' displays a right-skewed distribution with a concentration of values at the lower end and a tail extending towards higher values, which signals that while most of the crowd density measurements are lower, there are occasional spikes that could correspond to specific events or peak crowd times.

The latitude distribution is multimodal, indicating several peaks where data points are concentrated, which reflects multiple hotspots of geographic activity within the dataset as the dataset is created to mimic a world dataset from Nashik, the district in India, and is focused on a range of 60-90 km in and around the city. The longitude distribution shows multiple peaks, suggesting a non-uniform spatial distribution of the sites within the data, and this is because the dataset is spread sparingly in some locations to cover the highways and railways from about 90km away from the city's centre.

Scatter plots across 'Ref_Count' and 'Current_Count' reveal a linear pattern, indicating a strong positive correlation; higher reference counts tend to be associated with higher current counts. This relationship suggests that areas or times with historically high crowd densities will likely maintain these levels in real-time measurements.

When 'Latitude' and 'Longitude' are plotted against 'Ref_Count' and 'Current_Count,' no discernible pattern indicates any correlation. This lack of a clear trend suggests that crowd density counts are not strongly influenced by the specific latitude or longitude of the sites within the dataset's range. The scatter plot between 'Latitude' and 'Longitude' reveals a dispersed pattern with some clustering, indicating that sites are spread across various geographic locations, with some areas having a higher concentration. The dataset is expected to be concentrated in the city's centre and spared in the outer circle of the city.

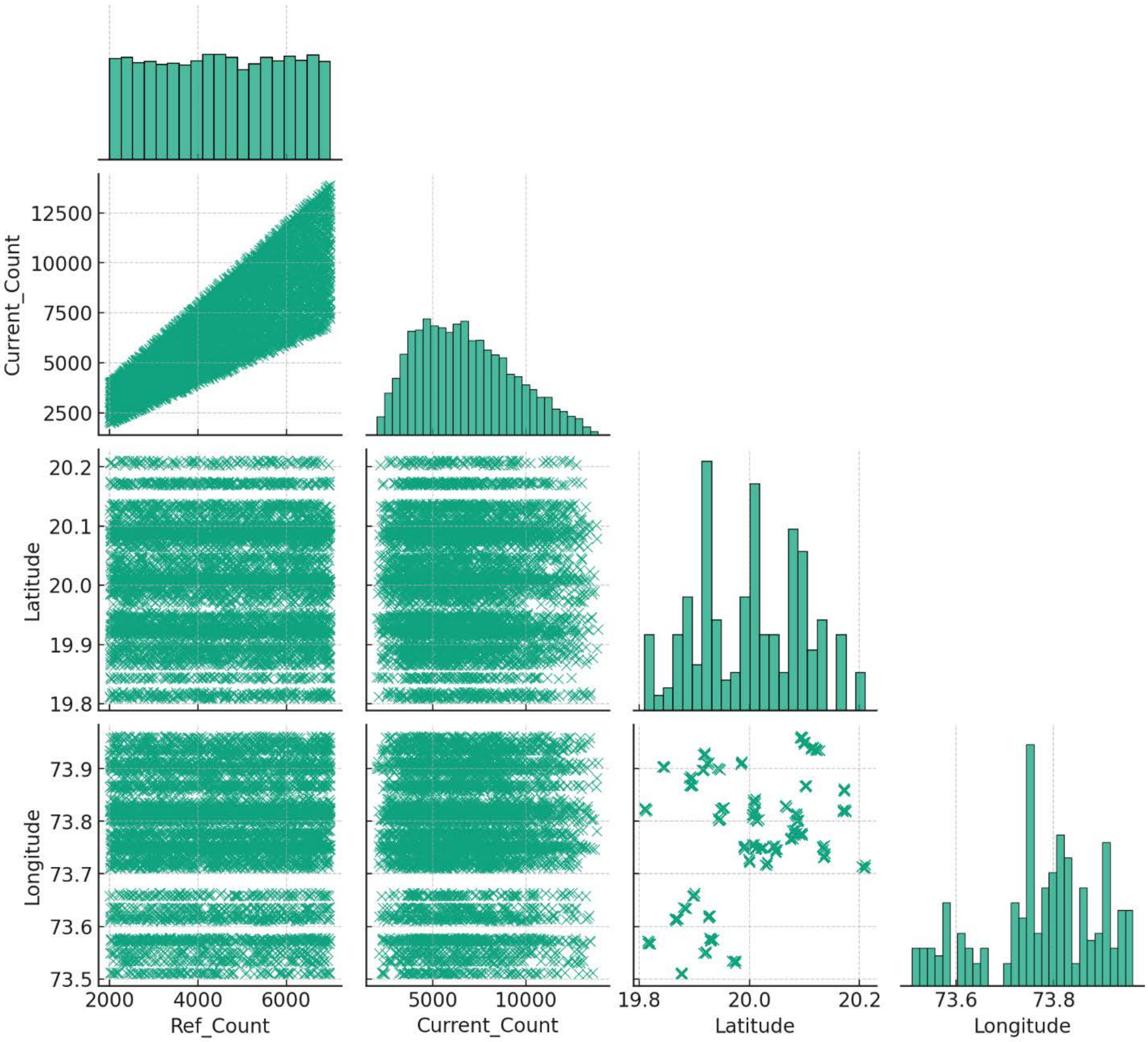


Figure 5-5: Paired Distribution for Key Data Columns

The nature of the dataset, as portrayed by the pair plot, indicates variability in crowd density with certain geographic clustering of data collection points. The strong correlation between the reference and current crowd counts a predictable pattern as this data set is created to replicate the real-world data set from Kumbh Mela, 2015 [1] , which was a

[1] https://a-lavanya.medium.com/crowd-steering-during-massive-gatherings-and-in-daily-life-with-cell-tower-pings-54c3612874ae

massive crowd gathering of 100 million people in the city of Nashik, Maharashtra, India over the period of 40 days, that can be leveraged for further crowd density analysis and modelling.

5.7 Outlier Computation

Detecting outliers in the crowd density and crowd count prediction is paramount for building trustworthy predictive models. In addition, outliers could signify an increase in the crowds, which cannot be explained through some unexpected event. These outliers in crowd density estimation can represent anomalies such as festivals, concerts, emergencies, or machine errors. These outlier understandings are important for several reasons. They may be alerted for unusual and atypical behaviour that needs reinforcement, like extra security or medical services when in crowds.

Moreover, outliers can largely distort predictions in model development, thus causing overfitting or underfitting, which may be corrected when properly handled. Identifying and understanding abnormal observations helps data scientists improve their models to fit the outliers or keep attention to normal situations for forecasting purposes. Therefore, outlier detection in crowd density data helps improve surveillance systems and enhances urban planning and other related aspects, including infrastructure improvements, as it highlights areas or times in which crowd behaviours deviate from the expected ones. This improves public services' and organizers' readiness and incident response policies, making crowd control and safe operations easy. Table 5-3 systematically identifies outlier locations and transitions, to understand the dynamics of crowd movement and distribution and make decisions about ordinary and emergent conditions. The table below illustrates the number of outliers based on hour, i.e., time.

Table 5-3: Spatio-Temporal Outlier Distribution

Hour	1	3	9	11	12	13	14	16	17	19	21	22
Number of Outliers per Location (areas covering about 5km of radius)	5	2	2	2	1	1	1	1	1	1	4	4

The figure 5-6 below illustrates hourly outlier detection across two statistical measures. The mean threshold method, shown in dark red, consistently identifies a higher number of outliers across all hours, ranging from just over 100 to nearly 125 (Sites = Micro location of MOBILE towers located at every half km in urban area) per hour. The median threshold approach, represented in light blue, detects fewer outliers, staying close to 100 for most hours. This contrast indicates that the mean threshold is more sensitive to extreme values, possibly due to its susceptibility to skewness in the data.

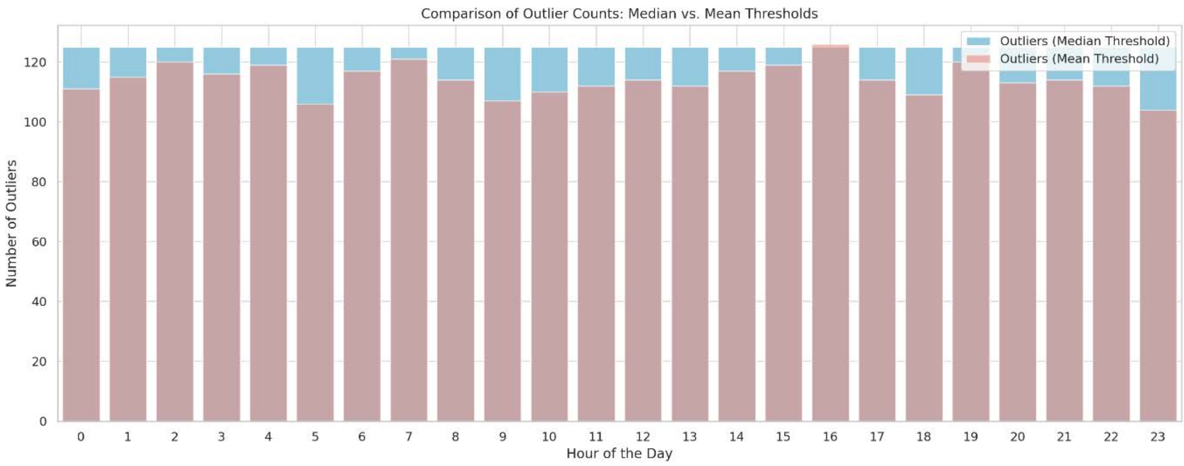


Figure 5-6: Mean Vs. Median Threshold Outliers

The figure 5-7 shows vertical lines represent outlier counts for different subregions across each hour. The plot shows consistent outliers across all subregions throughout the day, with occasional spikes. i.e., SubRegion 50 shows a notable peak around 11:00, suggesting an abnormal increase in crowd density at that hour. This granularity reveals the specific subregions and hours where outlier counts are particularly pronounced, which signals unusual crowd activity or data anomalies in those areas.

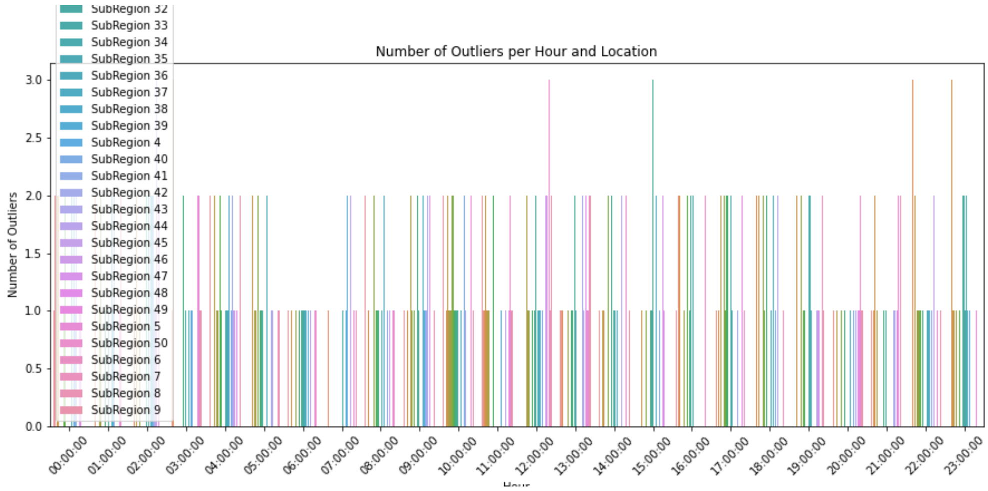


Figure 5-7: Spatio Temporal Outlier Distribution at Micro level

Analysis of outliers in the dataset throws light on aberrational crowd density across varied sub-regions and time intervals. These outliers are important for improving crowd management systems because they identify areas and periods different from normal practices. Understanding these deviations will enable better adjustment of predictive models, improving emergency preparedness and handling unexpected large crowd surges for public safety while allocating enough resources simultaneously.

5.8 Threshold Results

The first step in real-time crowd density estimation without benchmark data is establishing the threshold values. Threshold measures play key roles in recognizing abnormalities with reference points within which values are deemed acceptable. It is quantitative; it allows for issuing alerts, hence quick management of people-based crowd phenomena.

5.8.1 Threshold Algorithm 1 Results

The CCDF Figure 5-8 illustrates the complementary cumulative distribution function (CCDF) of threshold values for 'Ref_Count' and 'Current_Count'. The CCDF for the 'Current_Count' threshold (blue line) descends sharply, indicating a rapid drop-off in the frequency of higher threshold values. Most threshold values are concentrated in the lower range, with nearly all values falling below 200,000. The 'Ref_Count' threshold (orange line) shows a more gradual descent, reflecting a more even distribution across a broader range of threshold values, with none exceeding 600,000. This suggests that 'Current_Count' has a tighter clustering of threshold values while 'Ref_Count' spans a wider range.

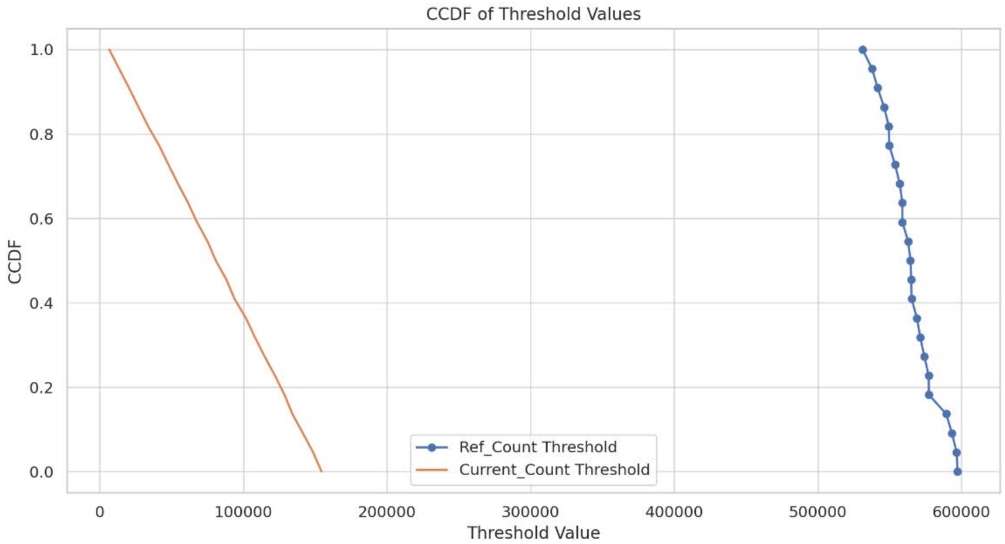
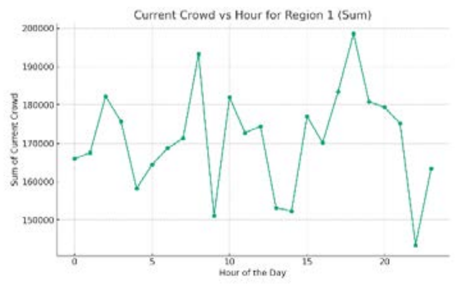
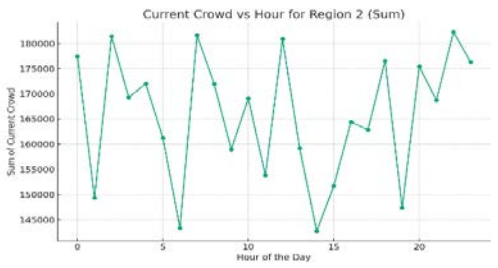


Figure 5-8: Ref_Count Vs Current_Count Threshold Comparisons

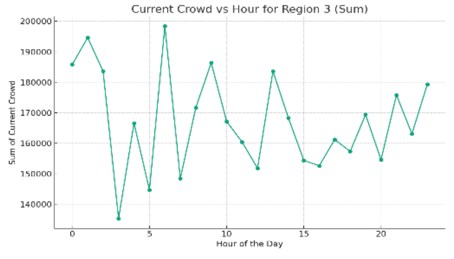
The figure 5-9 (A-J) below for each subplot represents a section with sum counts of current_count which shows the distribution of crowd over hours and every region, thus providing indications of the level changes in the population during a single day. As an illustration, there are rises and falls of crowd density within region number one, which could match rush hour periods or event times and have different characteristics for distinct regions. Remarkably, Region 6 is observed to have a hump, signifying that there was an incident or an anomaly in the crowd distribution.



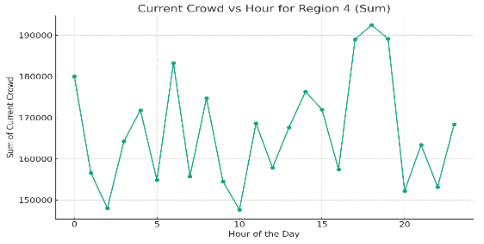
(A)



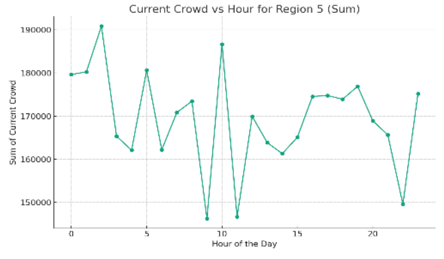
(B)



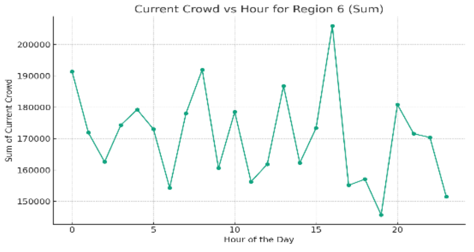
(C)



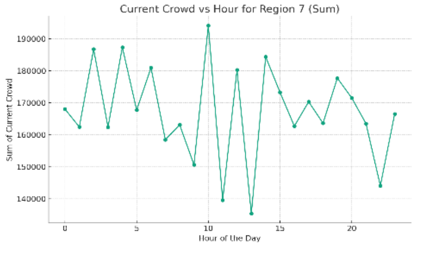
(D)



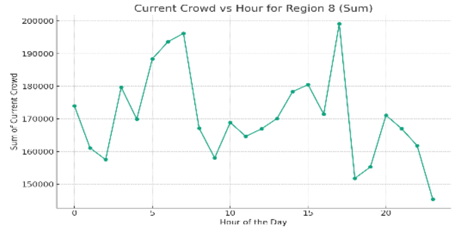
(E)



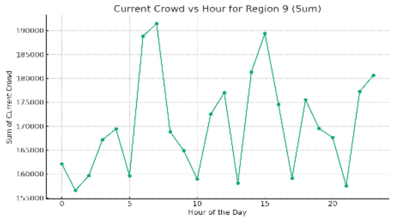
(F)



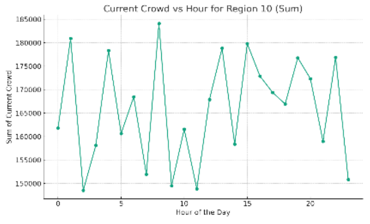
(G)



(H)



(I)



(J)

Figure 5-9: Current_Count Crowd Density Presentation Region Wise

The figure 5-10 below is the total number of people in all the regions per hour daily. The number of people is steadier, but not much so. The count values vary between 1.49 million and 1.73 million people, not one hour showing a dramatic change from that band, meaning fairly steady flow and crowd density is not high.

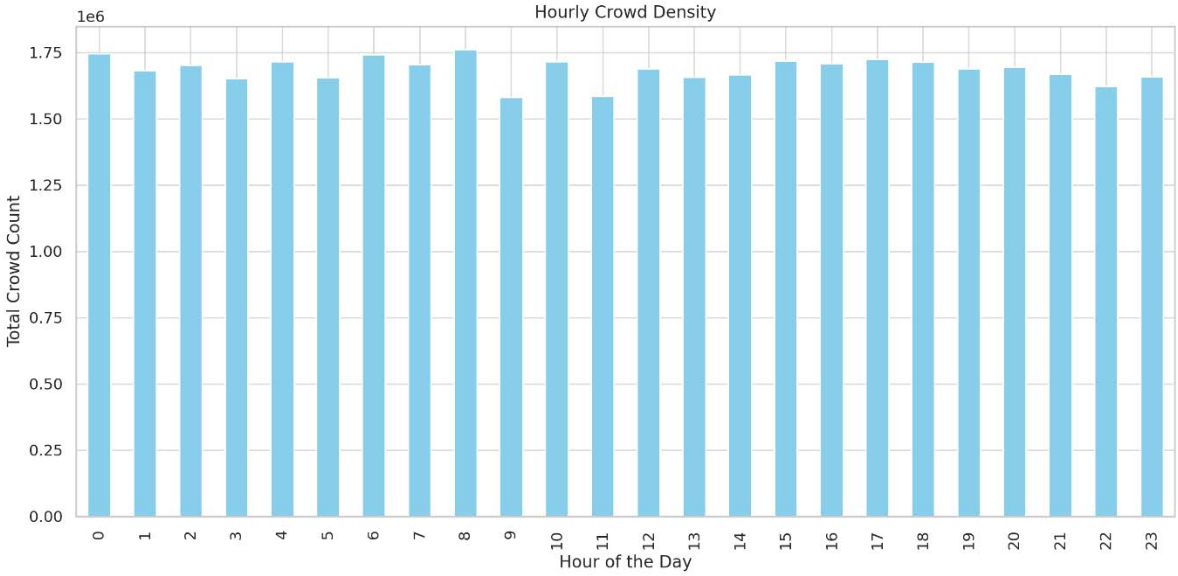


Figure 5-10: Temporal Distribution of Crowd Density

The assessment of crowd density using different thresholds and by-hours distributions in different regions has offered an understanding of crowd motions. This has emphasized the need to develop rigid thresholds to enable effective real-time crowd density estimates, which will benefit proactive crowd management and crisis preparation. The fact that total crowd density was consistent between hours and regional differences indicates the need for a localized monitor to deal with specific challenges in every region.

5.9 Median Vs. Mean Results

Crowd density is measured using median values because these values tend not to be skewed by extreme outliers that might influence mean values. The robustness of medians in measures of typical crowd conditions makes it hard to trigger a threshold-based system for anything that is truly abnormal and not mere outliers. Therefore, crowd monitoring is more precise with median-based thresholds. The figure 5-11 below illustrates the comparison between the mean and median threshold values shown with an hourly variation while the medians are constant. This orange line is referred to as a mean threshold, which is rather unstable, showing significant ups and downs, the last observed at night.

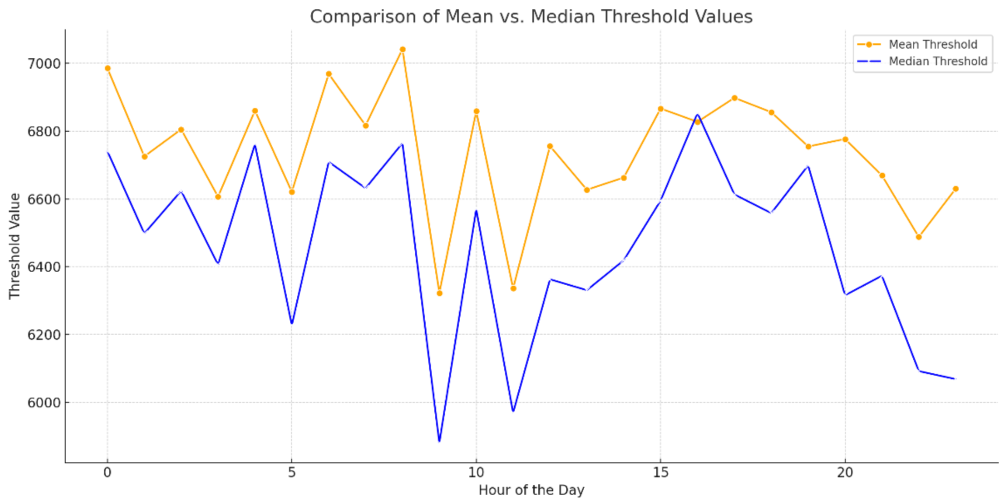


Figure 5-11: Comparison of Mean vs Median Threshold

The figure 5-12 below represents the range for median and mean crowd density thresholds in different subregions. Compared to the minimum and maximum values of the first thresholds (red), the latter exhibit fewer oscillations between their limits, falling close to a smaller area. The median thresholds, marked in blue, show comparatively narrow variations, with their respective maxima rising much above the medians. In this context, such conduct reflects the median's resistance towards outliers as an apt threshold standard of crowd density.

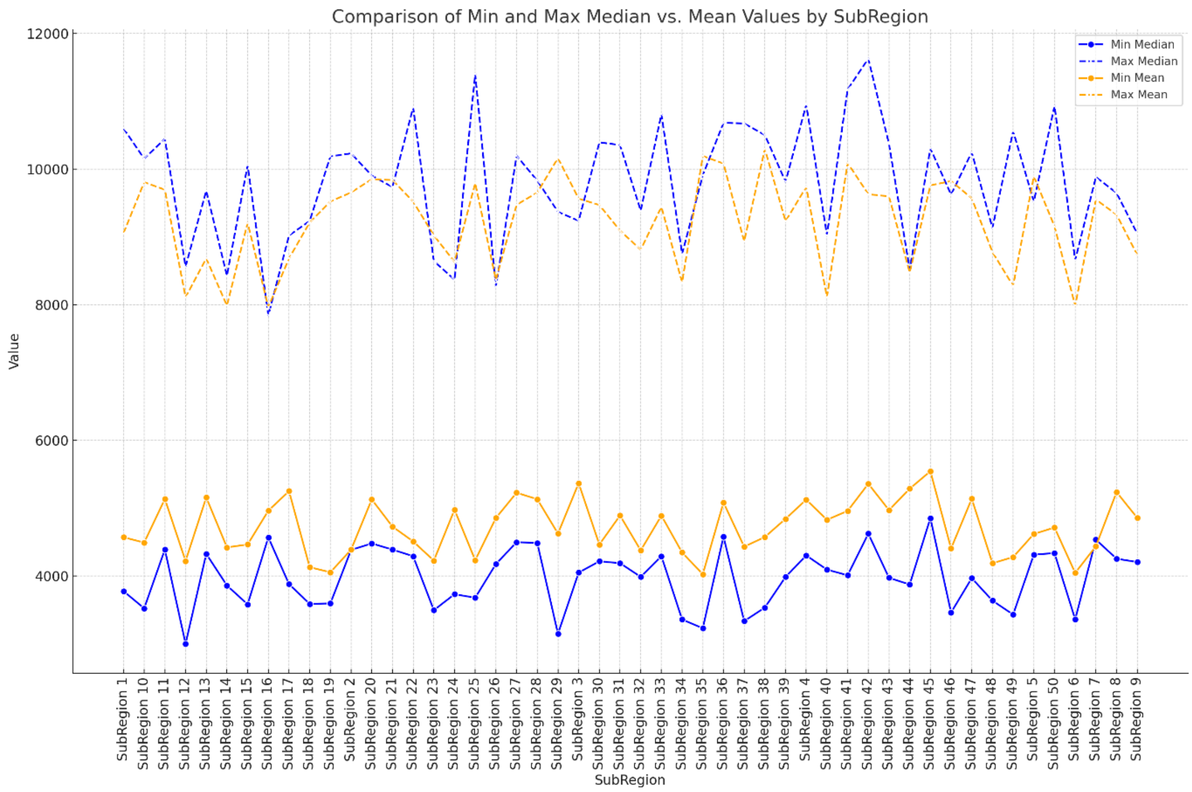


Figure 5-12: Min-Max Comparison Mean Vs. Median

Comparison between median and mean thresholds at different sub-regions and time frames reveals that the median is more useful in estimating mean crowd density. It gives a stable and less variable measurement that suits well with real-time applications, in which sharp changes caused by exceptional happenings are distinctly distinguished from typical changes. A median threshold's stable nature means it is the more trustworthy metric for managing crowd control systems; they must be strong enough towards data anomalies not to raise false warnings and respond adequately to real crowd density changes.

5.9.1 Threshold – Median & Median-Of-Median

Median Threshold plays a critical role in modelling for highly accurate determination of crowd density. It forms the basis for distinguishing ordinary crowd behaviours from extraordinary incidences. It is an objective yardstick against distortion through abnormal information points. It should be consistent enough to guarantee reliable real-time analysis and forecasting. The Median of Median Threshold comes up as a more improved strategy that further enhances the strength of the median approach. This approach involves averaging multiple median values of different data sets or specific periods that capture the essence of crowd dynamics in enhancing the model prediction capability of managing crowd density for improved precision.

The figure 5-13 shows the current count's difference between the hourly median-of-median. The hourly median levels, which differ significantly, signify variable counts for various hours. The median of the median line, on the other hand, tends to be much smoother, which means that it exhibits less sensitivity to short-run changes, hence capturing more reliable information on data trends over time.

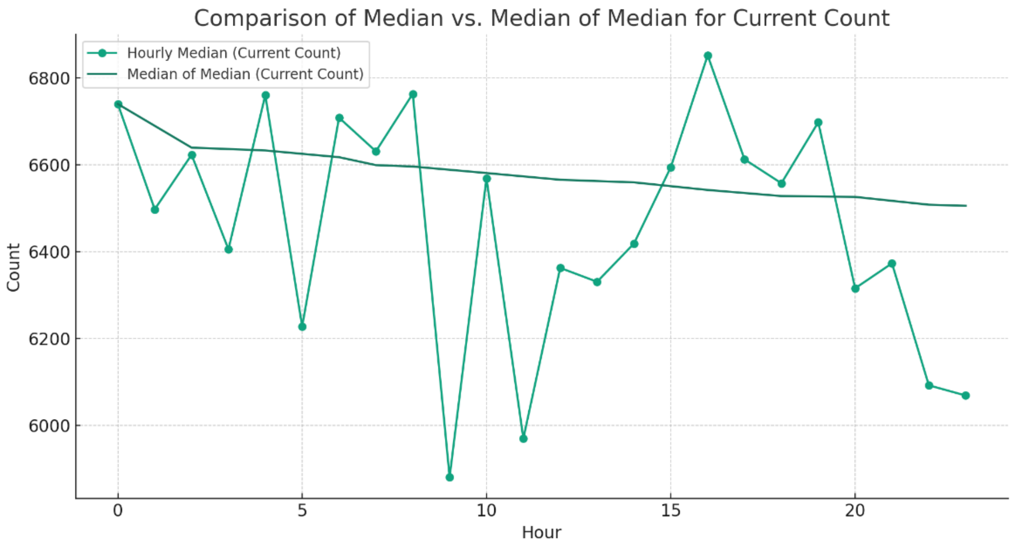


Figure 5-13: Comparison of Threshold Median vs Median-of-Median for Current_Count

The reference count Figure 5-14 displays the hourly median values for reference counts, and their fluctuations indicate the different reference counts for various hours. In contrast, the median line's median shows a flatter pattern, reflecting lesser responsiveness to changing fluctuations in the general data trend.

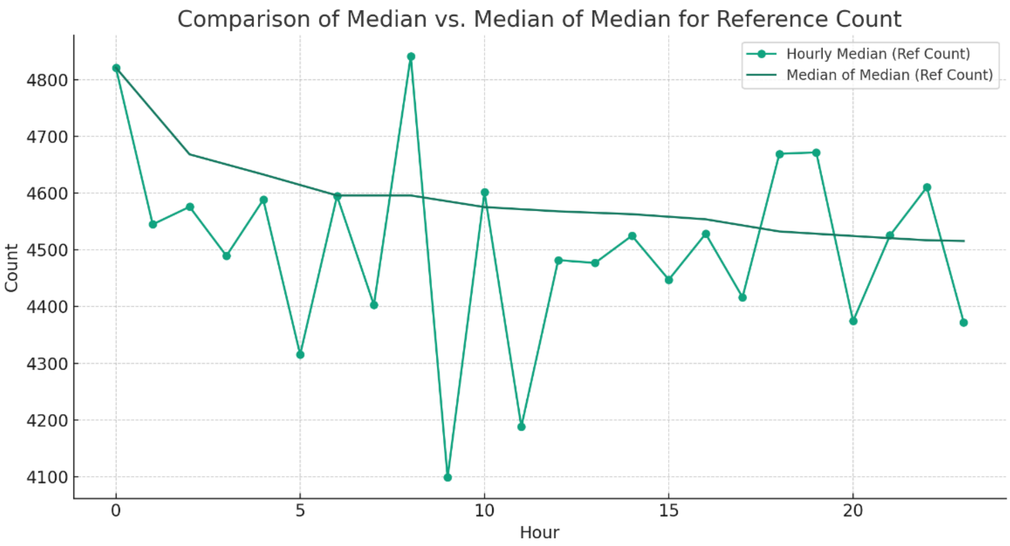


Figure 5-14: Comparison of Threshold Median vs Median-of-Median for Reference Count

While plot hourly medians can show some changes regarding immediate short-term trends, both prove that hourly medians vary a lot, and thus there are big changes in general. The median of medians, however, presents a better, if more stable, index, able to smooth over any short-term fluctuations, showing what lies behind the general trend. Thus, median-of-median is an appropriate metric for analyses inclined towards comprehending general tendencies and weeding out spurious deviations.

5.10 Retrospective Crowd Density Results

The retrospective crowd density analysis provides an overview of what is typically observed in different regional settings over time. The method relies on previously collected data for assessing crowd behaviour, fluctuations in densities, and temporal trends. Retrospective analysis of parameters such as reference, current, and floating counts can reveal information on peak density periods, spatial patterns, and events that may lead to crowd density fluctuations. These analyses provide useful information that can be used to develop better future crowd management strategies, improve public safety, and allocate resources in anticipation of occurrences and gatherings. With that, building on the predictive models and creating strategic plans relevant to urban and event management is easier.

5.10.1 Crowd Overview Heatmap

A heat map figure 5-15 shows current population densities for each area at a specific hour. The darker shade of blue denotes the areas with higher crowding, and the lighter one represents those with a lower population density. For some instances, the peak crowds within certain areas seem to occur during certain instances, which might be rush hours or events. However, in other areas, a similar distribution pattern is less consistent. However, the area designated as “Region 1” shows a high population in almost all periods as opposed to the rest of the region, suggesting that this is where people mostly gather.

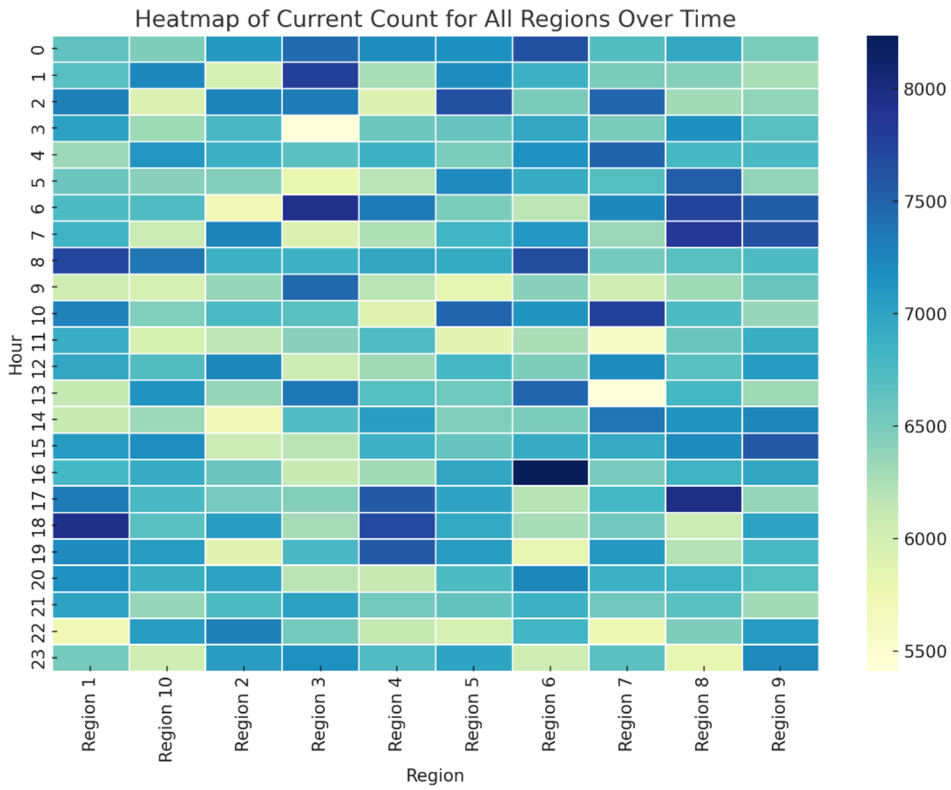


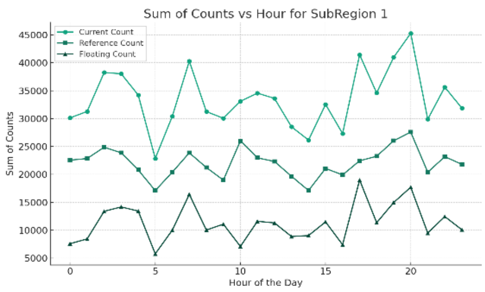
Figure 5-15: Heatmap Crowd Density Distribution Time vs Region

5.10.2 Crowd Assessment – Granular Level

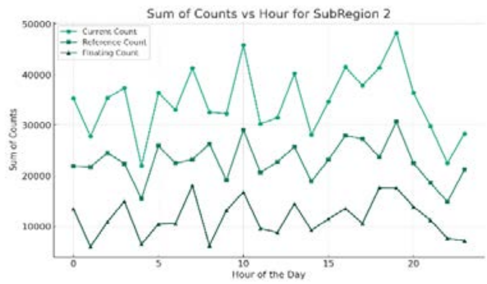
The following set of plots named figure 5-16 to Figure 5-25 display the crowd density statistics in different sub–regions under each covering region. They track three key daily metrics: Ref_Count, Current_Count, and Floating_Count. The temporal characteristics differ in subregions; however, some demonstrate sharp peaks that might indicate rush hours or event-directed crowds, whereas others express uniform daily counts. In each sub-region, the Current_Count may reach their peaks at particular points, implying that there could be repeating periods of congestion and events that attract many persons during such instances.

A. Crowd Desnity Distribution Resgion 1

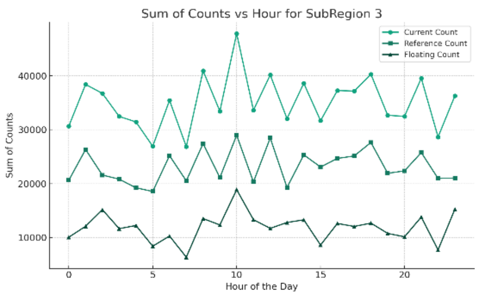
The distribution of crowd density in figure 5-16 (SubR-1 to SubR-5) across the sub-regions is also very much time based. In contrast to Reference Count and the less lumpy Floating Count, "Current Count" characteristically demonstrates spiky peaks suggesting strong levels of activity. These millimeter-high humps may signify particular hours or parts of a day when huge volumes mass together--or hives out suddenly--in response to work schedules or social functions. But the other figures indicate that people were about evenly distributed throughout the day. Understanding these patterns is essential for effective crowd management, infrastructure planning and service provision in each sub-region.



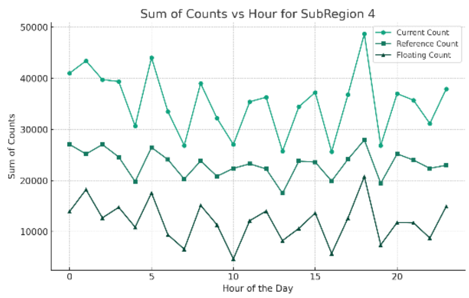
SubR-1



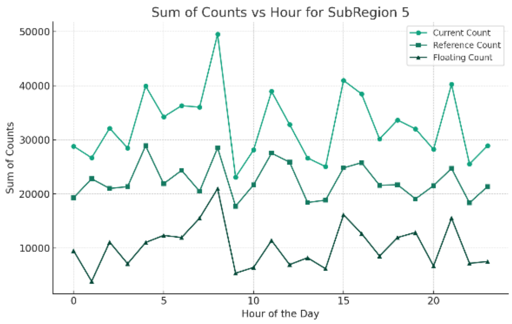
SubR-2



SubR-3



SubR-4

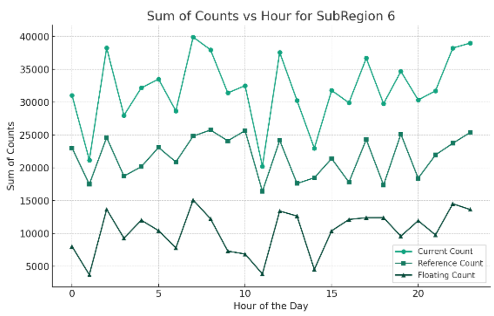


SubR-5

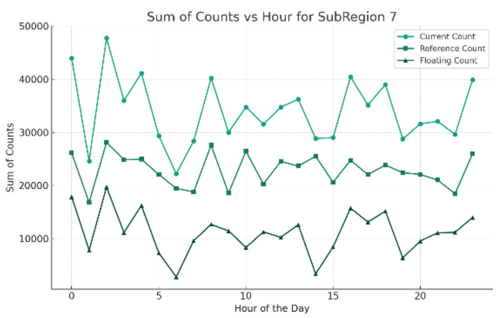
Figure 5-16: Crowd Density Metrics – Regions 1 (Sum_Count Vs Hours)

B. Crowd Density Distribution Region 2

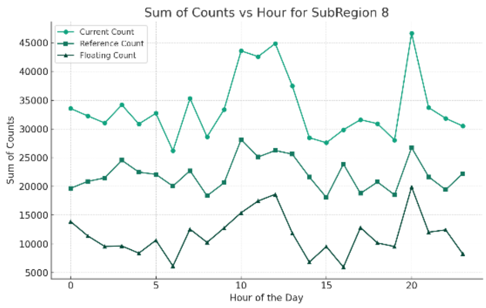
The line graphs in figure 5-17 for SubRegions 6 through 10 illustrate the hourly crowd density with three measures: Current Count, Reference Count and Floating Count. Current Count for SubRegion 6 has very pronounced peaks, especially during the early hours and at the end of the day. SubRegion 7's Current Count is similarly fluctuating with great jumps in numbers, showing that crowd movement can vary. SubRegion 8 shows a clear peak around the 18th hour in Current Count, while Reference and Floating Counts are still relatively stable. And SubRegion 9 comes to a nearly identical conclusion, with an especially jarring peak in Current Count at just past the twentieth hour. Lastly, Current Count in SubRegion 10 rises sharply about two hours after noon. In general, the Current Count shows more frequent and higher values than do others. Generally speaking, this hints at times of concentrated activity within a given interval.



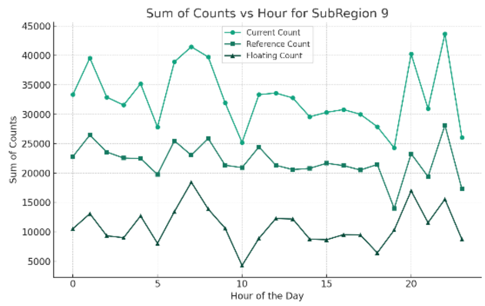
SubR-6



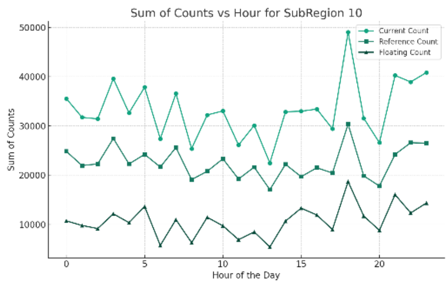
SubR-7



SubR-8



SubR-9

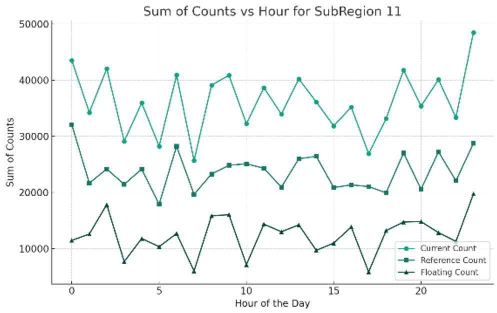


SubR-10

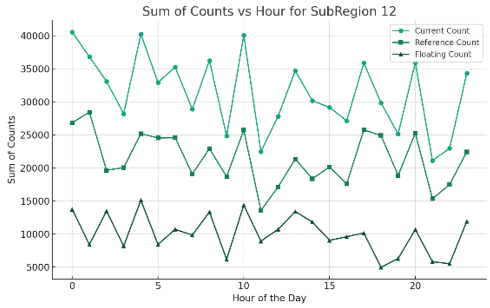
Figure 5-17: Crowd Density Metrics – Region 2 (Sum_Count Vs Hours)

[3]. Crowd Density Distribution Region 3

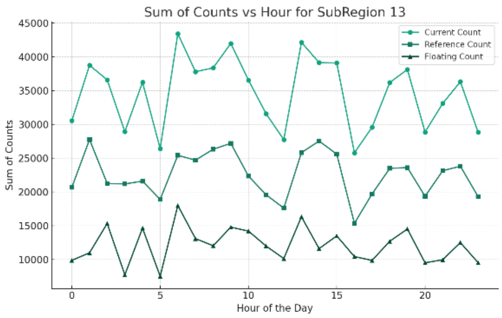
These data visualized in figure 5-18 for SubRegions 11 through 15 illuminate the fluctuation in crowd density over time of day; represented by numbers looking at Current Count, Reference Count and Floating Count. The Current Count also shows a clear peak at the end of the day in SubRegion 11, demonstrating that crowd activity surges later. The middle of the day shows a huge jump in Current Counting for SubRegion 12, while Reference and Floating examples remain more stable. The late evening Current Count is especially remarkable in SubRegion 13. According to the Current Count, a mid-day peak and another in the late afternoon. Finally, SubRegion 15 shows a strong upward spike in the Early Hours with another distinct increase towards late evening. In these sub-regions the Current Count fluctuates wider than either the Reference or Floating Counts, demarcating causations of crowding in time.



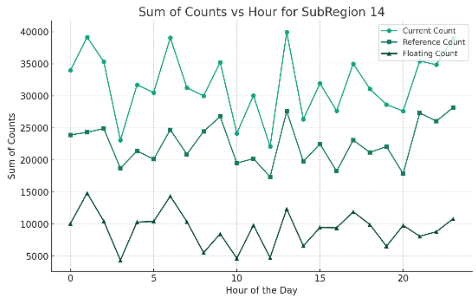
SubR-11



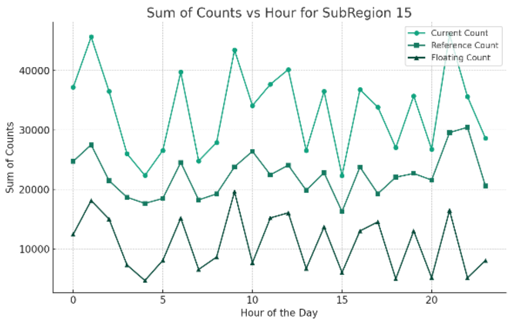
SubR-12



SubR-13



SubR-14

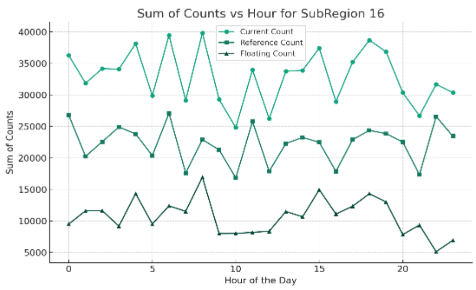


SubR-15

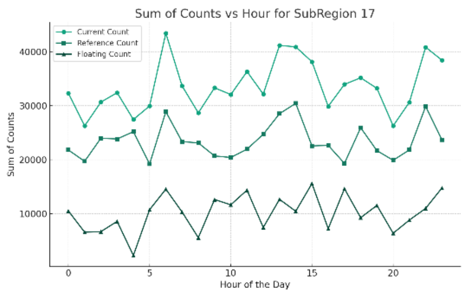
Figure 5-18: Crowd Density Metrics – Region 3 (Sum_Count Vs Hours)

[4]. Crowd Density Distribution Region 4

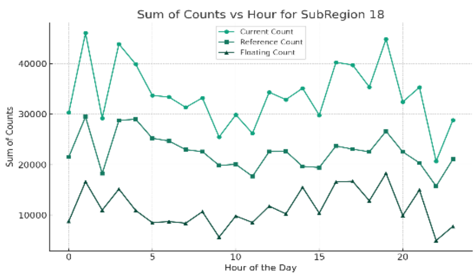
The figures below show the fluctuation in crowd counts by hour for SubRegions 16 through 20. Within these subregions, numbers vary noticeably over the course of a day. For each subregion, three types of counts are compared: Current Count, Reference Count and Floating Count. In particular, Current Counts peak at specific hours. It is speculated that this may be due to routine arrival and departure from work or transformation stations. The Reference Counts, stable indicators of historical baseline. Floating Counts, which would seem to be lowest in all sub-regions and could represent a short term or less stable segment of the population. There are also differences in the peak times and amounts of counts between subregions, which represents unique distribution and movement patterns within each area. This data shows us just how dynamic crowd movement and density can be. For urban planning, resource allocationsand understanding human behaviour in these areas this could become very important information.



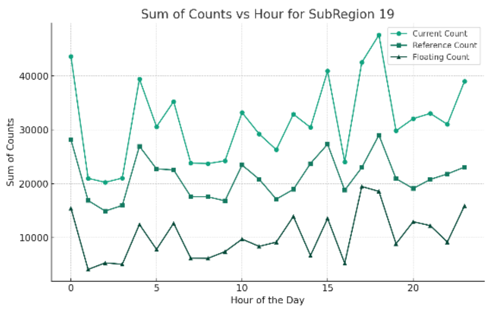
SubR-16



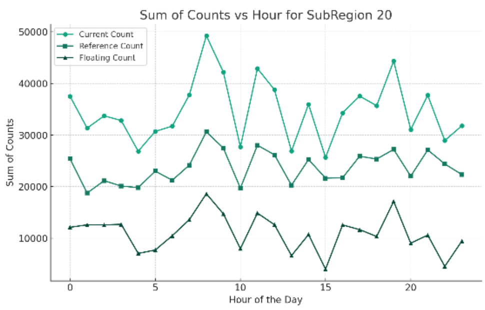
SubR-17



SubR-18



SubR-19

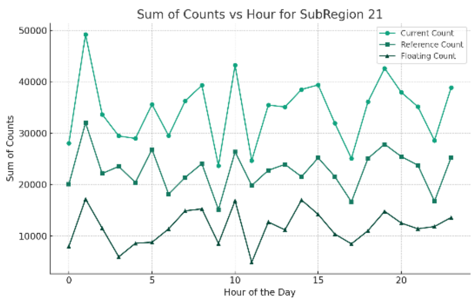


SubR-20

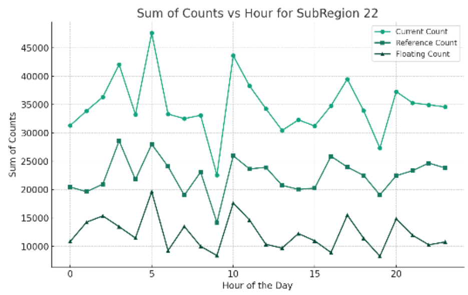
Figure 5-19: Crowd Density Metrics – Regions 4 (Sum_Count Vs Hours)

[5]. Crowd Density Distribution Region 5

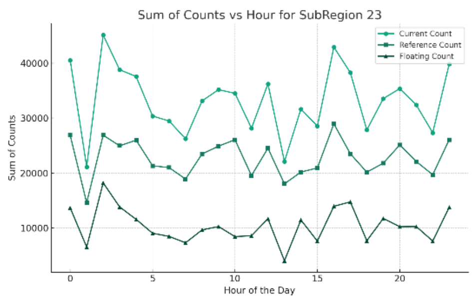
The addition crowd count graphs for SubRegions 21 through 25 continue to show the hourly fluctuation in density of population. The Current Count stresses large peaks and troughs, which could be representative of a standard day's activities, such as getting to work or playing. Reference counts, probably representing an historical average, show less pronounced variations. A relatively stable pattern is indicated over time. Still the lowest are Floating Counts, signaling perhaps a subset of society that is sporadic or mobile. Different subregions have different temporal patterns; some with sudden jumps upward or downward at particular hours. These patterns are of enormous importance in deciphering the features of temporal crowd dynamics which can help optimize management, on a subregional level at least and also permit more efficient traffic planning as well as better emergency response preparation within any given national or cultural locality.



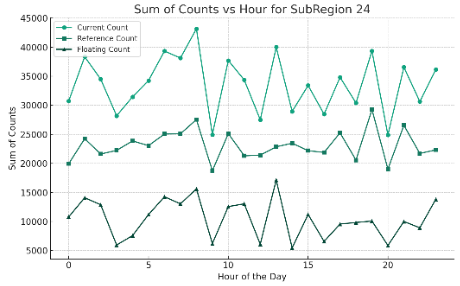
SubR-21



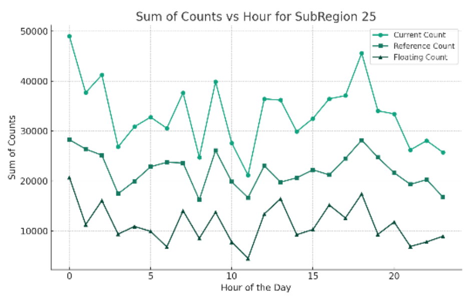
SubR-22



SubR-23



SubR-24

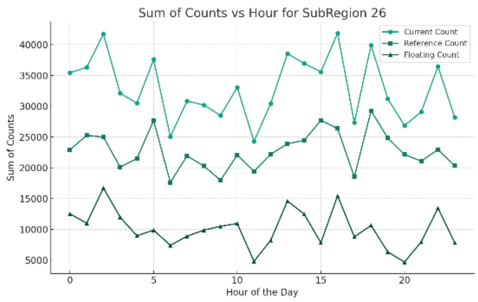


SubR-25

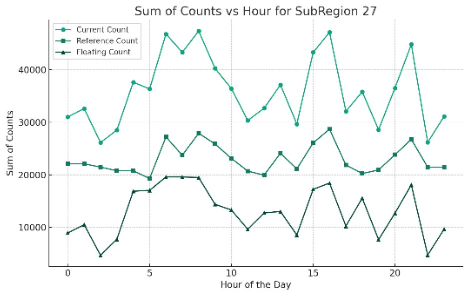
Figure 5-20: Crowd Density Metrics – Regions 5 (Sum_Count Vs Hours)

[6]. Crowd Density Distribution Region 6

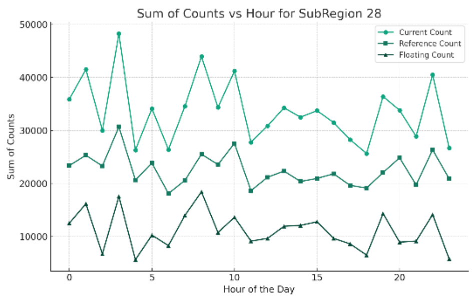
SubRegions 26 through 30 A series of graphs showing crowd counts by the hour breaks down efficiently into three separate groups with clear patterns of fluctuation for Current, Reference and Floating Counts. The current counts vary widely, with some of the swells during certain hours perhaps representing logically those which subserve indeed have often taken place but up to date time even if they are just about things that happen every day or by gathering groups. Crowd density is most commonly represented through reference counts, which show more steady trends. These could be representative of a normalized or enough crowd density range based on historical data. Suggesting a degree of uncertainty or unexpectedness and chance (it's not the same people every time), Floating Counts fell without exception in all sub-regions. These figures show how swirling the crowd flows and where it crawls--information of great value for both urban administration and event planning, not to mention design saving effort digging streets like rabbit burrows.



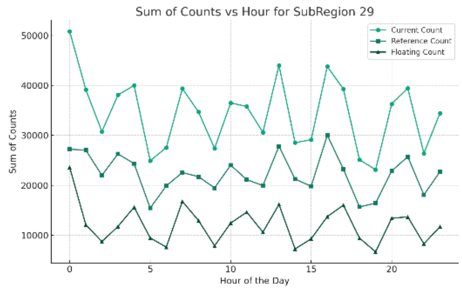
SubR-26



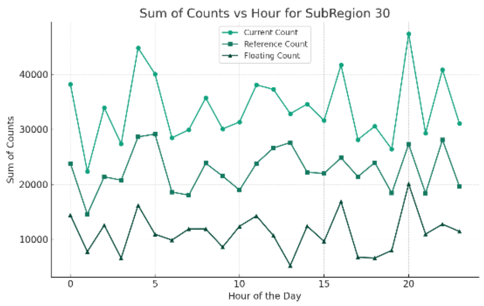
SubR-27



SubR-28



SubR-29

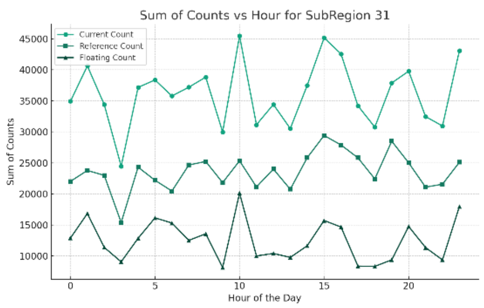


SubR-30

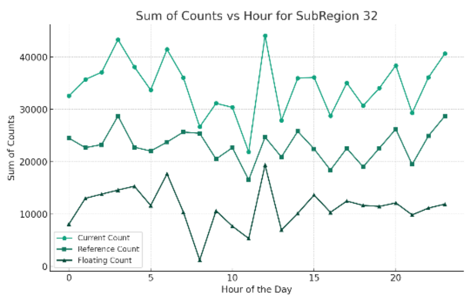
Figure 5-21: Crowd Density Metrics – Regions 6 (Sum_Count Vs Hours)

[7]. Crowd Density Distribution Region 7

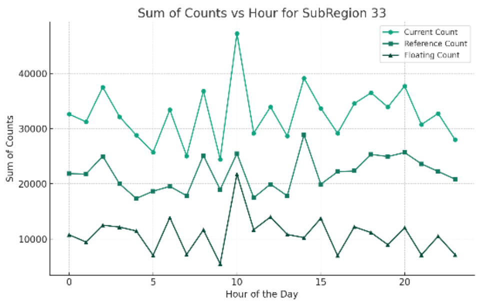
The graphs for SubRegions 31 through 35 depict the hourly crowd counts, showcasing the variations within each category: Count, reference count and floating count. Current Count shows a pattern oscillating daily, with peak surges probably coming during rush hours or other times of high activity. The troughs might represent slower days or quieter dead time. The Reference Count shows a more stable pattern, and seems to represent an average or expected count based on past surveillance. The Floating Count, consistently lower than any of the others indicates a portion of society that is always fluctuating and in flux. These across-the-board patterns within the subregions could reflect very different types of socioeconomic activity and migration flows unique to each area that can be helpful in planning and managing operations.



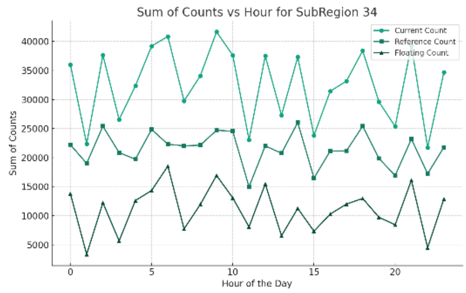
SubR-31



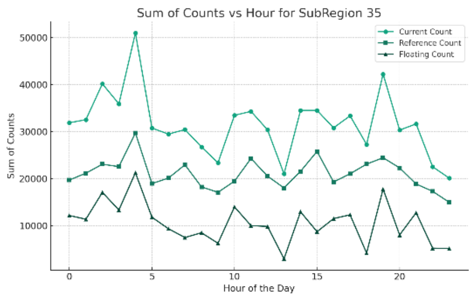
SubR-32



SubR-33



SubR-34

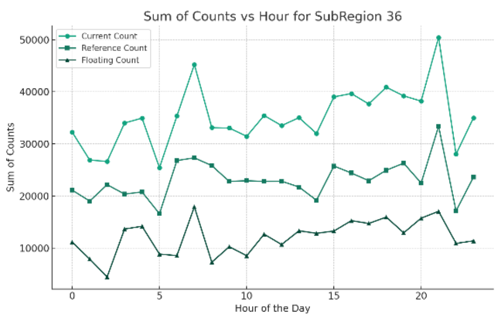


SubR-35

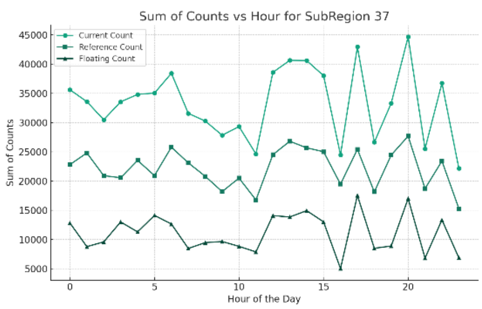
Figure 5-22: Crowd Density Metrics – Regions 7 (Sum_Count Vs Hours)

[8]. Crowd Density Distribution Region 8

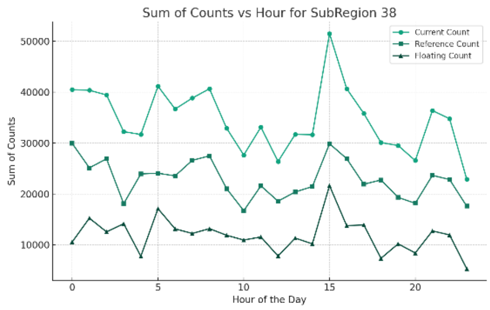
Hourly fluctuations in crowd density are shown by data from SubRegions 36 to 40. There is a different pattern for Current Counts peaking in each subregion, perhaps due to local events or rush hours. Reference Counts trace out more of a stable trend line--probably an average expectation based on historical data. League Floats are the most inconstant, being obviously a fluid or occidental population subgroup. These variations in counts are of course important for understanding the nature of crowd dynamics, and can help plan municipal services, emergency response or event organization.



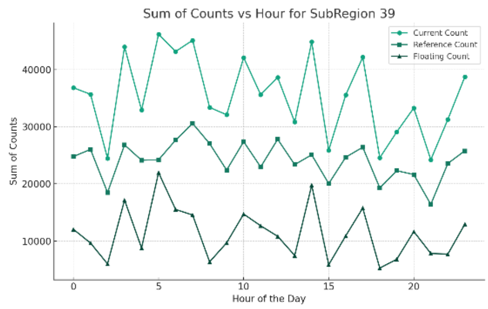
SubR-36



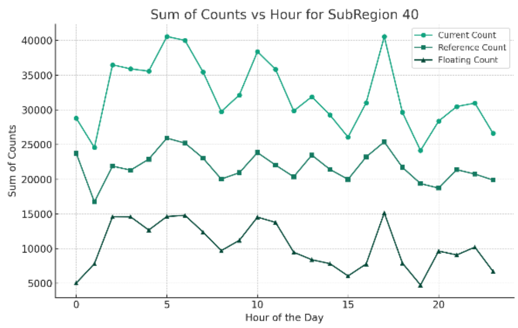
SubR-37



SubR-38



SubR-39

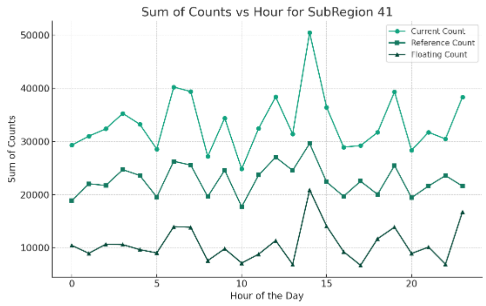


SubR-40

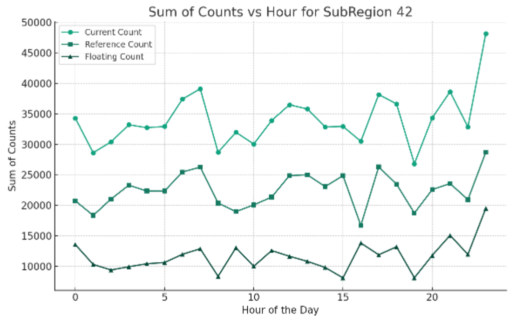
Figure 5-23: Crowd Density Metrics – Regions 8 (Sum_Count Vs Hours)

[9]. Crowd Density Distribution Region 9

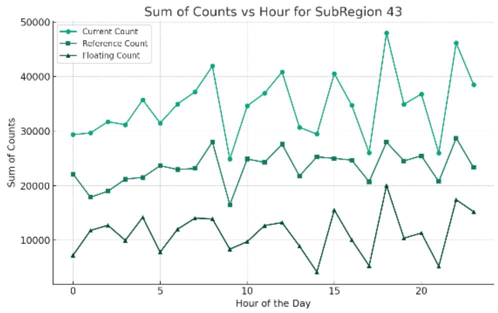
The crowd count graphs for SubRegions 41 to 45 show significant diurnal fluctuations. Each subregion has sharp peaks on the Current Counts graph which may correspond to certain incidents, or max activity times. And troughs seem probably related with quieter periods of time. A relatively steady pattern in Reference Counts implies an average at a stable level throughout history. Floating Counts are smaller and more volatile, implying that they perhaps record a temporary or mobile portion of the population. These patterns reflect relatively local activities and can help predict and respond to crowd-related demand in these areas.



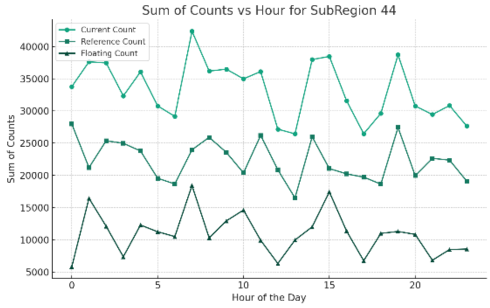
SubR-41



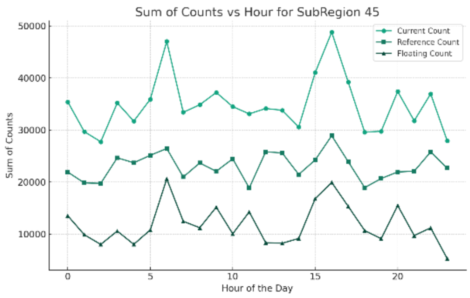
SubR-42



SubR-43



SubR-44

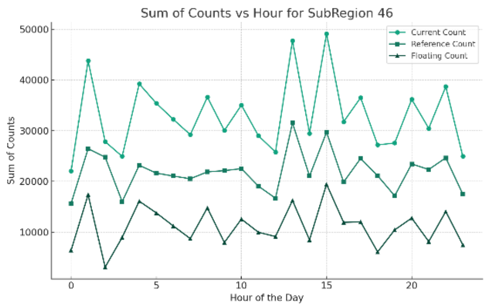


SubR-45

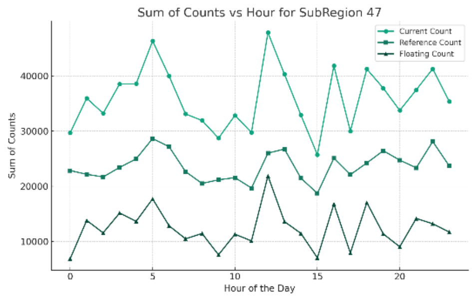
Figure 5-24: Crowd Density Metrics – Regions 9 (Sum_Count Vs Hours)

[10]. Crowd Density Distribution Region 10

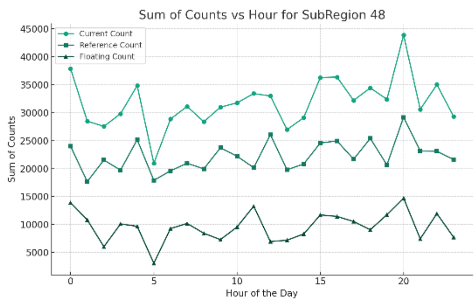
Counts per hour across subregions (current, reference and floating) are shown in these line graphs. There are also some noticeable changes, and currents for instance peak at certain hours. This could point to trends or patterns of activity in the future. For example, in SubRegion 41 a prominent peak exceeding the current count of more than 4500 is reached. The floating and reference counts vary to only quite an extent across one day. However, SubRegion 50 is more purely regular in the entire region. The lowest overall values remain within a narrow range-not over thirty thousand. These maps illustrate the temporal nature of such counts, displaying both similarity and difference among different subregions on any given day.



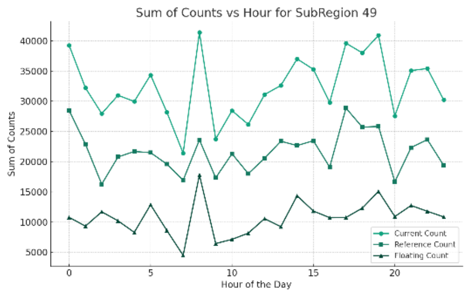
SubR-46



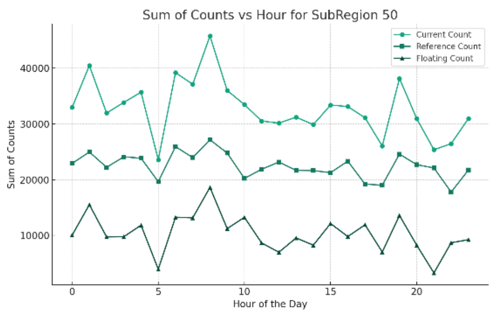
SubR-47



SubR-48



SubR-49



SubR-50

Figure 5-25: Crowd Density Metrics – Regions 10 (Sum_Count Vs Hours)

The data over time shows rather great variations in activity patterns. Reality is such that there are currently very frequent peak spikes; whereas when I as well have looked at reference and floating counts, they do show greater continuity throughout the day up until this point.

5.11 Cumulative Distribution – All Regions

The figure 5-27 present a summarized graphical representation of ‘Current Count’ data for several regions. Similarly, each Histogram shows the extent to which crowd sizes are alike in each area and has the tallest bars corresponding to the most common counts. These distributions most likely correspond to a certain area of typical crowd density within each region. In some areas, there is a tight dispersion, which means that the crowd size remained constant or similar for most of the period, while other areas had broader spreads, implying higher variation in the number of crowds in those locations. Such a pattern illustrates how specific events associated with a certain place determine the typical movement of people in a given area.

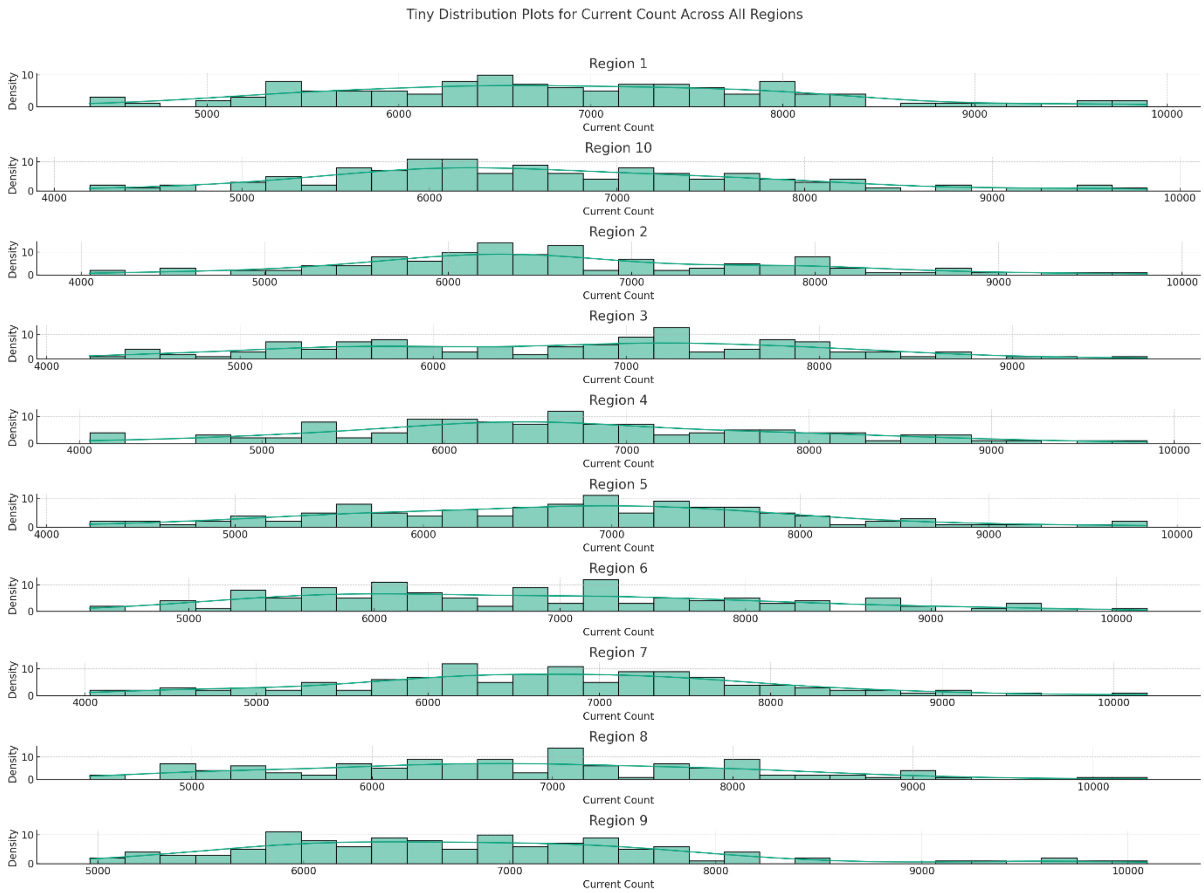


Figure 5-26: Cumulative Distribution for Crowd Density All Regions

Analysis of crowd density metrics using various visualization techniques shows varied patterns of crowd distribution in different parts and hours. Inference from this information shows an important fluctuation in population density and intensity. Such findings have crucial informative value for improving crowd control policies and understanding space-time crowd dynamic processes.

5.12 Algorithm 6: Multivariate Crowd Density Analytical Framework (CD_AF) Results

The primary goal of this research is to utilize the Historical Threshold Crowd Density Classification methodology. First, it seeks to capitalize on the normally overlooked power of historical data in providing an additional dimension to contemporary crowd ratings beyond the surface meaning. The figure 5-28 shows that algorithm analyses the historical data to discover the hidden patterns and trends of crowd density, which is not obvious in real-time data. The method adopted aligns to improve predictive precision regarding urban planning and crowd control measures. Urban Planners and authorities can use this information to plan by making informed decisions to predict future challenges and implement preventive measures.

Figure 5-27: Class Diagram Algorithm 6

5.12.1 Threshold Current_Count

The maximum median reference count is observed between the ninth and eighteenth hour, with values exceeding 4800. The median reference count dips noticeably around the 8th and 16th hours, falling below 4300. The data show that there is a daily fluctuation of reference counts with peaks of activity.

5.12.1.1 Distribution of Current_Count

The figure 5-29 presents medians estimated at approximately 6,000 to 9,000, with a change based on hourly intervals. As the day ends, more sporadic spikes of outliers can be noted. There is variation in the various hourly count numbers, as indicated in respective IQRs.

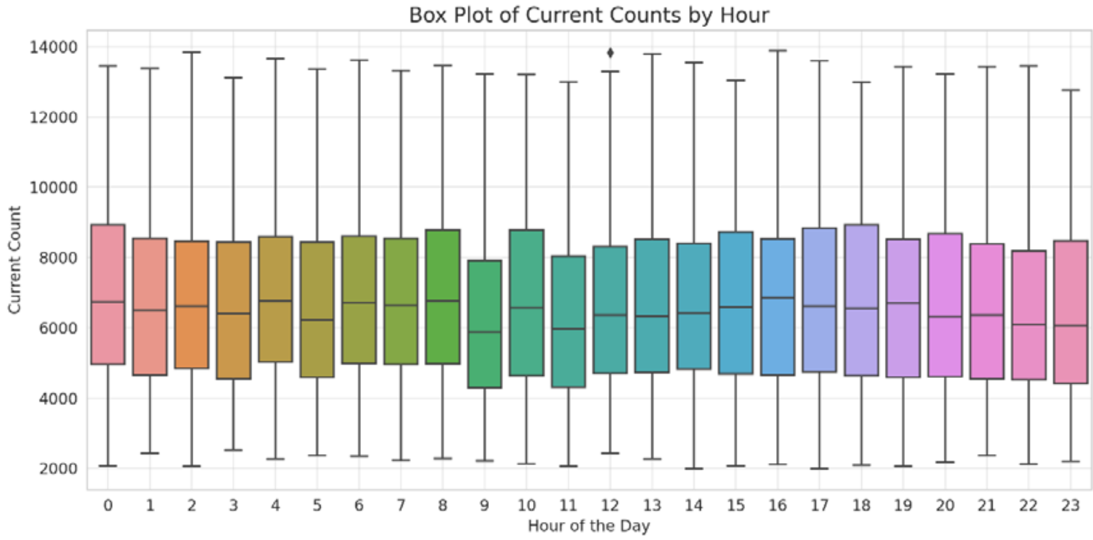
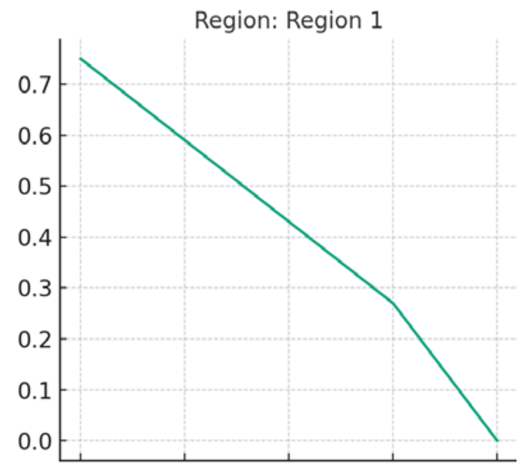
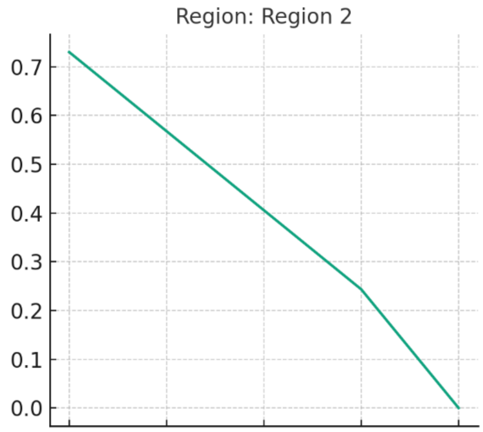
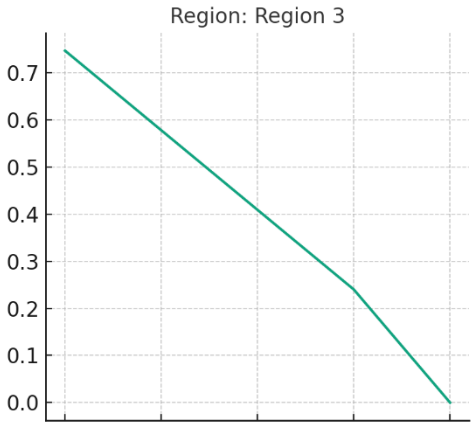
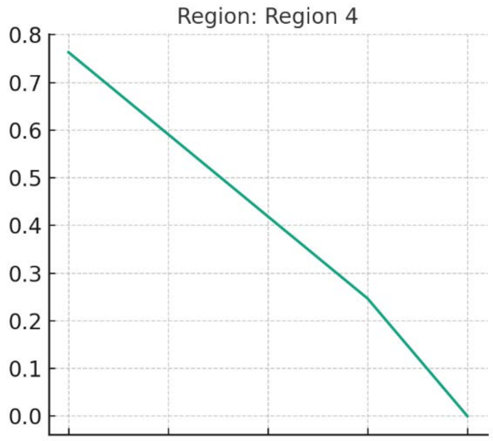
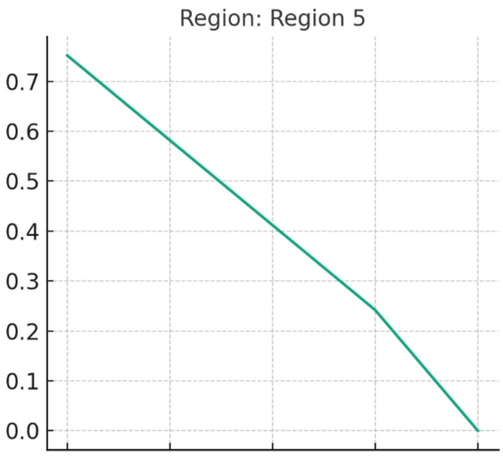
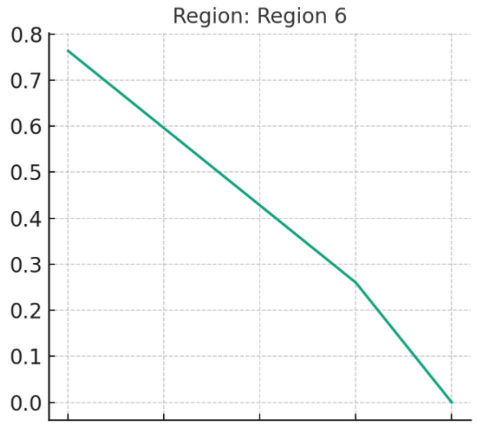
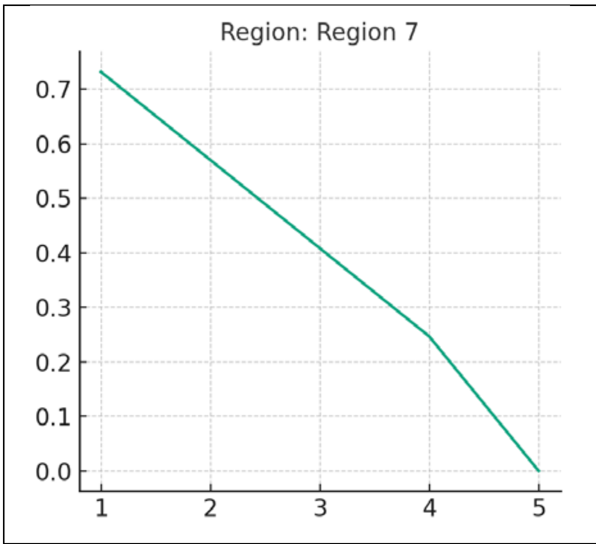


Figure 5-28: Current Count vs Hour Distribution

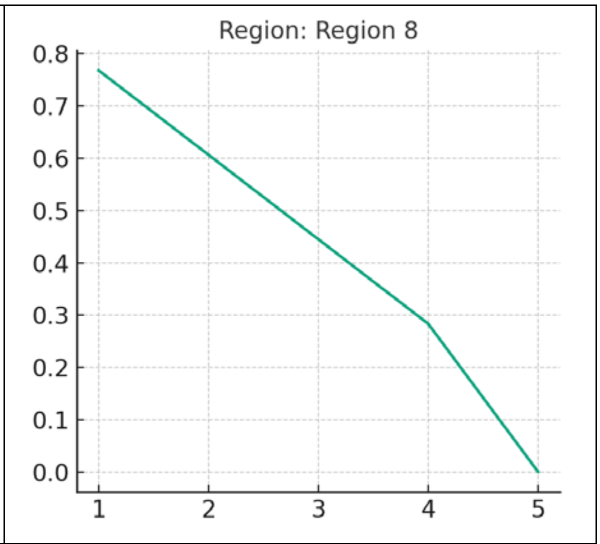
5.12.2 Crowd Classification

CCDF plots in figure 5-30 (A to J) depict the probability of classifying a region with particular or higher values. The higher the crowd classification code, the higher the crowd density. The graphs depict downward movements, indicating fewer highly dense areas. The CCDF value is 0.2, at a ‘high’ level; this means that out of all the observations, 20% are high and above. Different slopes and sizes of steps demonstrate that crowd densities vary between regions.

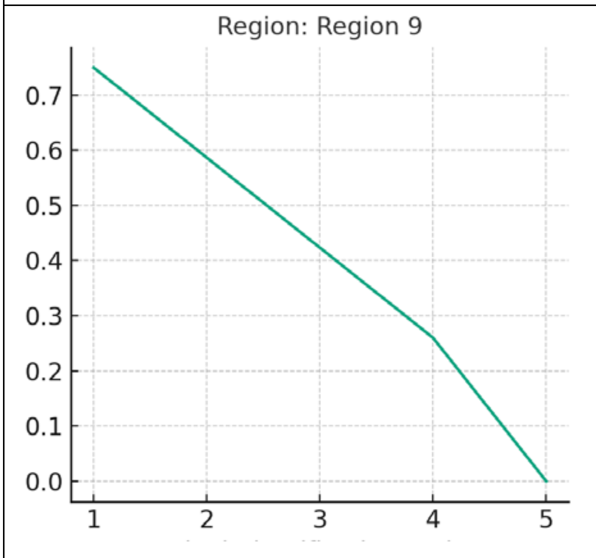
 Region: Region 1 0.7 0.6 0.5 0.4 0.3 0.2 0.1 0.0	 Region: Region 2 0.7 0.6 0.5 0.4 0.3 0.2 0.1 0.0
(A)	(B)
 Region: Region 3 0.7 0.6 0.5 0.4 0.3 0.2 0.1 0.0	 Region: Region 4 0.8 0.7 0.6 0.5 0.4 0.3 0.2 0.1 0.0
(C)	(D)
 Region: Region 5 0.7 0.6 0.5 0.4 0.3 0.2 0.1 0.0	 Region: Region 6 0.8 0.7 0.6 0.5 0.4 0.3 0.2 0.1 0.0
(E)	(F)



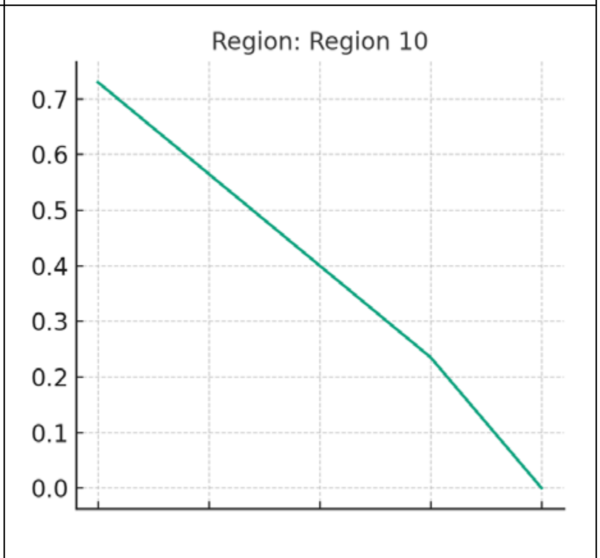
(G)



(H)



(I)

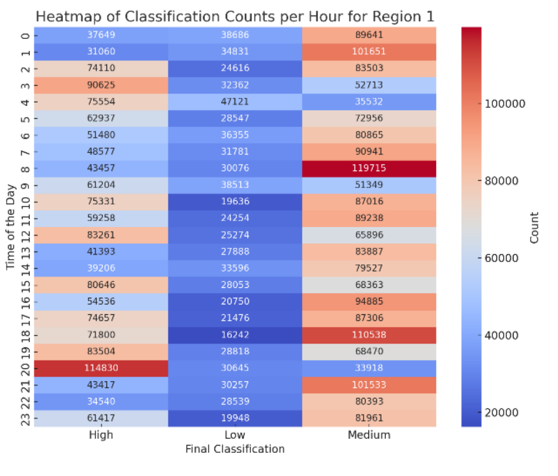


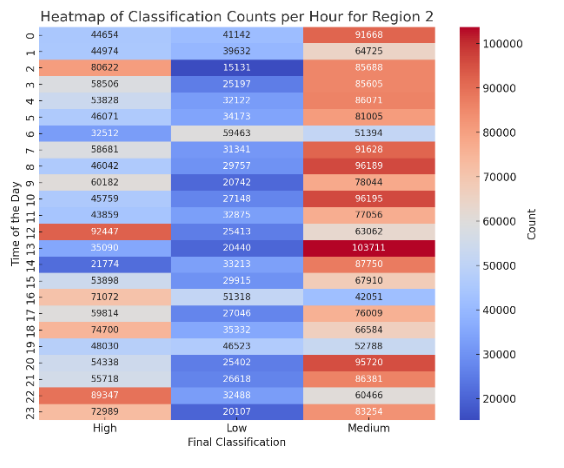
(J)

Figure 5-29: CCDF for Crowd Classification vs Region

5.12.3 Crowd Density Heatmaps

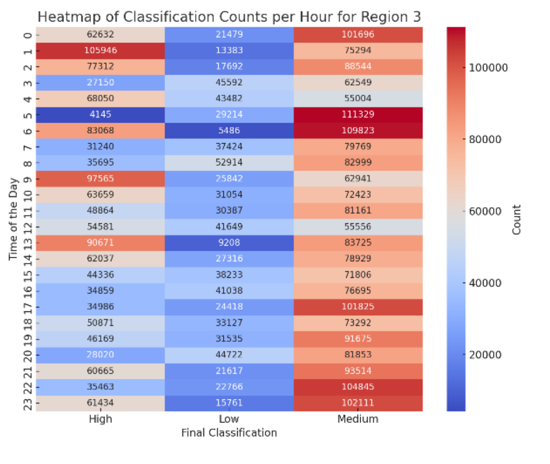
Faceted heat maps show in figure 5-31 (A to J) shows the distribution of crowd density classifications such as High, low, and medium at different hours in multiple areas. The intensity of the color in each heatmap cell correlates with the classification frequency: darker shades indicate higher occurrences. Dark-blue cells indicate occasions with higher ‘High’ cases, implying high crowd concentration. On the other hand, light cells denote low counts compared to yellowish ones that portray minimum crowding. Every subplot represents one particular region, facilitating crowd density trends in regions.

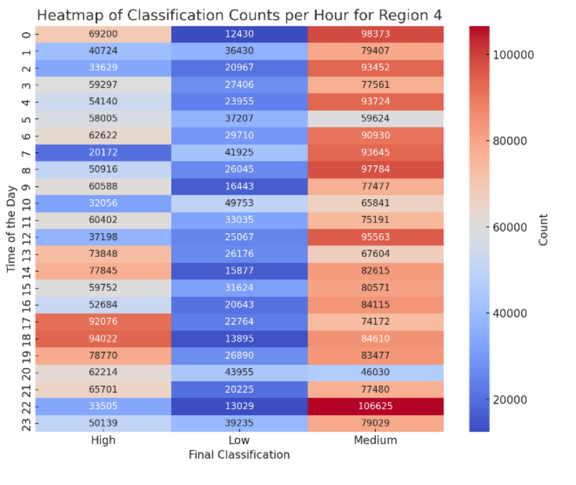

Heatmap of Classification Counts per Hour for Region 1
Time of the Day
Final Classification
High
Low
Medium
Count


Heatmap of Classification Counts per Hour for Region 2
Time of the Day
Final Classification
High
Low
Medium
Count

(A)

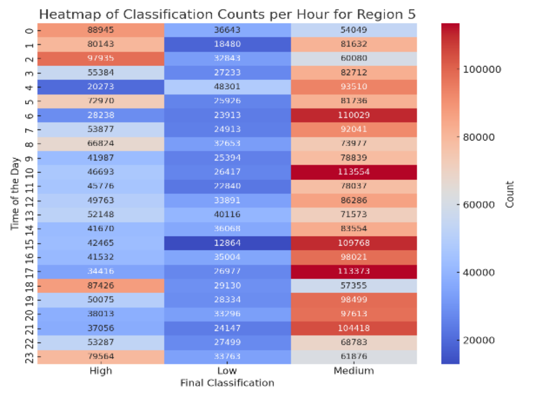
(B)

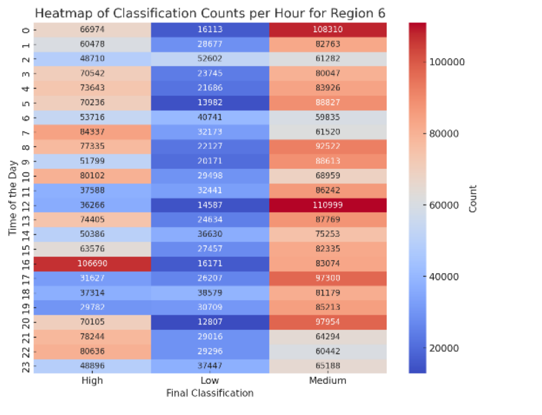

Heatmap of Classification Counts per Hour for Region 3
Time of the Day
Final Classification
High
Low
Medium
Count


Heatmap of Classification Counts per Hour for Region 4
Time of the Day
Final Classification
High
Low
Medium
Count

(C)

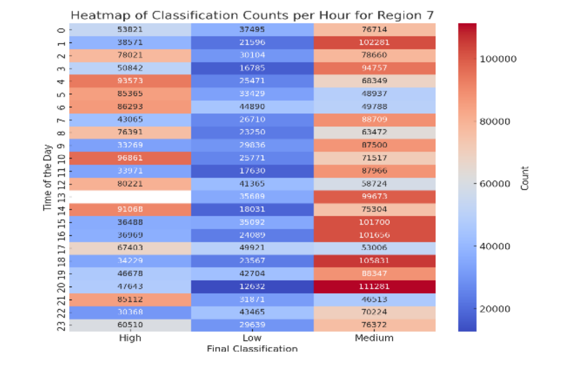
(D)


Heatmap of Classification Counts per Hour for Region 5
Time of the Day
Final Classification
High
Low
Medium
Count

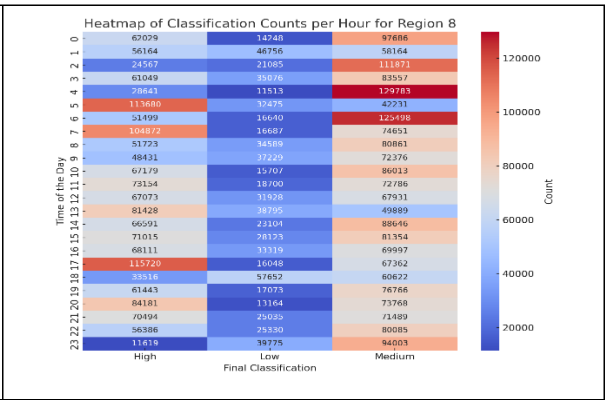

Heatmap of Classification Counts per Hour for Region 6
Time of the Day
Final Classification
High
Low
Medium
Count

(E)

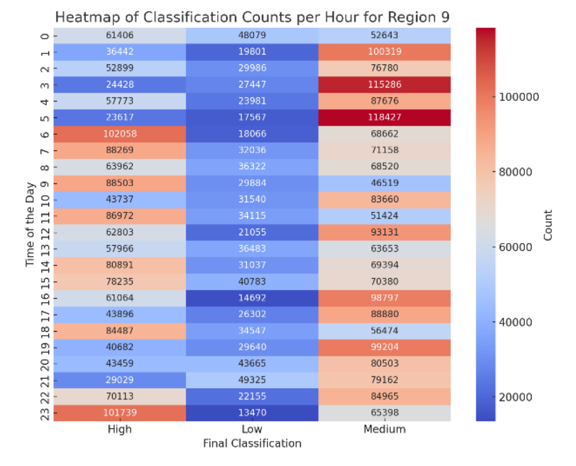
(F)



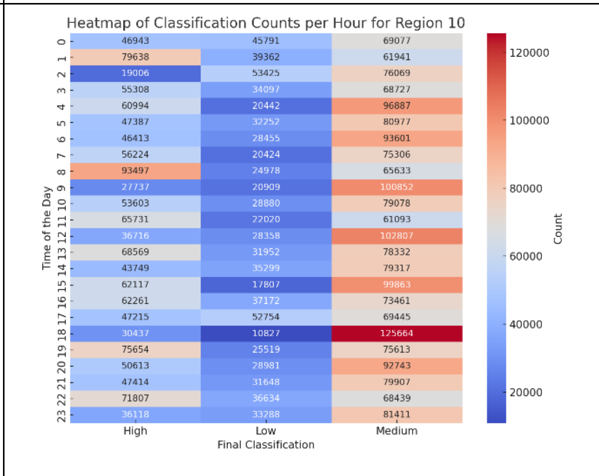
(G)



(H)



(I)



(J)

Figure 5-30: Classification Distribution Time vs Region

Analysing hourly reference counts, current counts' distribution, and classification heatmap explicitly shows each region's density. The peaks, in many cases, their median were high. This indicated that in some instances, the density may increase with the varying fluctuations within the existing range of values of the present count. The classification heatmaps with constant high-density events occurred separately in each region as their temporal patterns. This portrays a complex network between issues, resulting in varied crowd behaviour depending on time and place.

5.13 Algorithm 7: Proactive Crowd Management Paradigm (PCMP) Results

The class diagram in figure 5-22 shows classifying crowd density results using a dataset containing 'Time,' 'Ref_Count,' and 'Current_Count' as records. The algorithm processes these data within every hour, distinguishing the records as either "Above Median" or " The subsequent step involves another quadratic categorization where all records are finally allocated under 'Low,' 'Medium,' and 'High' categories of crowd density. The methodological approach presents a thorough analysis of crowd dynamics focusing on temporal changes of spatial variability over different periods.

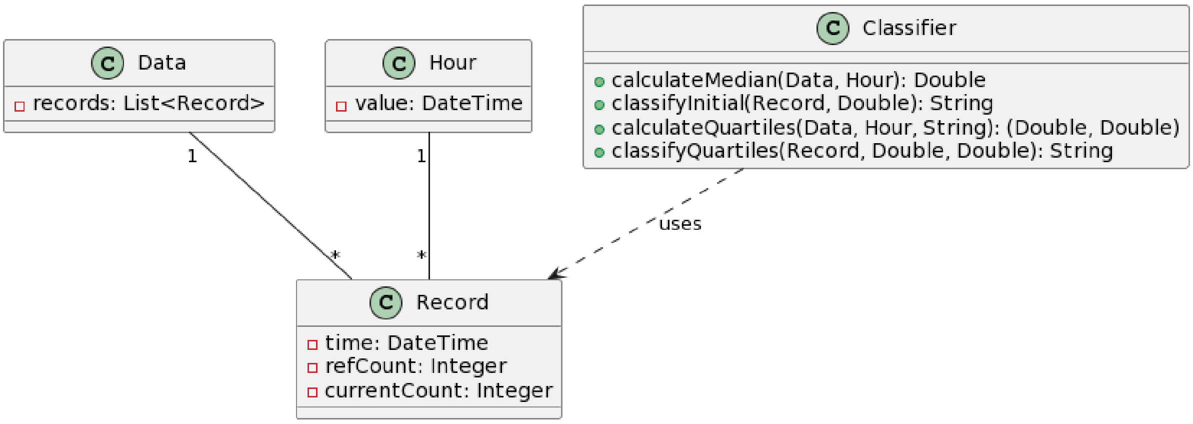


Figure 5-31: Class Diagram Algorithm 7

5.13.1 Historical Threshold

The figure 5-32 shows the hourly thresholds of the median crowd count during the twenty-four hours. Median crowd count changes over time, peaking at an average of about 4,800 minutes before hour 5 and showing significant valleys going down to about 4,200 people between hours 9 and 10. The figure also shows varying crowds since there are other significant peaks around 10 hours, between 15 hours, and after 20 hours.

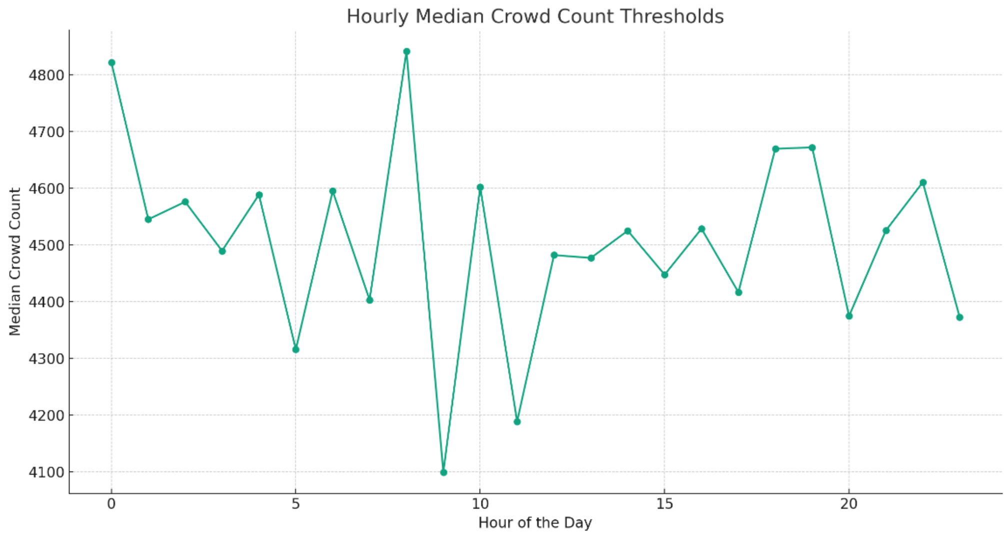


Figure 5-32: Threshold Median - Reference Count

5.13.2 Crowd Density Classification

The figure 5-34 displays the current count's complementary cumulative distribution function (CCDF) for three classifications: Low, Medium, and High. CCDF indicates that the higher count drops for all three classes with the current value increase. Approximately ninety-five percent of the ‘low’ classifications have a count of less than two thousand; above fourteen thousand is the count for most in the ‘high’ classification. This means there are many different values for crowd counts, and the three classifications are distant.

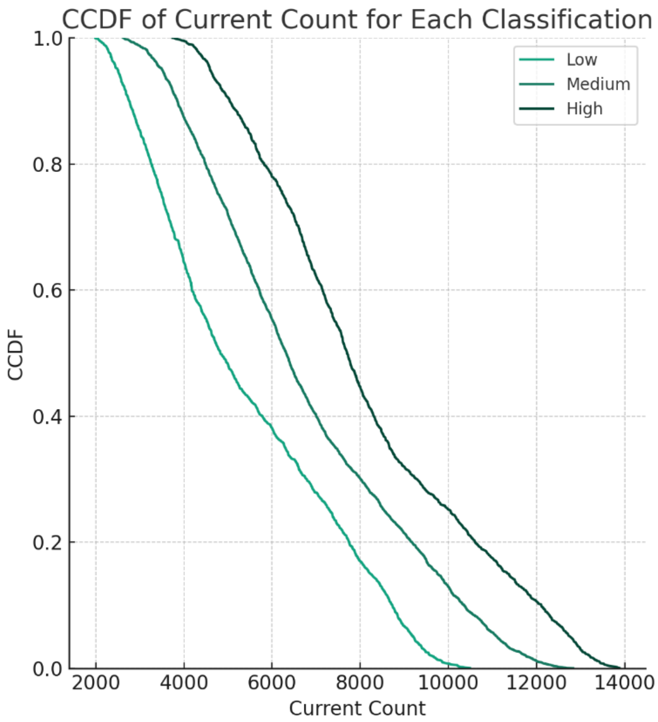


Figure 5-33: Crowd Classification Distribution

The hourly median crowd count chart reveals wide fluctuations in crowd density, ranging from as high as approximately 4,800 to less than 4,200 on varying occasions. The CCDF chart complements this by showing the probability distribution of crowd counts across three classifications: Low, Medium, and High. ‘Low’ mainly encompasses counts below 2000 while ‘High’ extends to more than 14,000, thus depicting distinct crowd density categories. The proposed algorithm effectively captures the changing spatiotemporal fluctuations in crowd density.

5.14 Algorithm 8: Results of Streamlined Crowd Density Taxonomy (SCDT) (Median Threshold)

The figure 5-35 presents the class diagram for the implementations of crowd data D, on which the sophisticated algorithm is applied involving the records containing 'Hour' and 'Current_Count.' It begins with partitioning the dataset into different hours. After that, median thresholds are calculated for these hours, forming a threshold for every entry and further classified as 'above threshold' or 'below threshold.' It also does a quartile ordering of 'Current_Count' values, placing records into corresponding quartiles. These categories and classifications are integrated with the thresholds calculated. Below is a class diagram illustrating the implementation structure of the algorithm.

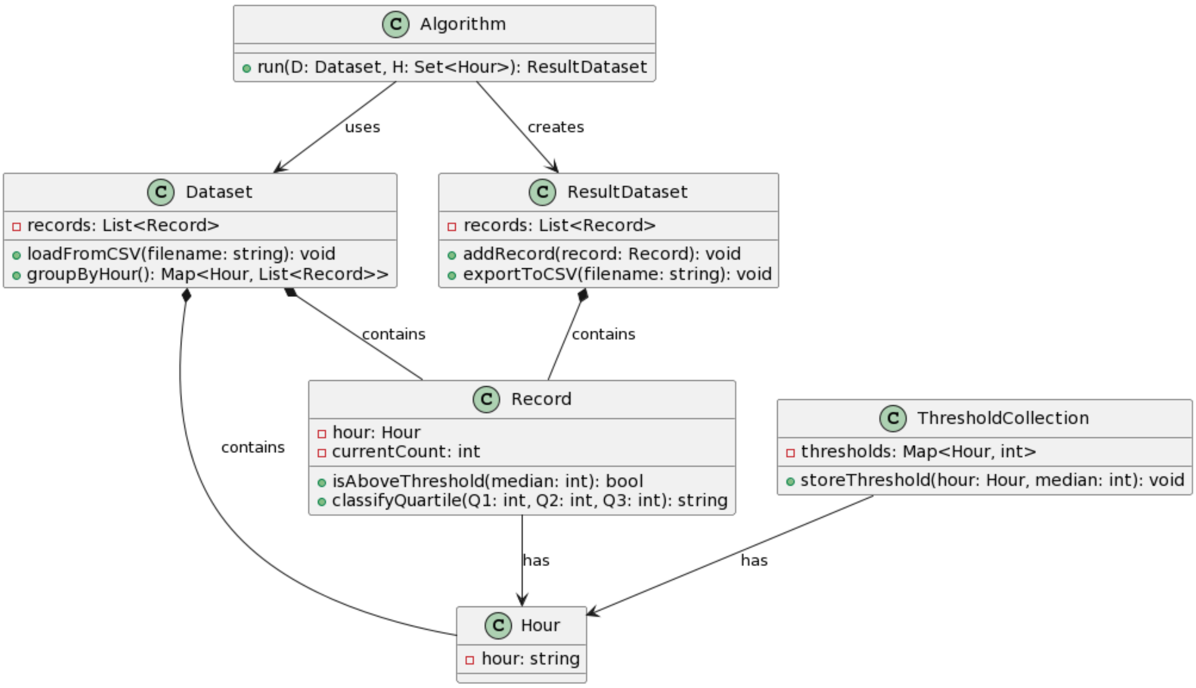


Figure 5-34: Class Diagram for Algorithm 8

5.14.1 Median Threshold Results

The figure 5-36 illustrates the median threshold values over the course of a day. A single point on the graph represents the median 'Current_count' for every hour and sheds light on the changes in these counts over time. This mirrors the daily trends within the data, like hourly peaks and troughs. Varying features in the plot signify that the dataset records various activities and events during different hours (days).

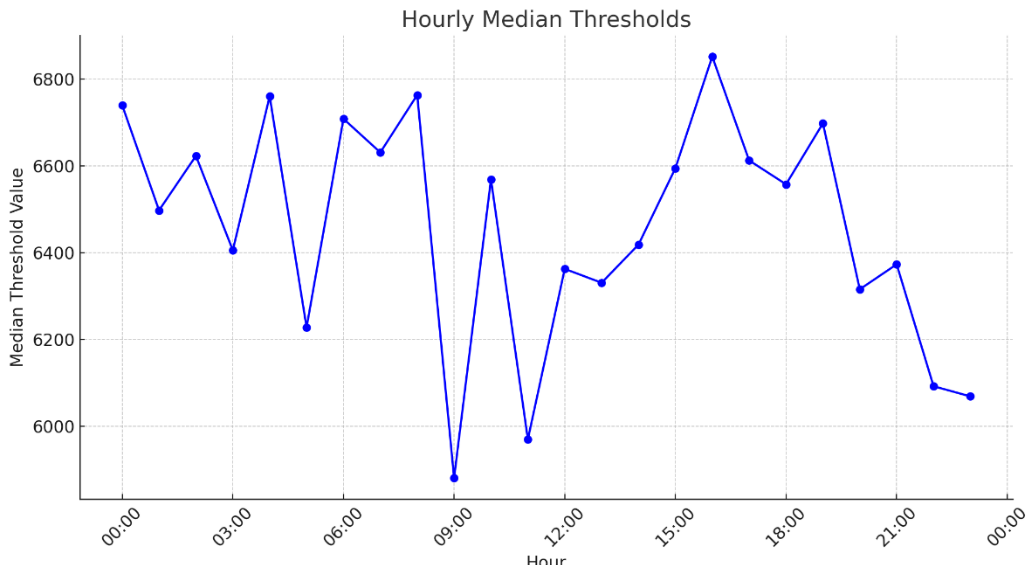


Figure 5-35: Threshold Median Current Counts

5.14.2 Crowd Density Distribution

The figure 5-37 illustrates the CCDF presentation of two-level classification

Above Median: The CDF corresponding to crowds' count over the threshold. The sharper drop in this line shows that high crowds occurred less frequently.

Below Median: Shows the CCDF for crowd counts under the median cutoff value. This curve has a slower slope, signifying an increased evenness of lower crowd counts.

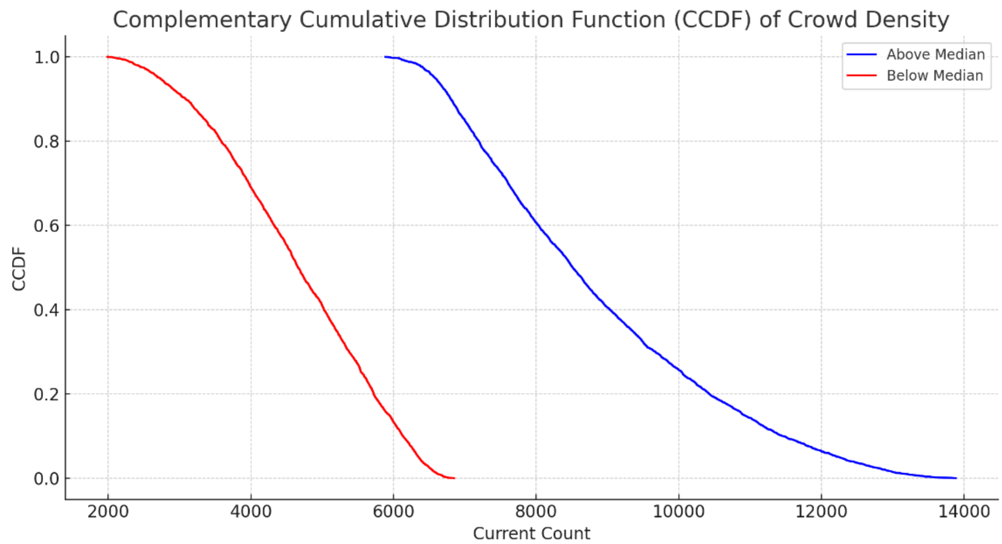


Figure 5-36: CCDF Threshold Crowd Density Classification

5.14.3 Crowd Density Granular Distribution

CCDF plot in figure 5-38 depicts an elaborate description of the crowdedness, highlighting the features in diverse categories. The blue curve shows that the number of people is greater than the median and, therefore, classified as ‘High Density.’ It reveals how often high crowd counts happen in this particular density range. The Cyan curve shows counts above the median threshold classified as ‘Very high density.’ The “Medium Density” graph counts with values less than the median beneath the Median Threshold Curve. The curve illustrates a greater level of consistency in the distribution of crowd counts, which implies that medium-density occurrences were more evenly distributed. The Orange Curve represents the count; though they are under the median level, they still exist in ‘High Density.’ It shows a sharp difference against the higher-than-average low-density numbers, depicting disparities in crowd density beyond mere average levels.

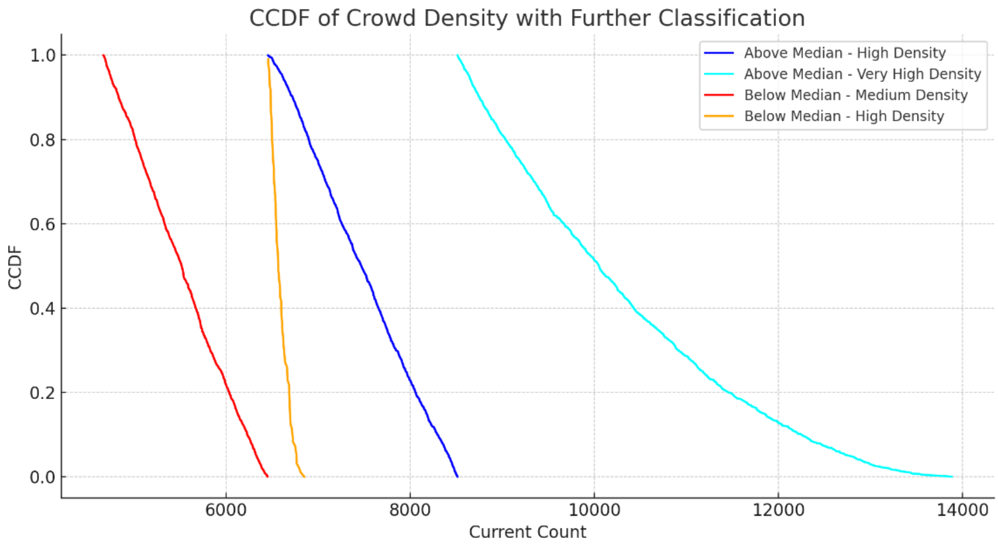


Figure 5-37: Granular Level Crowd Density Classification

The algorithm for processing the dataset and an extensive study of the crowd’s density pattern have been made. The analysis showed that by grouping the data into hourly increments and the thresholds with median, there is a clear distinction of the temporal change in crowd density because the counts were categorized as either above or below these average threshold values. The application of quartile classifications improved the comprehension, as revealed on the CCDF plots, depicting the dispersal of crowd densities among different levels and periods. Adopting this robust process helped to get a multi-faceted interpretation of the data, essential when making informed decisions about urban planning traffic flow, among others.

5.15 Conclusion

The chapter investigates various synthetic data sets for crowd density estimation to replicate real needs related to these growing fields. The created dataset, carefully elaborated to imitate authentic MOBILE information, is one of the main sources used to evaluate crowd density estimation algorithms. This dataset has complex attributes and structures that facilitate comprehensive testing of algorithms in various circumstances, making it an important tool in advancing methodologies in crowd density estimation. These findings contribute to guiding actionable implementation processes that can potentially transform urban planning, traffic management, and emergency responses in an opportunistic environment.

CHAPTER 6: ALGORITHM 14: INTEGRATED CROWD DENSITY ESTIMATION (SYNTHETIC VS REAL MOBILE)

6.1 Real-World MOBILE Data Details

A crucial project was implemented in 2015, where over one hundred million pilgrims were expected during the Kumbh Mela in Nashik, India. I was also allowed to use MOBILE tower log-meta data from eight telecom companies, of which I was the sole innovator. This was a joint project with MIT's Kumbhathon, Nashik Municipal Corporation, and the TRAI that would use the data to prevent possible stampedes and ensure public security. We used latitude, longitude reference count and current count data from every telecom site as a basis for this crowd management analysis. This is the first time that data is being used after 2015. An understanding agreement was discussed to preserve the data for at least a few years.

By 2023, out of 8 telecom operators, only Airtel and BSNL companies are strong in the market until now. Reliance (CDMA + MOBILE) was closed, and later, Reliance Jio was launched as a fresh start. Vodafone and Idea Telecoms are merged, and a new company is formed Vi, Tata (CDMA + MOBILE) is closed for good, Uninor, which was later turned into Telenor, is closed for good and the last telecom Aircel is also closed for good. Thus, the dataset is no longer in violation of any sort. The full data set consists of 1 full day reference taken randomly in June 2015 and later from July to Sept 5 days with + 12 hours each before and later are collected from the most important days of Kumbh Mela. For this research, only 24 hours of data from one of the most crowded data sources is considered. The choice of considering only 24 hours of data is to have macro and micro understanding of real-world MOBILE tower connection log-meta data in context to crowd density estimation and prediction. Here is the comprehensive summary table 6-1 that integrates the key metrics from the MOBILE data set used for the Kumbh Mela project:

Table 6-1: Real World MOBILE Data Metrics

Metric	**Value**
Current Count (Mean)	1084.41
Current Count (Std Dev)	2436.88
Current Count (Min)	0
Current Count (25%)	339
Current Count (Median)	569
Current Count (75%)	1149
Current Count (Max)	61058
DateTime (Unique)	13-09-2015 01:00
Latitude & Longitude (Count)	851
Reference Count (Mean)	1433.81
Reference Count (Std Dev)	1084.92
Reference Count (Min)	0
Reference Count (25%)	568
Reference Count (Median)	1182
Reference Count (75%)	2142
Reference Count (Max)	4878
Site ID (Count)	1514
Telecom (Count)	8
Telecom (Top Subscription)	AIRTEL

6.2 Input Data Format

The table 6-2 presents the input dataset is structured as a collection of data points, each capturing specific information related to a telecom site. Below is the tabular representation of the dataset:

Table 6-2: Real-World MOBILE Data Presentation

Data Point	Latitude (L_i)	Longitude (Lo_i)	Reference Count (R_i)	Current Count
1	L_1	Lo_1	R_1	C_1
2	L_2	Lo_2	R_2	C_2
3	L_3	Lo_3	R_3	C_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$N-1$	L_{N-1}	Lo_{N-1}	R_{N-1}	C_{N-1}
N	L_N	Lo_N	R_N	C_N

Description:

- Data Point: Each row in the table represents a unique data point corresponding to a telecom site.
- Latitude (L_i) : The latitude of the site's geographical location.
- Longitude (Lo_i) : The longitude of the site's geographical location.
- Reference Count (R_i) : The reference count associated with the site provides a reference metric.
- Current Count (C_i) : The current count represents the site's crowd density during a specific observation.

6.3 Descriptive Analysis (Real MOBILE Data)

Figure 6-1 presents the Distribution of Reference Count: It would seem that the distribution of Reference Count is multi-modal in several peaks, like in the normal crowds of different types of sites and events located on the territory of the Kumbh Mela.

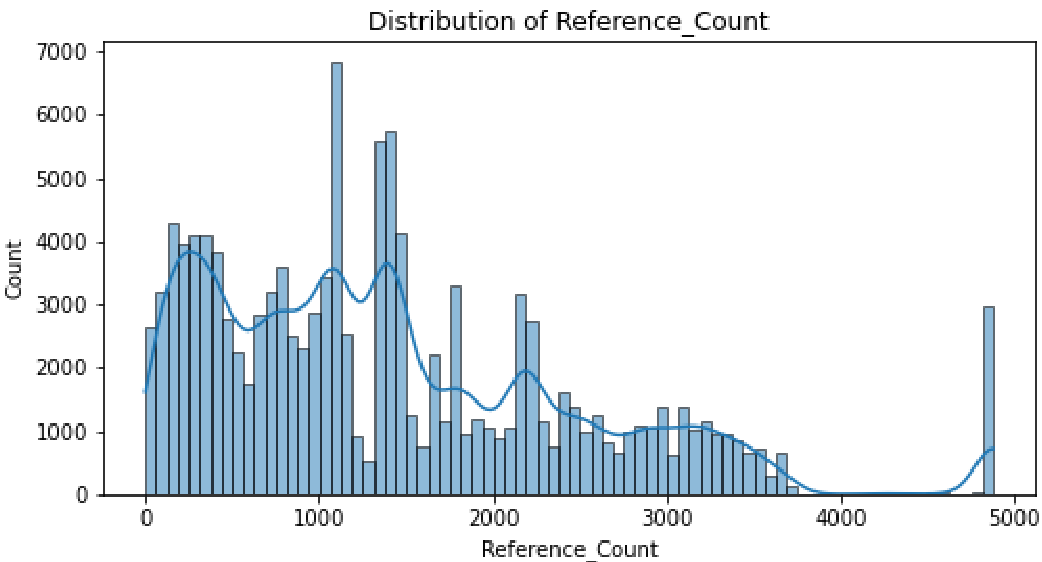


Figure 6-1: Reference Count Distribution Real MOBILE Dataset

The figure 6-2 illustrates the Distribution of Current counts: A right-skewed histogram indicates that most telecom sites have lower current counts while others have high current counts. The dataset is mostly collected from routes, highways, and the actual riverside areas where the crowd is expected. However, apart from the Kumbh area, the highly crowded location is the city's local crowd who communicate around.

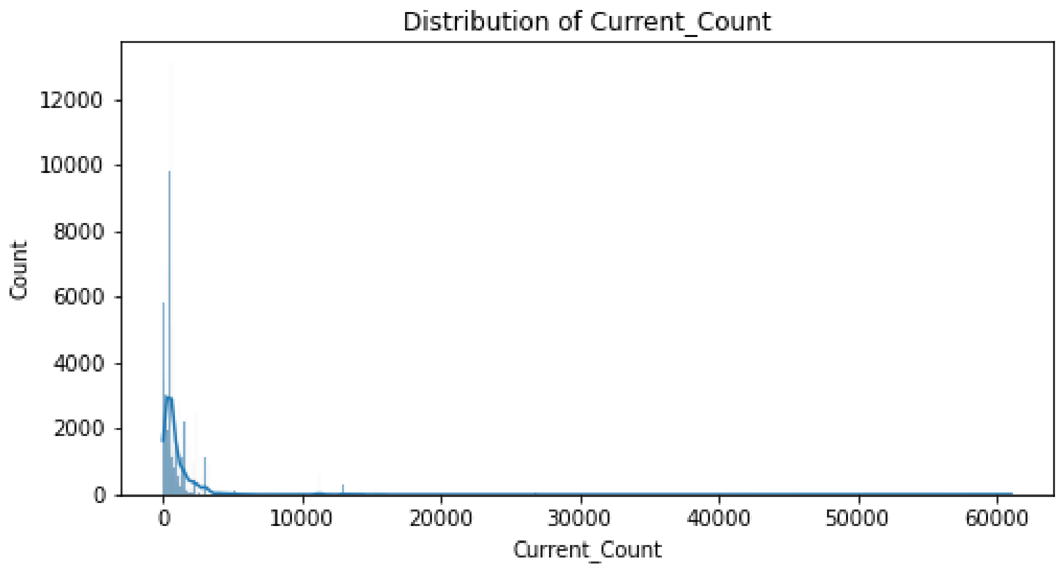


Figure 6-2: Distribution of Current_Count Real MOBILE Dataset

The figure 6-3 illustrates the Distribution of Floating Count: Like the Current Count, the histogram on the right side of the figure is strongly right-skewed, implying that many areas have low floats, which may be due to some of the transient population. These are mainly the people passing by the city or who have entered the city, completed their rituals at the riverside, and left the city the same day. Floating crowd allows us to map the steering of the crowd within the given location and across the city.

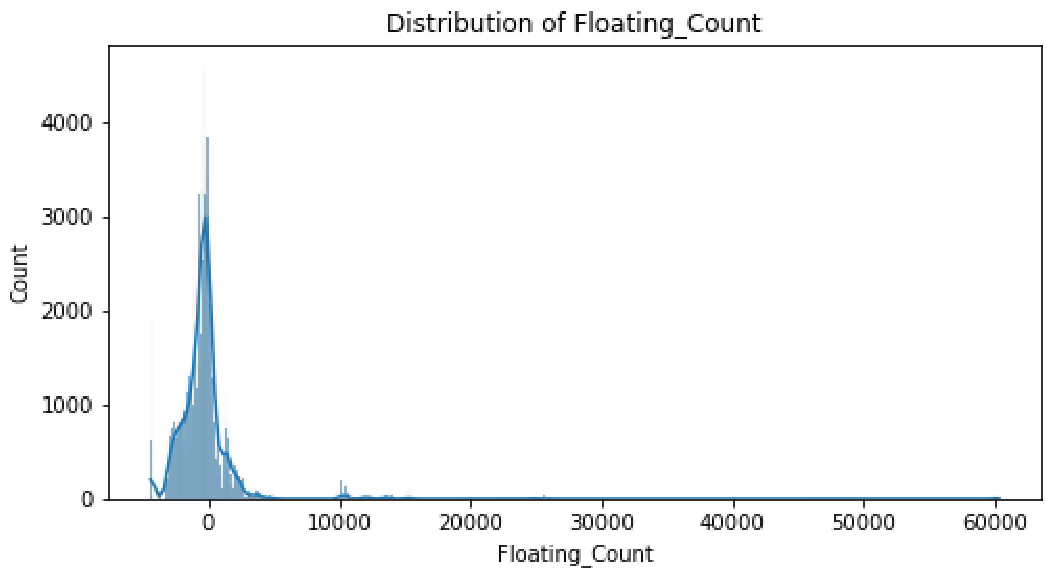


Figure 6-3: Floating Count Distribution of Real MOBILE Dataset

Figure 6-4 illustrates the Local Geospatial Distribution of Sites: A scatter plot of the latitudes and longitudes of the telecom points, which also helps to draw the order to reveal clusters that could indicate regions of high population densities or popular places within the Kumbh Mela. This local geospatial distribution helps mark the paths of crowds entering the city and collecting in the center.

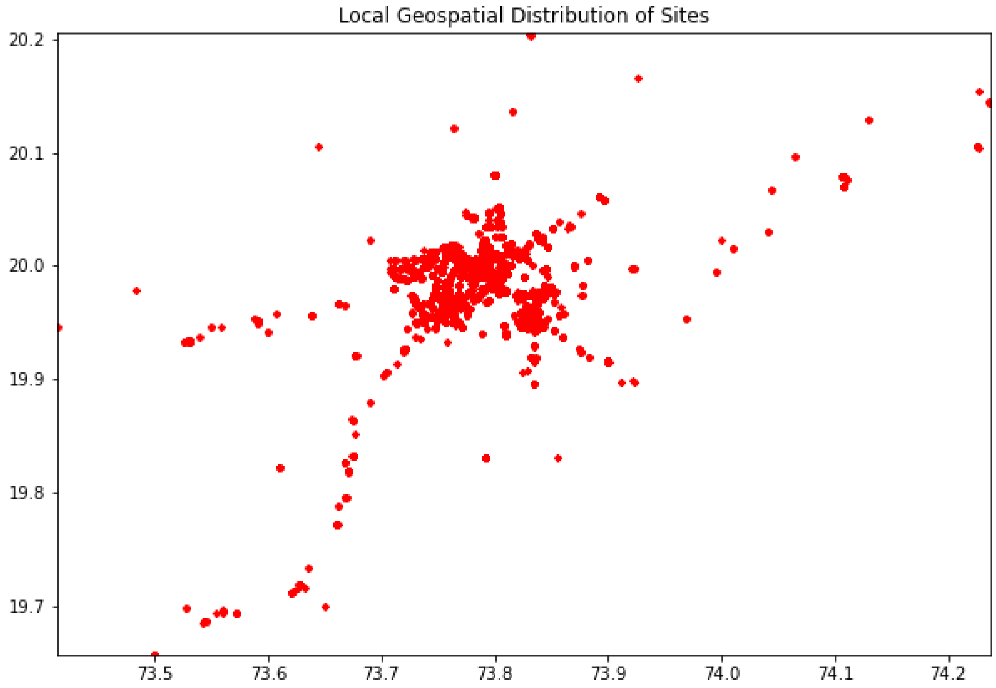


Figure 6-4: Geo Spatial Distribution Real MOBILE dataset

The figure 6-5 below Pairplot with Clusters: The complex plot comprises many subplots that pair and compare certain parameters such as time, number of records, reference count, and geospatial data. Different colored clusters indicate that data points sharing common features have been arranged to reveal trends in crowd dynamics.

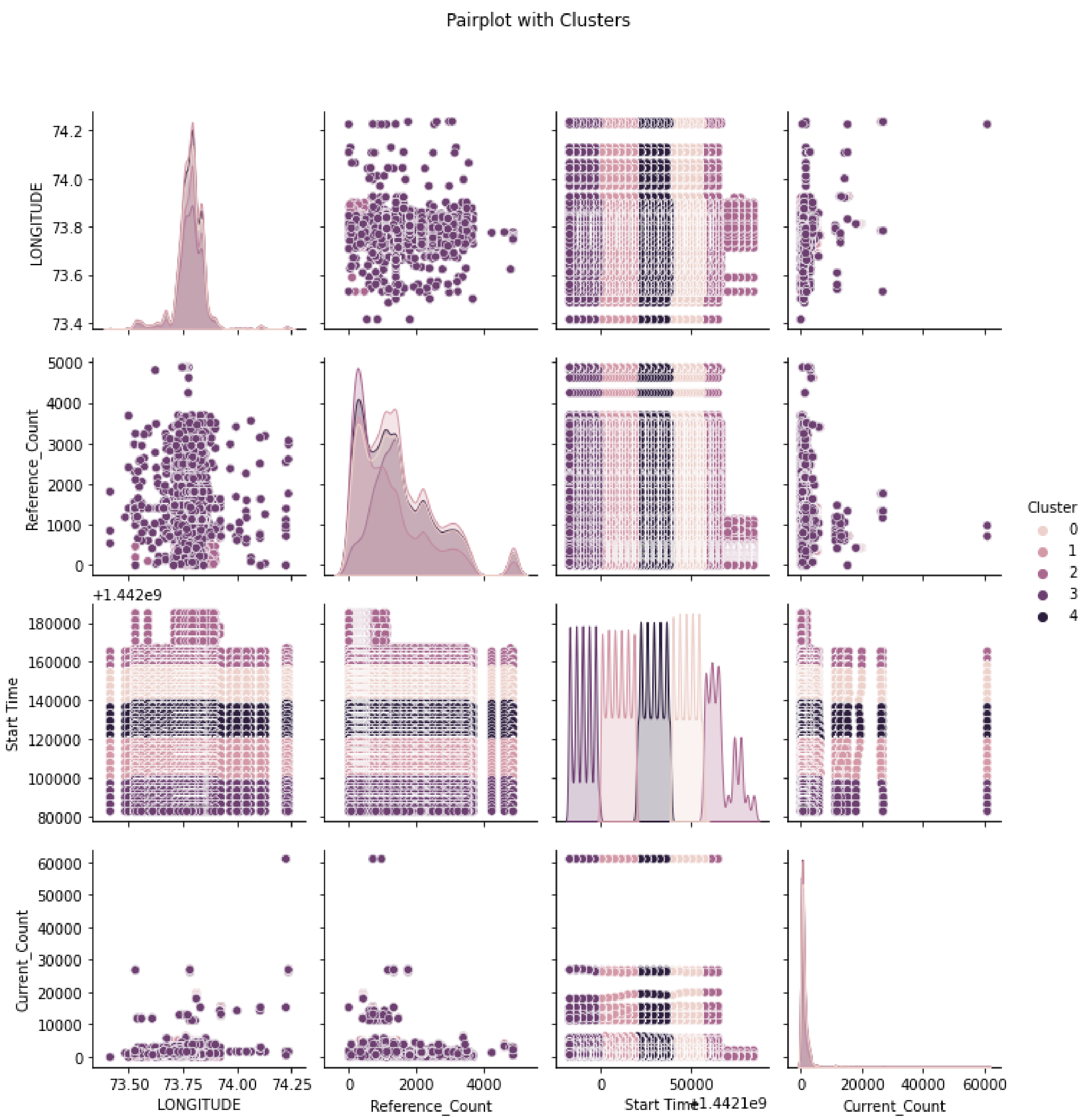


Figure 6-5: Paired Cluster Distribution Real MOBILE Dataset

6.4 Outliers (Real MOBILE)

The figure 6-6 illustrates the Boxplot "Hourly Outliers by Location," which describes the distribution of current counts declared by MOBILE towers across the day. The boxes represent the IQR, while the markers point out the outliers. Two outliers are noted: Always Outliers in red, possibly persistent high-traffic spots, and Never Outliers in Black, potentially occasional outliers, happening elsewhere at a different hour. The presence of some outliers shows that specific places and times recorded unusually high crowd congestion levels and call for special management solutions. This is the first point where this research felt important; when only crowd density was applied in real-time analysis, there were different areas apart from the Kumbh spot, which would show red in the heatmap.

The most expected areas are kept under surveillance when surge crowds enter the city via buses and trains. There is the resident crowd of the city who has to commute to the office or elsewhere. When the main roads were blocked only for the Kumbh pilgrims, the residents took inner routes and alleys running between and connecting several residential areas, and those areas were shown traffic jams. That is when all the crowd density monitoring algorithms failed as they would normalize the data or pick up the average number. When the dataset is skewed to the right or left in the real world, how can a normal distribution be expected to do justice? The same was true with many high and low peak fluctuations, taking an average value as the threshold and giving bias classification. That is how the median and median-of-median concepts were brought into the picture for more reliability on the density classification.

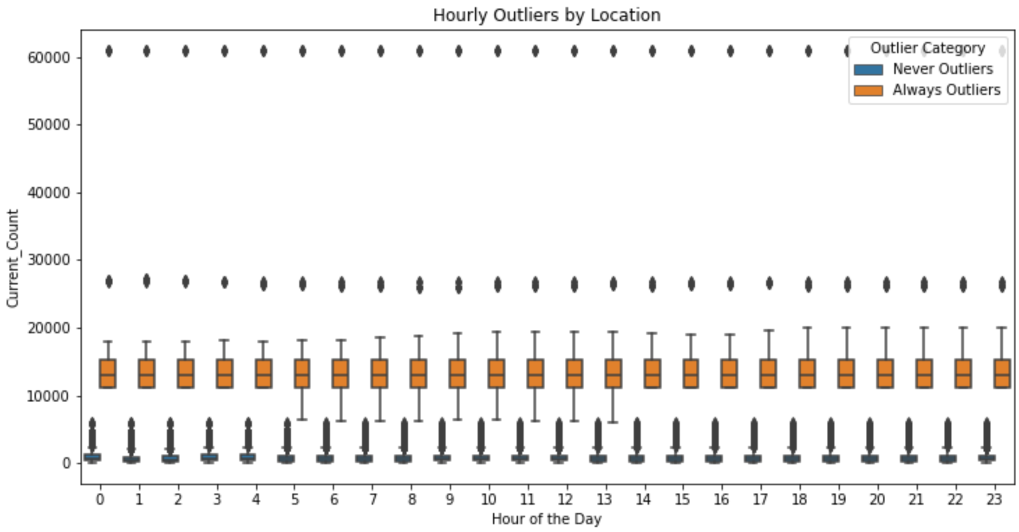


Figure 6-6: Hourly Outlier Distribution (Real MOBILE)

6.5 Algorithm 14: Insights & Comparison (Synthetic vs Real MOBILE)

Figure 6-7 the class diagram, that is a process of breaking down the data set through several parameters. First, the algorithm calculates median thresholds every hour before determining a median-of-medians threshold. These counts are categorized according to this M-o-M threshold and subdivided into quarters. The prediction model employs current activity weights, historical density weight, and threshold comparison for expected counts. All this information is collected and exported to a CSV file, giving one an analytical overview of the collected data for informed decisions.

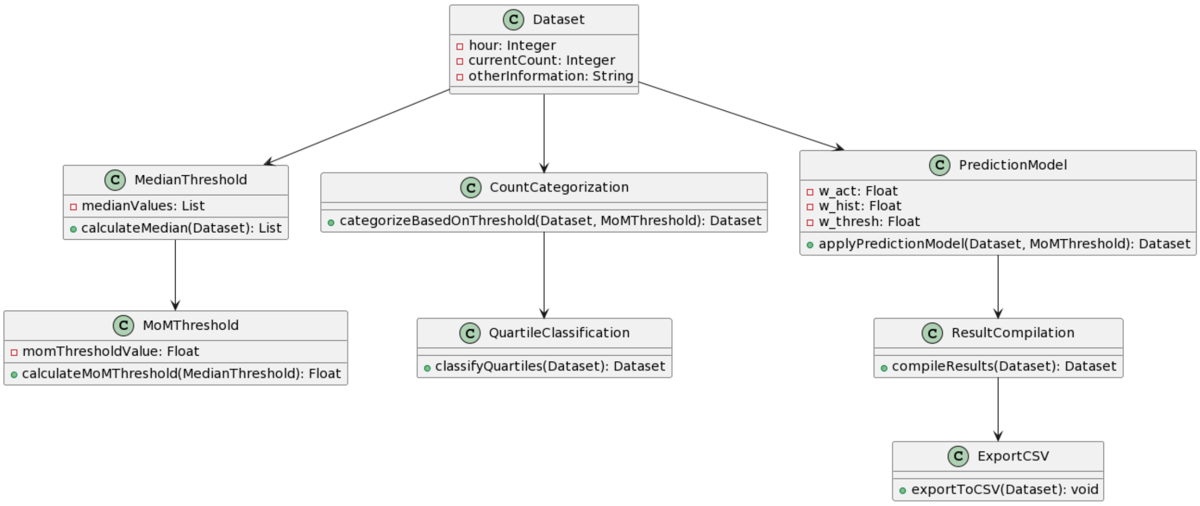


Figure 6-7: Algorithm 14 Class Diagram

6.6 Median-of-Median Threshold (Synthetic vs Real)

Figure 6-8 presents that a great difference is seen in the threshold values of synthetic and real MOBILE data, as displayed by the box plots. The threshold of the synthetic data is narrowly clustered into a lower interquartile range, showcasing less variation and possibly fewer elaborate patterns in data. However, the thresholds of the real MOBILE data's values are much lower but with a wider band spread that implies great variety, showing a complex dataset. The variance observed across real-world data may yield more meaningful inputs into prediction analyses since such data is likely better to reflect actual behaviour patterns than controlled research. The same argument holds true about outliers as part of real-world data, but these could be extreme ones, which are equally important for testing analytical methods.

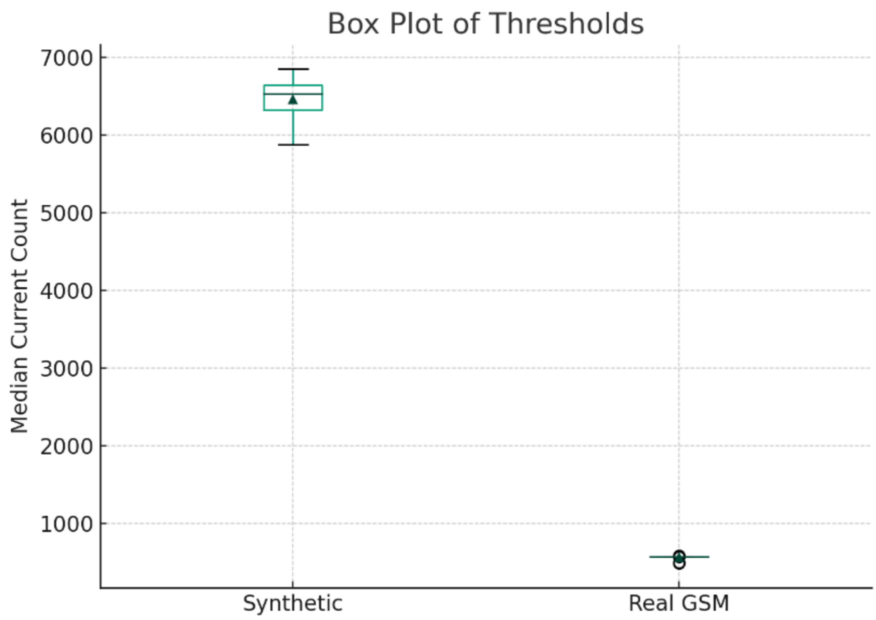


Figure 6-8: Median Threshold Comparison - Synthetic vs Real MOBILE

6.7 Classification (Synthetic vs Real MOBILE)

Figure 6-9 and Figure 6-10 presents a Synthetic and real MOBILE graph showing the relation of ‘Ref_Count’ versus ‘Current_Count’ within Quartiles. In the synthetic data, there is a straight line in which closely clustered points reveal the connection of reference and present counts. However, the real MOBILE data appears randomly distributed, especially in higher quartiles, demonstrating variability and potential underlying complications. The true MOBILE data spread indicates various cases and offers more substantial benefits in predicting models and trend analysis. It is worth noting that there are large variations in ‘Current_Count’ in the highest quartile of the real MOBILE data for similar ‘Ref_Count’ values as can be expected.

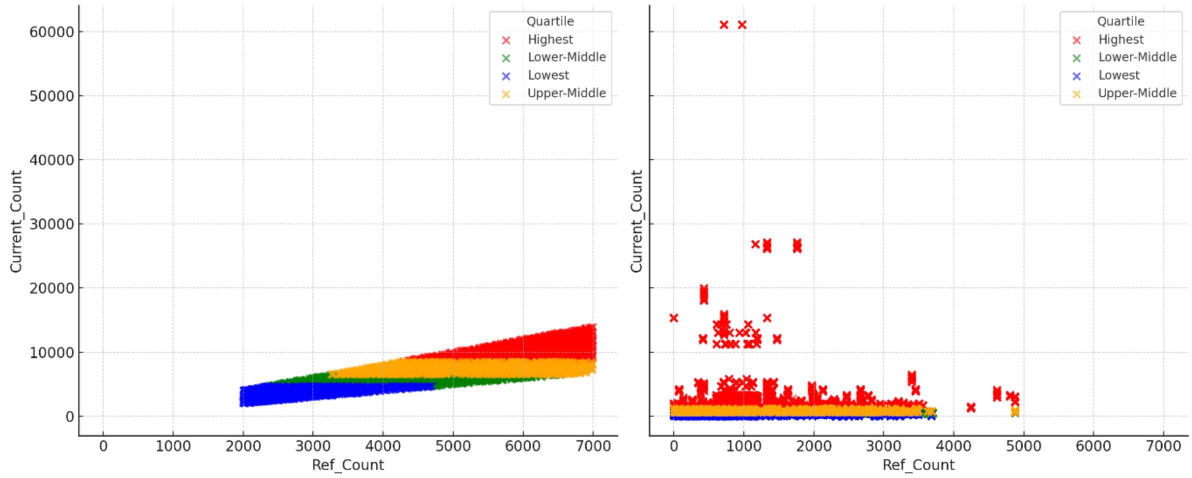


Figure 6-9: Crowd Density Classification Comparison - Synthetic vs Real MOBILE

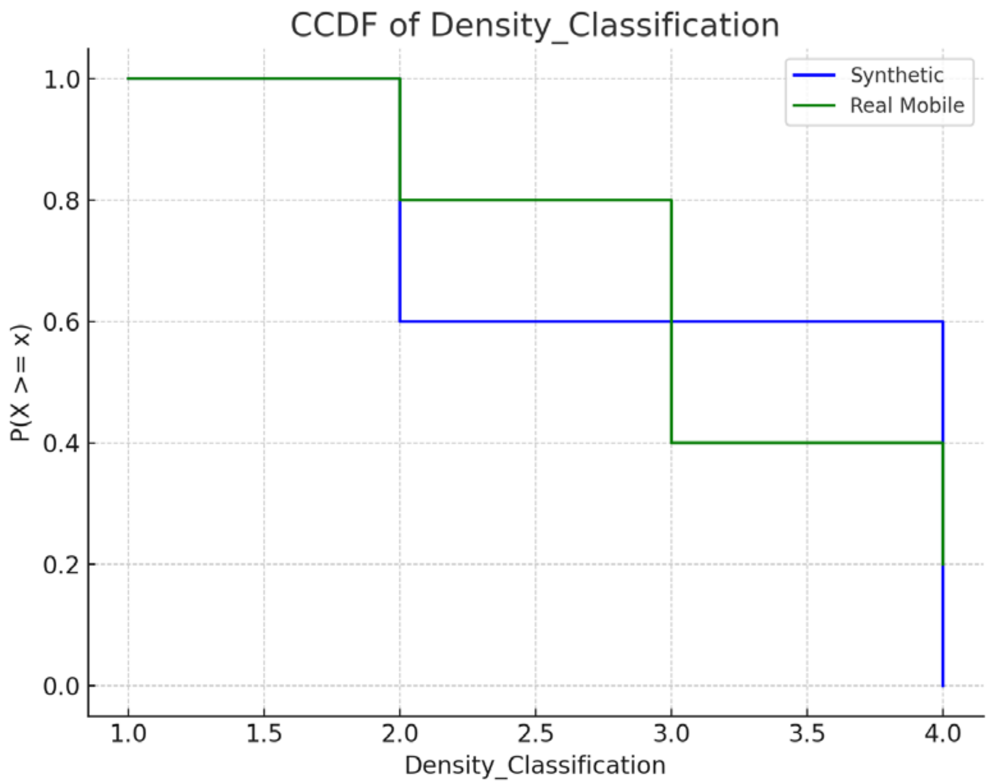


Figure 6-10: Density Distribution (Synthetic vs Real)

6.8 Datasets (Synthetic vs Real)

CCDF plots from Figure 6-11 ((A), (B) (C) and (D)) compare 'Ref_Count' vs. 'Current_Count,' 'Floating Count vs. 'Prediction,' and 'Real MOBILE dataset vs. Synthetic MOBILE Dataset.' Synthetic data demonstrates a steeper fall in the CCDF curve, reflecting higher values within smaller intervals. On the other hand, the graph of real MOBILE data gives rise to a less steep tendency that implies more dispersed points scattered at different ranges of values. A real MOBILE dataset that covers a broader extent of counts and predictions is likely to depict real-world variability and, as such, gives better results in analysis where a total understanding of data phenomenon across different scenarios, including the uncommon ones, must be considered.

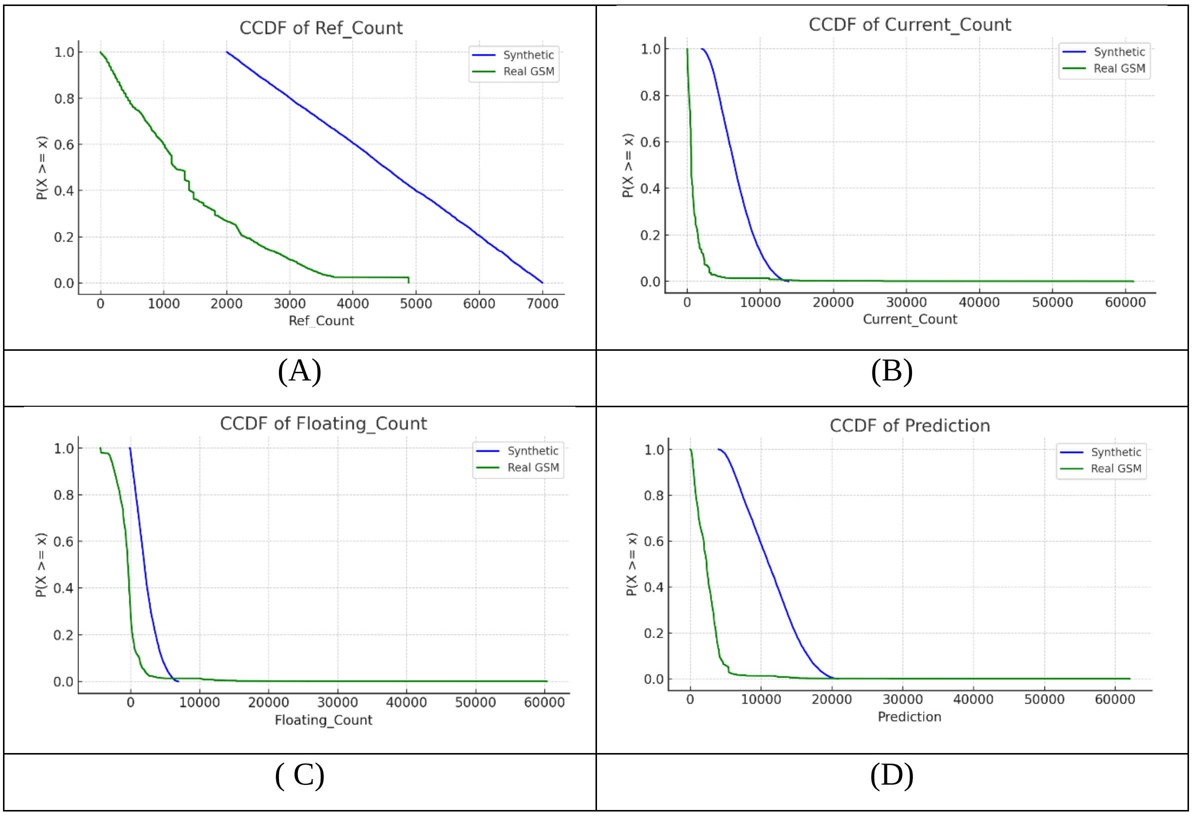


Figure 6-11: Data Counts & Prediction Comparison - Synthetic vs Real MOBILE

6.9 Conclusion

Kumbh Mela in Nashik provided vital information regarding crowd management and dynamics. The project identified crowd density patterns and possible congestion points by looking at various metrics, including reference count, current count, and geospatial distribution. Such made it possible to distinguish ordinary crowd behaviour from outliers, allowing for effective crowd control techniques. Thus, these innovations indicate that telecom data can be utilized to manage major functions during high-profile activities and improve public security.

CHAPTER 7: INDIVIDUAL AND GROUP MOBILITY RESULTS

7.1 Individual Mobility Dataset

The next research stage is to identify individual and group mobility patterns in the given crowd. For this section of the result, the dataset structure in table 7-1 has been changed from cumulative counts to Individual mobility traces. The nature and structure of the data are discussed in this section. The dataset summary indicates 1,200 entries with 50 unique individuals and 10 unique sites. The most frequently occurring individual in the records is 'Individual 1', and the most common site is 'SiteID 2'.

Table 7-1: Individual Dataset Summary

Stats	**IndividualID**	**Time**	**SiteID**
count	1200	1200	1200
unique	50	24	10
top	Individual 1	00:00:00	SiteID 2
freq	24	50	133

7.2 Data Distribution

The figure 7-1 exhibits uniform distribution, meaning every one out of 50 individuals has 24 records in this dataset. This means that the data covers a full 24-hour period for every individual, which is perfect for investigating daily mobility regimes.

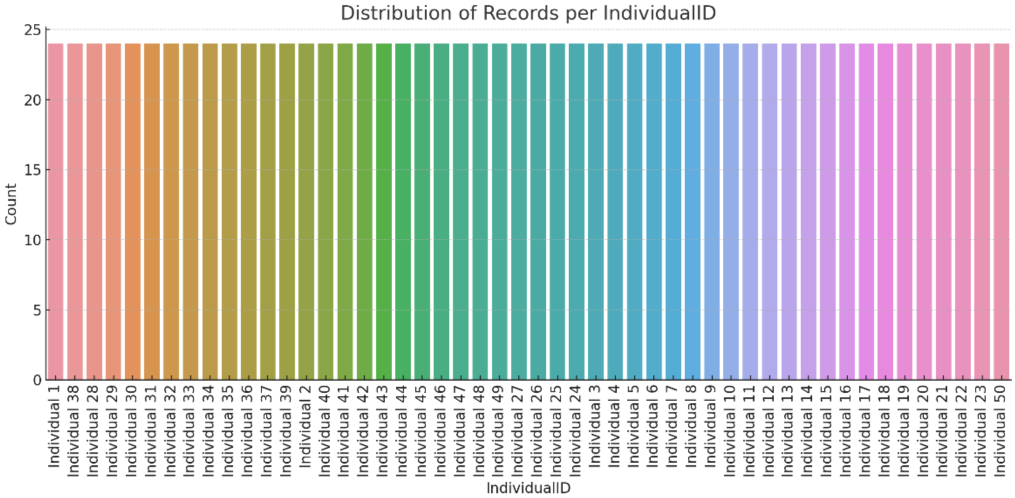


Figure 7-1: Distribution of records per Individual ID

The figure 7-2 shows that visitors were distributed among the sites fairly. SiteID 2 showed the highest count at more than 120 visits, while siteID 1 had only close to 100. This shows that variations exist in the popularity of the sites, but the difference is not very great on the different sites.

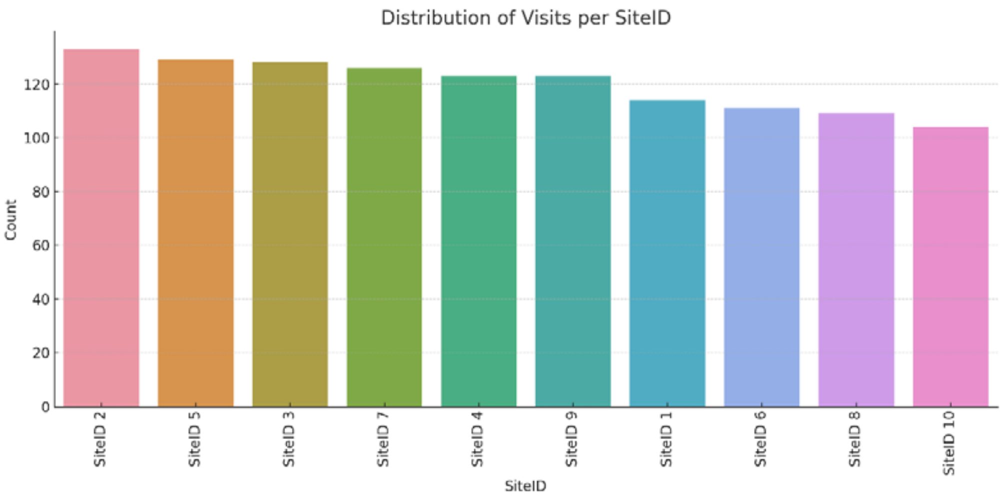


Figure 7-2: Distribution of Visits per SiteID

7.3 Structure of Data

This dataset presented in table 7-2 consists of several records, each uniquely identifying an individual position at a specific time. The data is organized into three primary columns: The unique ID of the individual, ‘time’ referring to the time when their locality was recorded, and the site ID. This entails distinct site connectivity for each row and paints a complete picture of movements between sites across time.

Table 7-2: Individual Mobility Data Strucutre

Index	**Notation**	**Time**	**SiteID**
1	IndividualID_1	Time_1	SiteID_1
2	IndividualID_2	Time_2	SiteID_2
3	IndividualID_3	Time_3	SiteID_3
...	...	...	...
n	IndividualID_n	Time_n	SiteID_n

The input dataset contains an extensive history log of individual movements that their specific identifier codes have tagged and time stamps as observed in subsequent site visitations. Due to this structured and uniform format, it becomes possible to analyze the mobility patterns of population movements over time and predict future location trends using historical information.

7.4 Algorithm 9: Individual and Social Dynamics Integration (ISDI)

The results obtained by implementing the algorithm diagrammatically presented in Figure 7-4 through the algorithm is analyzed in detail by providing insights into the patterns of mobility and social relations among persons. First, it involves investigating mobility profiles showing vital information regarding movement trends and most sought-after destinations for the analysis duration. The processes through which groups emerge are then analyzed, highlighting their effect on developing socio-cultural and spatial ties. The predictive models also used for forecasting future mobilities and social interactions are evaluated critically, demonstrating their applicability for disclosing complicated human behaviour. Such a complete analysis reveals more about social dynamics, thus providing avenues for future works in urban planning or social network analysis.

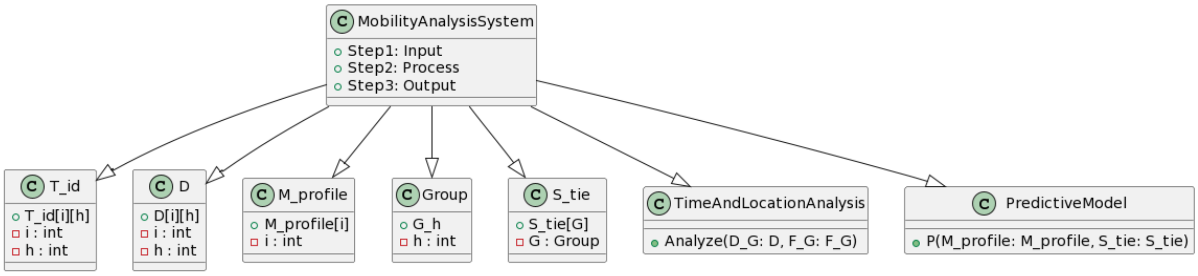


Figure 7-3: Class Diagram for Algorithm 9

7.4.1 Threshold – Median

Figure 7-4 depicts the median half-hour readouts of MOBILE cell towers throughout a day/24-hour period. Some significant spikes appear at various points and exceed 5.5, which may indicate times of increased movements or gathering. However, the largest trough appears to dip below 4.5, portraying a low linkage or activity timeframe. Such data display considerable variation; sometimes, certain hours stay close to five. This can result from the subjects' daily activities and movement patterns, such as beginning or ending a work day or nighttime rest.

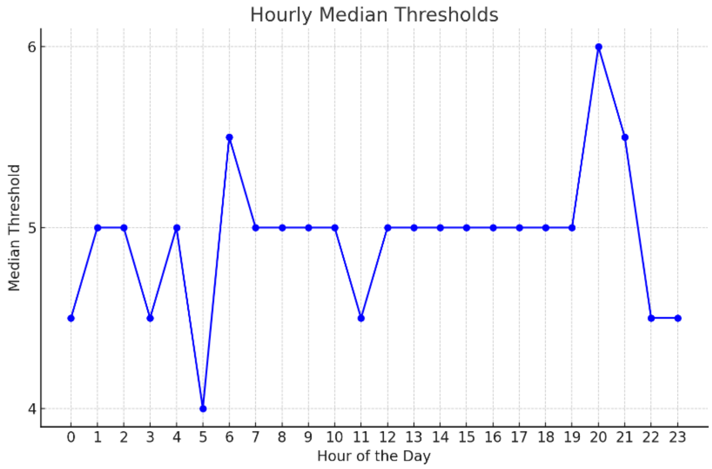


Figure 7-4: Hourly Median Threshold

7.4.2 Individual Network Connection

The figures 7-5 show a network diagram in which the most important links among people are represented according to their joint visiting of sites. Nodes represent individuals, while edges show that many people made many visits between themselves. The network layout seems to have been a relatively interconnected group, with the visits made by some of its members being identical, as shown by a multiplicity of links connecting the members of this network layout.

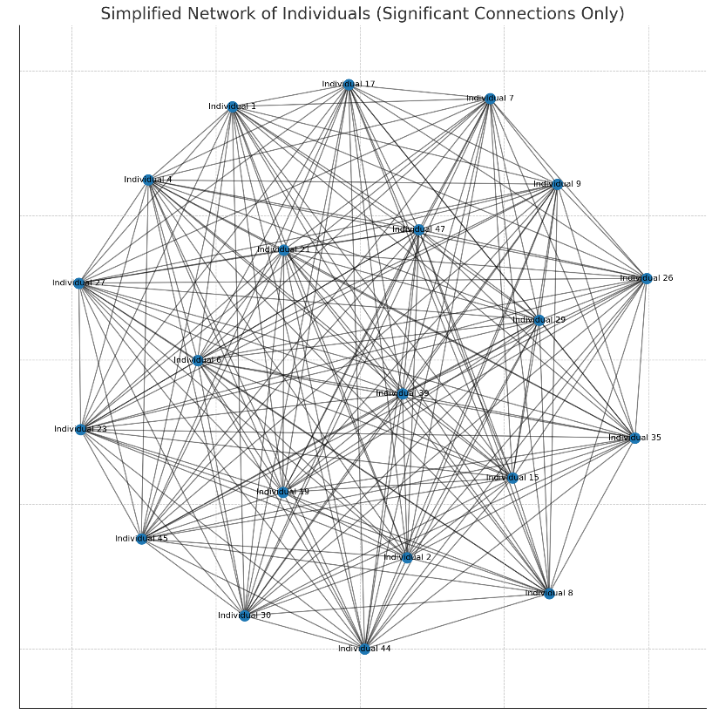


Figure 7-5: Simplified Individual Network Connects

7.4.3 Individual & Group Visit

The Figure 7-6 for "Individual 1" regarding the frequency of accessing the various sites. On the graph, the x-axis shows the site IDs, while the y-axis corresponds with the frequency of visits. "Individual 1" seems to have visited SiteID 1, SiteID 4, SiteID 10, and SiteID 5 thrice a piece. Sites ID 8 and ID 9 have received a single visit each.

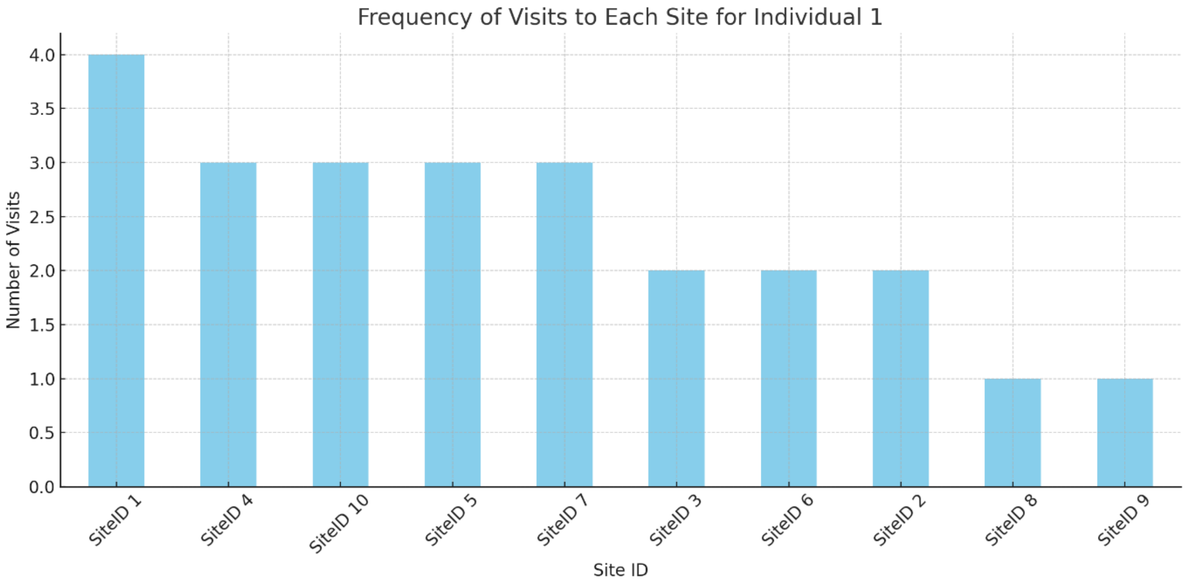


Figure 7-6: Frequency of Visits to each Sites

7.4.4 Prediction Model

With successfully tracked each individual's connection to different MOBILE towers during the 24 hours to outline their general route through space. This computed the hourly median thresholds of power law distributions for tower connections and used this as a baseline for normal movement behaviour within the dataset. A complete analysis of individual spatial-temporal patterns was undertaken using the constructed data matrix D and the mobility profiles M_" profile" [i]. Based on the profile, the prediction model is built, and the Figure 7-7 below illustrates the predicted mobility pattern.

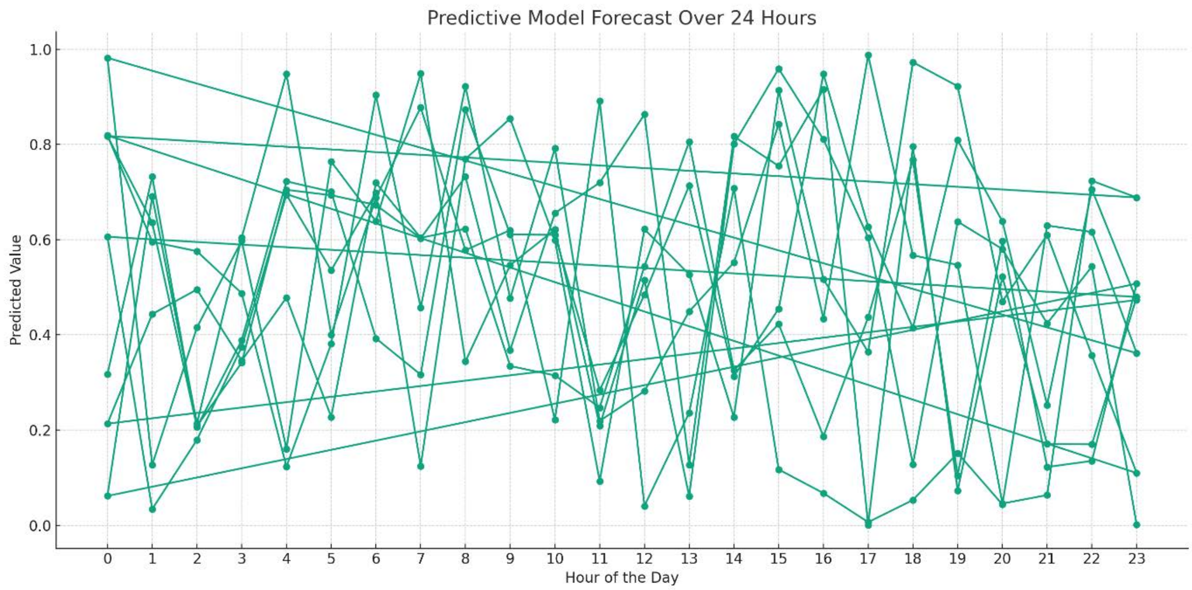


Figure 7-7: Predictive Model forecast over 24 Hours

Clusters G_h, formed by intersections in common locations during certain times, could be identified and deduced. Since these groups were formed in real physical spaces through common tower connections, they likely reflected shared social ties or group behaviours. It was possible to estimate the strength of social connection among various groups based on "Site[i]." This took into account the dynamic character of socioeconomic linkages. The temporal/spatial analysis provided useful information regarding how long and how often members of each group, G, came together to ascertain social solidarity and the nature of interactions within that group. The study used predictive modelling, which incorporated mobility profiles and the strength of the social ties, to predict future mobilities and their interactions. This predictive ability might be helpful for issues of urban development, road control, and healthcare in advance, especially by providing estimates about future demands on infrastructure and servicing. The analysis described the mobility patterns and group dynamics and created a model on which forecasting of subsequent behaviours would be based. Decision makers can use this framework to assist them with resource planning and allocation, while researchers can use it further to understand human movement patterns and social structure patterns.

7.5 Algorithm 10: Dynamic Urban Crowd and Social Interaction Model (DUCSIM)

A dual analysis approach uses MOBILE tower data, crowd counts, and individual tracking data to examine urban dynamics in this research. The class diagram figure 7-8 shows the implementation steps of Macroscopic investigation is into crowd density, using daily and weekly thresholds and quadrants. On the other hand, microscopic analysis focuses on micro-level mobility and interpersonal practices among individuals. This combines the two approaches, allowing one to develop predictive models for future urban crowd dynamics and social behaviour patterns. The implementation class diagram is illustrated below:

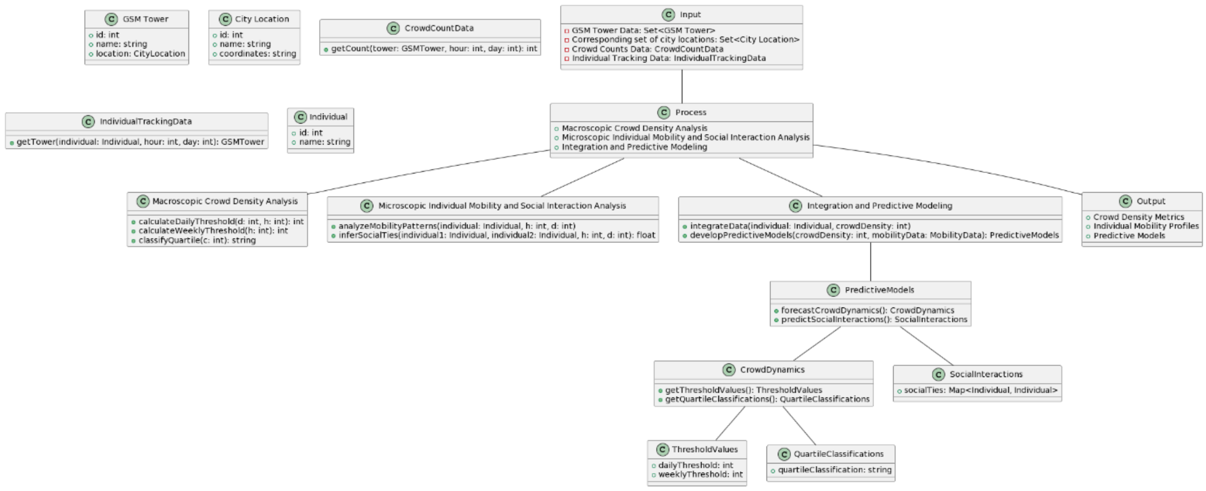


Figure 7-8: Class Diagram Algorithm 10

These findings offer a comprehensive perspective on urban mobility and urban rhythms. The daily average and weekly thresholds were also determined as crowd density measurements, along with creating a quartile classification. They also mapped individual mobility profiles that reveal people’s movement patterns and social networks. The predictive models formulated from the amalgamated mass of data can correctly forecast future urban crowds and social relations.

7.5.1 Hourly Crowd Density Thresholds

Figure 7-9 presents the Median crowd counts exhibit variations across the day and considerable spikes at some points. This suggests that there are occasions when populace intensity exceeds normal and could play a vital role in urban planning and resource distribution.

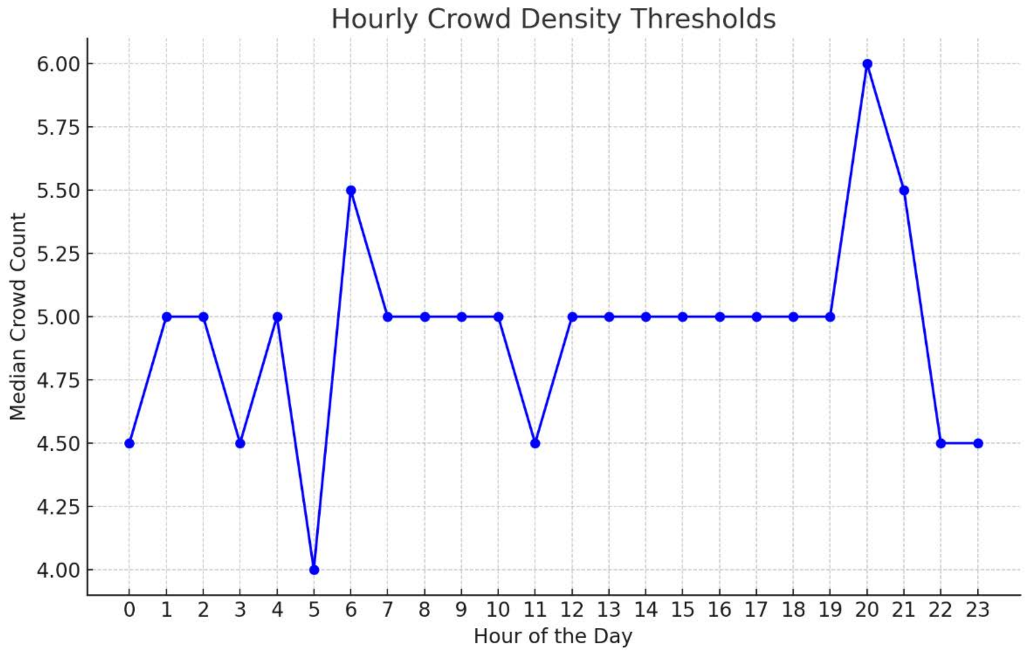


Figure 7-9: Hourly Crowd Thresholds

7.5.2 Crowd Count Heatmap

This heatmap showed in figure 7-10 is about where and when crowds were counted in specific places on certain days. Crowd counts tend to be higher during some hours, implying the time of highest activities. SiteID 3 displays greater counts, especially during the 8th and 15th hours, pointing to a hot spot.

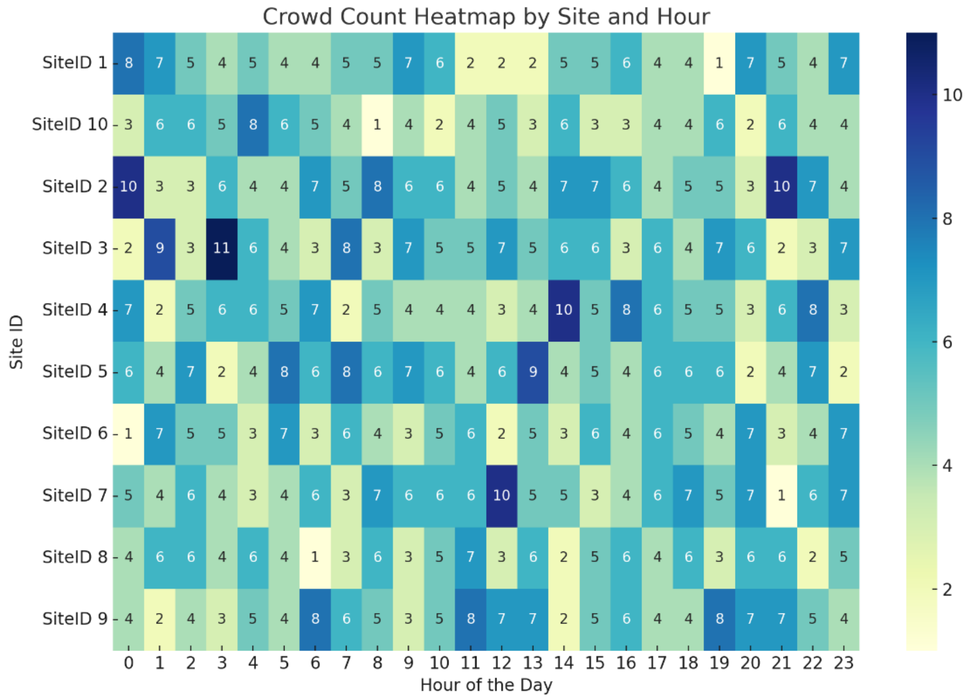


Figure 7-10: Crowd Density Heatmap SitID vs Hours

7.5.3 Social Ties Network

The complex network diagram figure 7-11 shows a high density of such social ties, with some people being central and connected to many others. The complicated nature of society's relations and the possibility of broad circulation in the network have led to this complex issue.

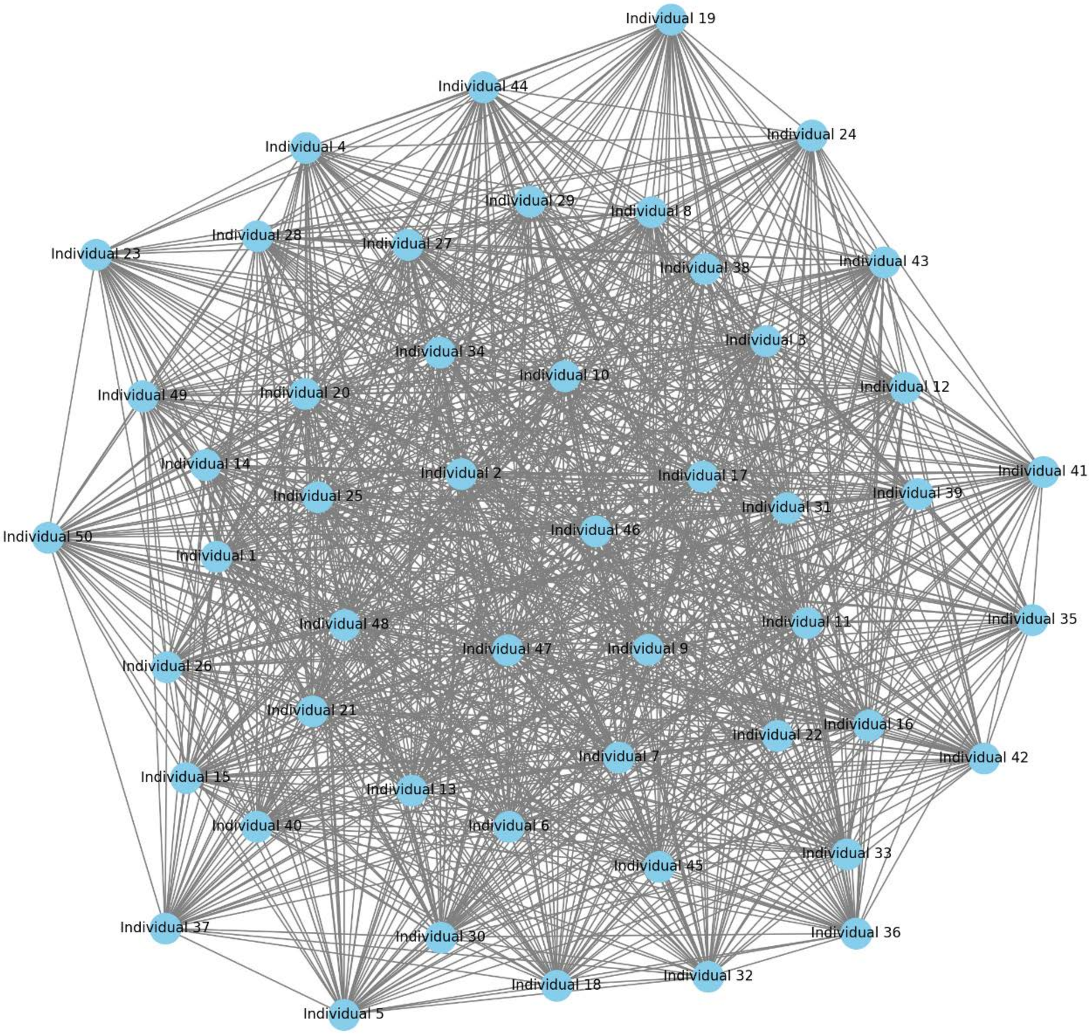


Figure 7-11: Complex Network Connection of Individuals

7.5.4 Simplified Social Ties Network

Figure 7-12 presents the individuals have different levels of connectivity in a simplified network chart. They have several links to people who seem like important players in social bonding and information diffusion.

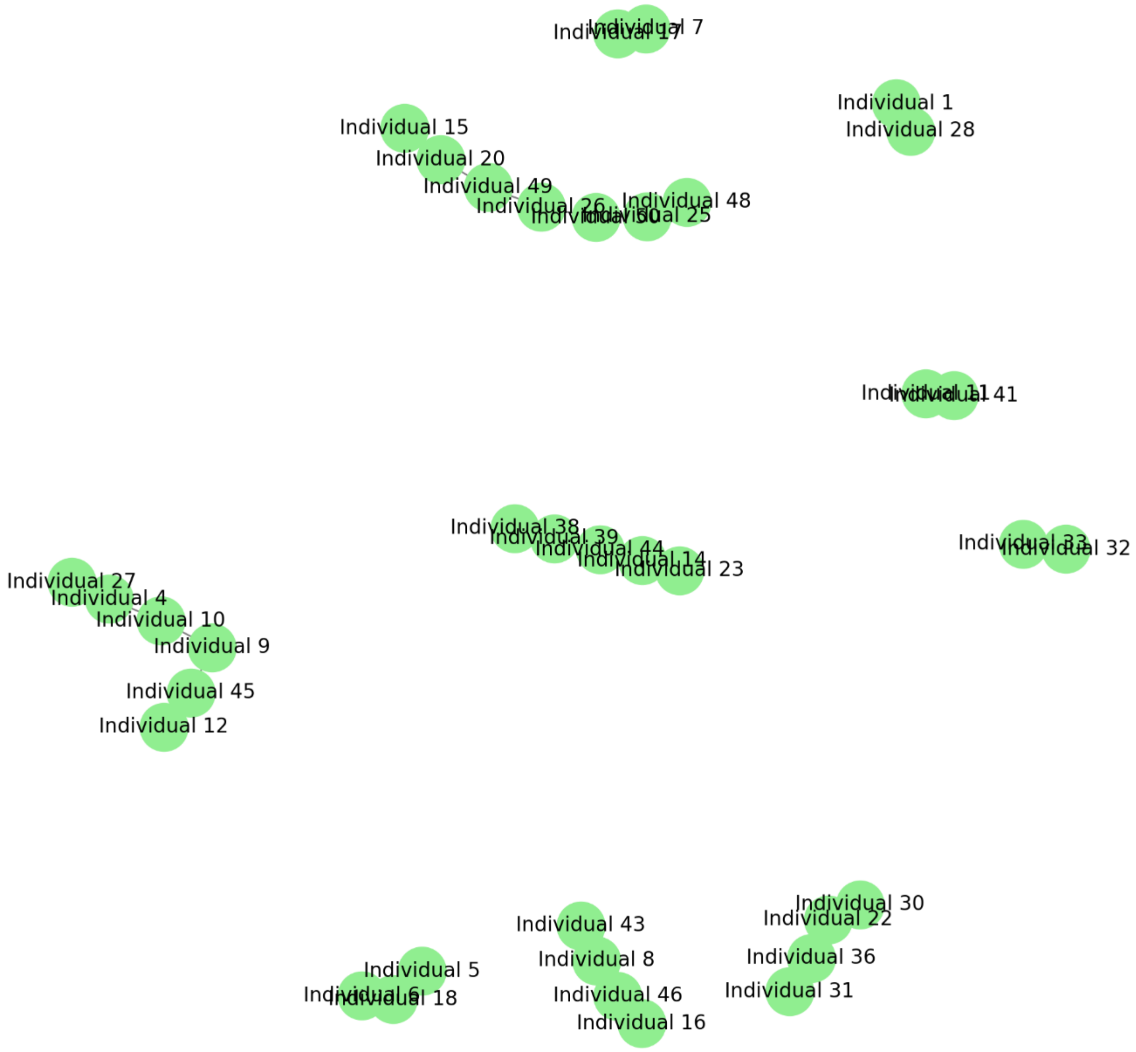


Figure 7-12: Social Ties in Opportunistic Environment

A mixed approach of MOBILE tower and person tracking is used in the analysis to reveal trends in crowd behaviour and social interactions. Heat maps and threshold plots represent a macroscopic understanding of crowd density, while the inner details of social ties are depicted in the network diagrams. These results drawn from complicated datasets emphasize the capability for predictions that can help city dynamics and the management of crowds in the urban planning domain and social structure understanding.

7.6 Algorithm 11: Comprehensive Mobility and Social Interaction Model with Enhanced DUCSIM

The class diagram figure 7-13 presents crowd dynamics from macroscopic and microscopic perspectives concerning crowd density and group or individual movement patterns in crowds. This algorithm uses an hourly movements data set across MOBILE towers, calculating crowd densities, movement thresholds setting, and group behaviour clustering for discovering intricate individual and collective movements' patterns engineered. The class diagram for the implementation details is illustrated in Figure.

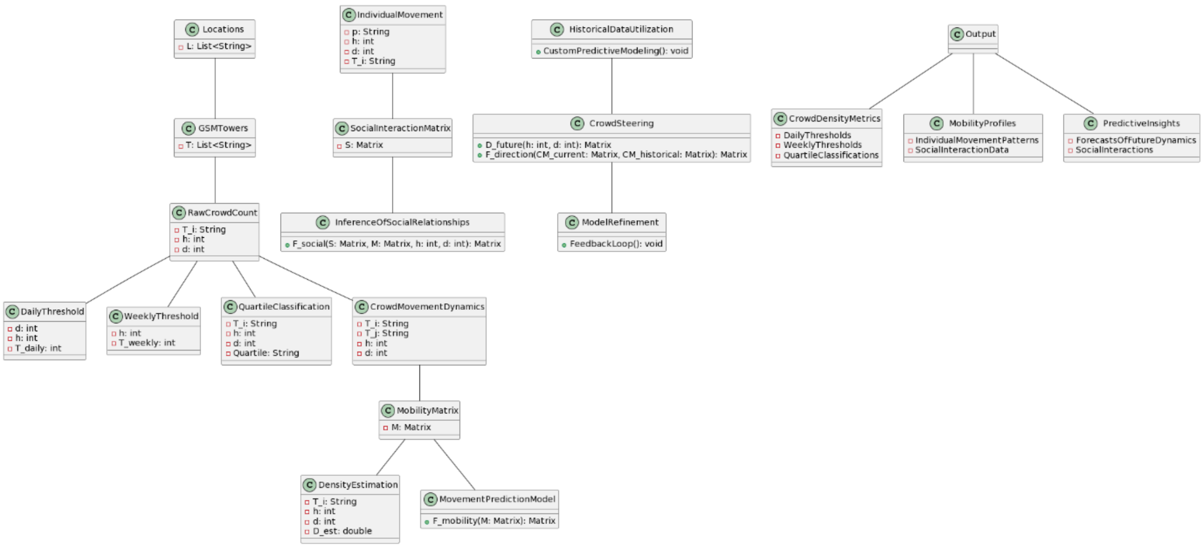


Figure 7-13: Class Diagram Algorithm 11

7.6.1 Median Thresholds Across Sites by Hour

The figure 7-14 below depicts the corresponding medians of the crowd counts of all these sites achieved for every minute. In this context, peaks in the plot signify particular hours of high medians of crowds, indicative of high movement or active hours of the day. On the other hand, troughs imply low points such as off-peaks. It can also serve as an indicator for forecasting sit congestions and resource planning.

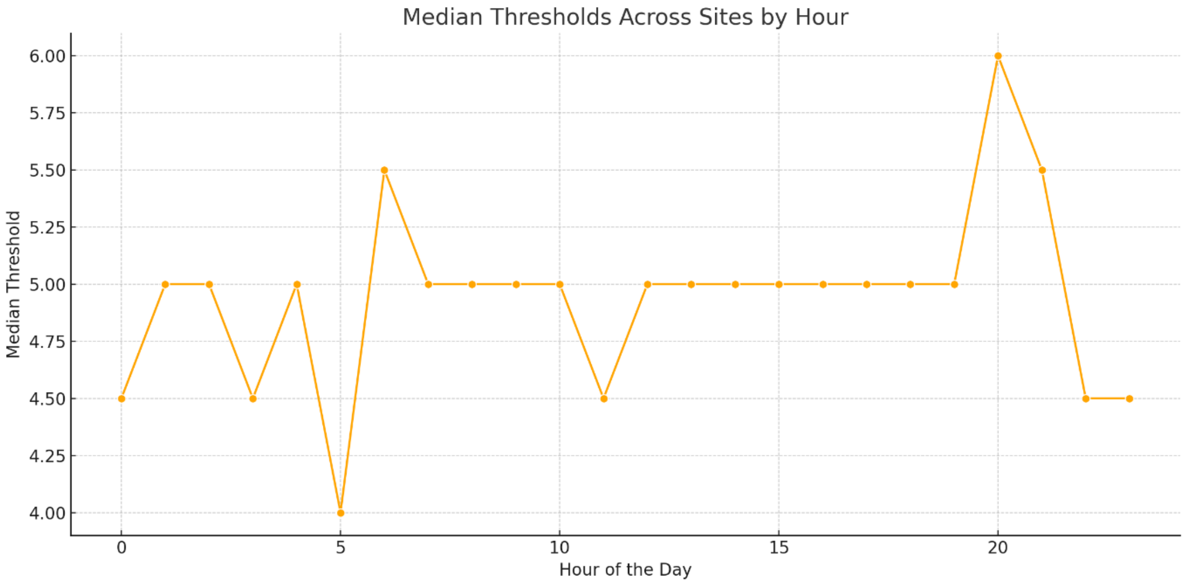


Figure 7-14: Median Threshold Across SitID vs Hours

7.6.2 Heatmap of Density Estimation at Sites by Hour

The Figure 7-15 represents the net density estimates of people in one day at different MOBILE towers. Higher and lower densities are represented by warm and cool colours, respectively. This means there was a high influx of eight units at SiteID 2 in the second hour, which probably indicated a peak period for that site.

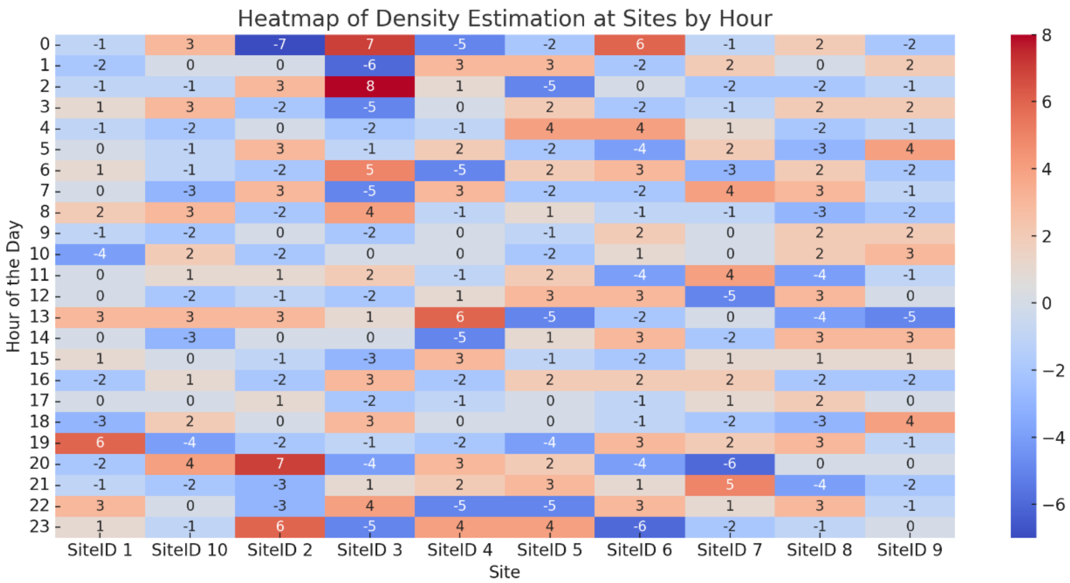


Figure 7-15: Crowd Density Estimation

7.6.3 Macroscopic Crowd Density by Site and Hour

Visual representation in figure 7-16 of the unprocessed crowd numbers for every GM's basepoint through one day. A peak appears on each line for every site, and another trough indicates ups and downs in the number of visitors. Peak values in the sites suggest a higher level of activity. This might be because of traffic or other local attractions.

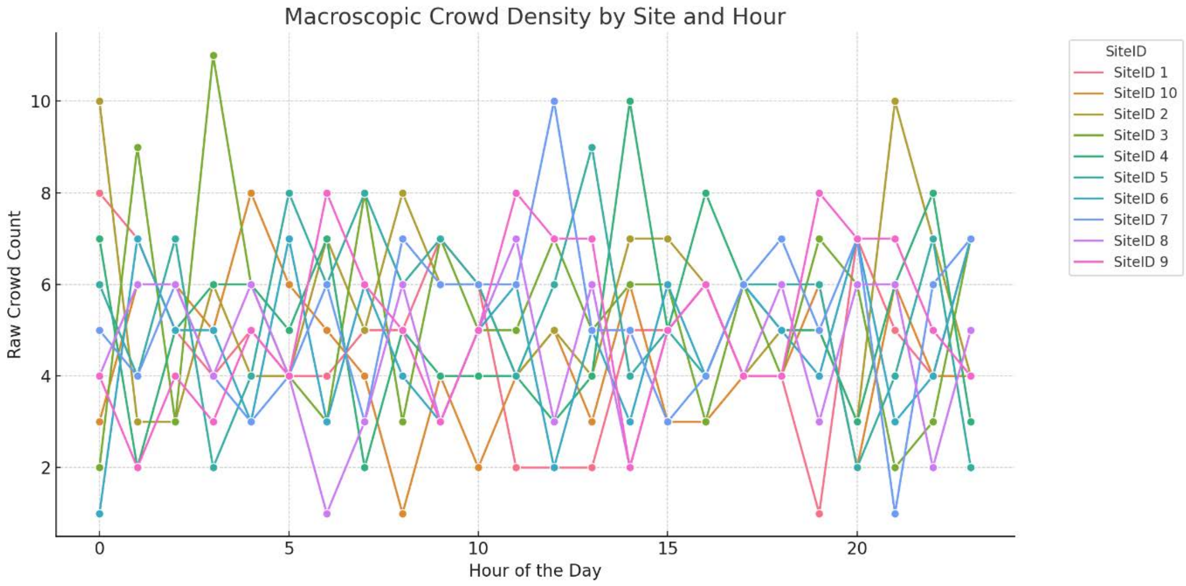


Figure 7-16: Macroscopic Crowd Density Distribution

7.6.4 Density Scatter Plot for Group Mobility Analysis

Figure 7-17 present that some people tend to occupy specific spots during different times of the day, as depicted in the density scatter plot. The dark area shows the highest density levels of men in one cluster zone during various hours. As such, Cluster 2 could be a gathering of people passing through in the middle of the day, indicating a similar routine or destination.

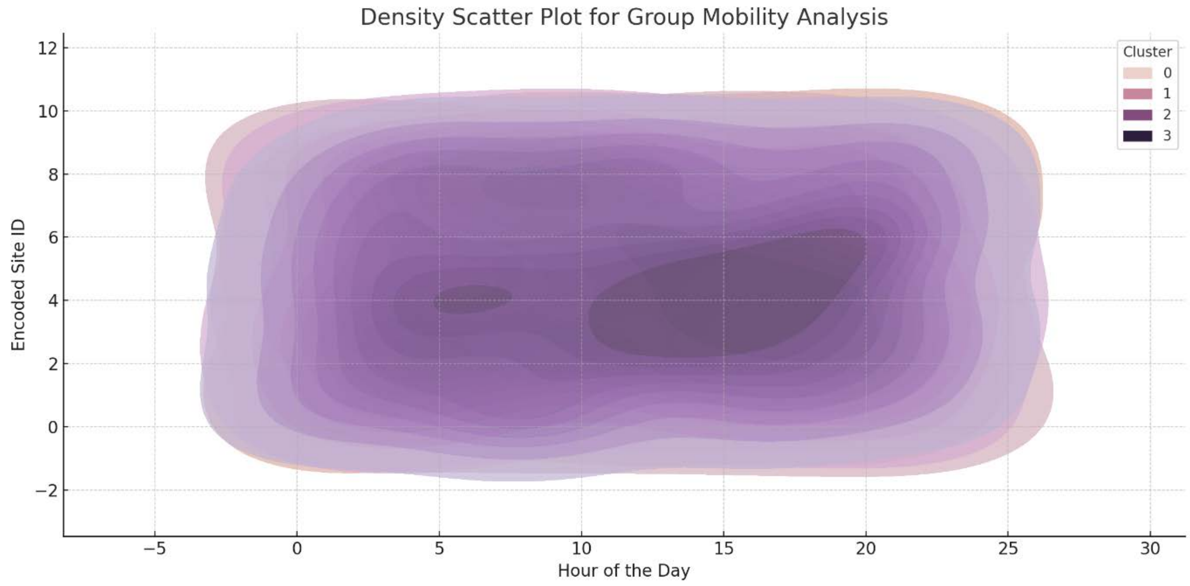


Figure 7-17: Group Mobility Density Distribution

7.6.5 Density Scatter Plot for Individual Mobility Patterns

This scatter graph figure 7-18 illustrates the travels of select people at various locations. For every individual, points denote the spreads representing the frequency of visits or the amount of time spent at various places. This could include an individual's points being closely located on-site ID3 for an evening, signaling a regular activity or habit.

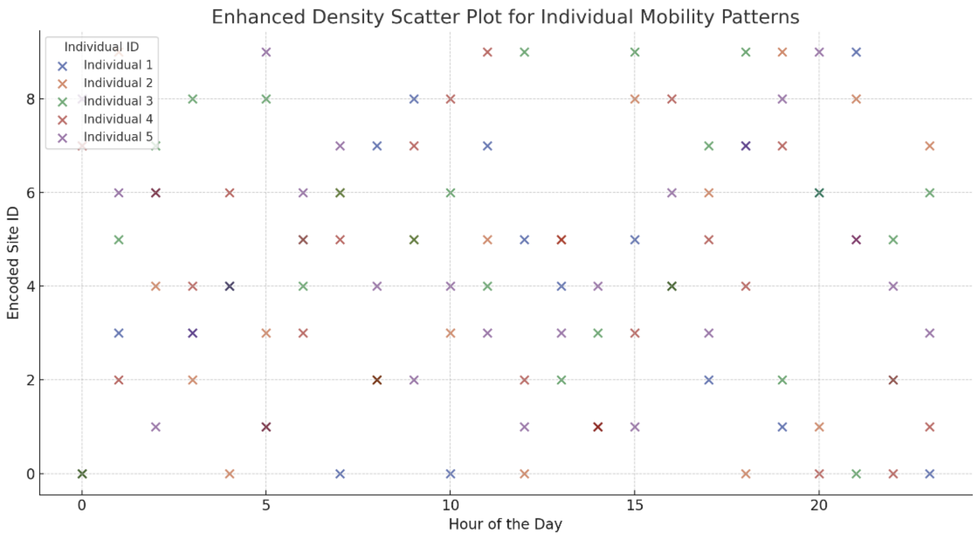


Figure 7-18: Crowd Density Scatter Plot

7.6.6 Prediction of Crowd Mobility

Heat map showing figure 7-19 the predicted numbers of persons moving from each FromSite to each ToSite in the coming hour. Darker patterns represent large populations, whereas lighter patterns signify reduced populations. Additionally, the intensity of the selected colors denotes a larger amount of predicted movement. A cell will be darker on, for example, a certain FromSite-ToSite intersection if the prediction says more people will travel from that site to that site in the coming hour. It outlines merely the movement that can be used for traffic management and resource allocation.

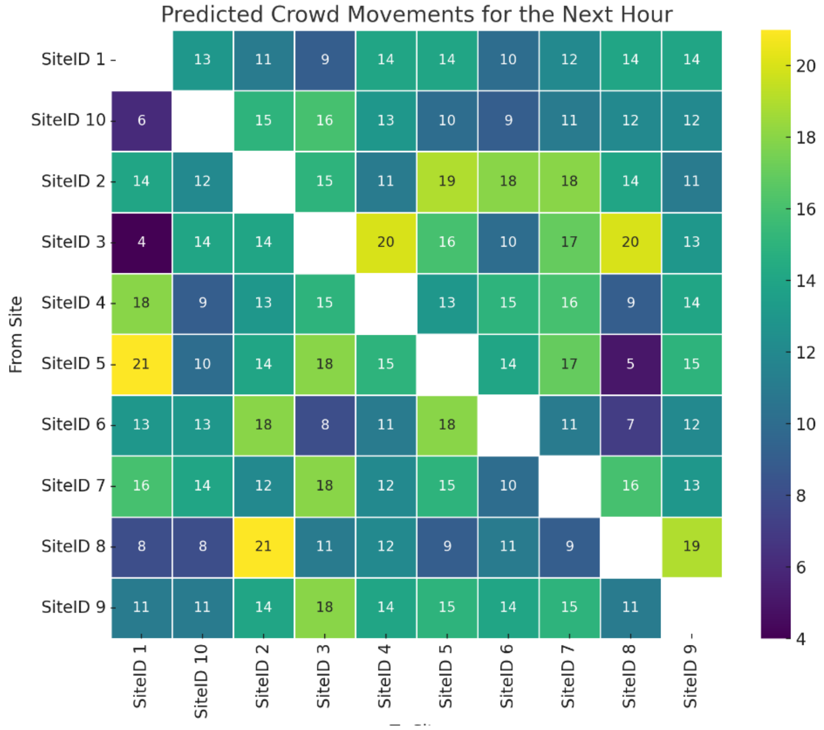


Figure 7-19: Prediction Crowd Movement Next Hour

Mobility behavioural traits were observed individually and in groups with varying crowd densities depending on sites and time. Density scatter plots showcase how heavily populated some places are. These insights are important for improving city crowd management, planning infrastructure development that considers people's movement complications, and the utility of data-driven methodologies.

7.7 Algorithm 12: Adaptive Learning and Customized Predictive Analytics With DUCSIM-$T_{M\text{-}o\text{-}M}$

The purpose of this study is to analyse the tower-based data for MOBILE for determining patterns of crowd movement and individual mobility patterns and to have an in-depth look at the issues raised by this study. Using sophisticated data management methods based on raw mob data (crowd count), we compose mobility matrices and get into mobile telephone connections (GSC Data). The method includes investigating macroscopic crowd density, micro mobility for individual behaviour, and dynamic predictive modelling. The results shed light on how people respond to certain situations and provide important input into issues such as traffic flow planning, urban planning, and other critical areas in the urban environment. Such a multi-faceted approach captures present dynamics and predicts future changes, delivering holistic data-driven decision aid in urban environments. The implementation class diagram is illustrated is figure 7-20.

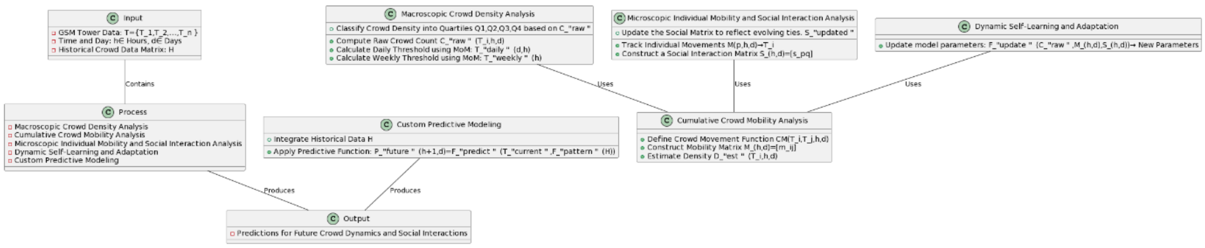


Figure 7-20: Class Diagram Algorithm 12

7.7.1 Median of Median Crowd Count Per Hour

The figure 7-21 illustrates the medians of median crowd counts per hour for each site on a cumulative basis overtime during the whole event (cumulatively). Interestingly, there’s an evident spike at night, probably indicating a commensurable occurrence or community’s assembly time frame that could demonstrate the community's behaviour regarding temporal patterns.

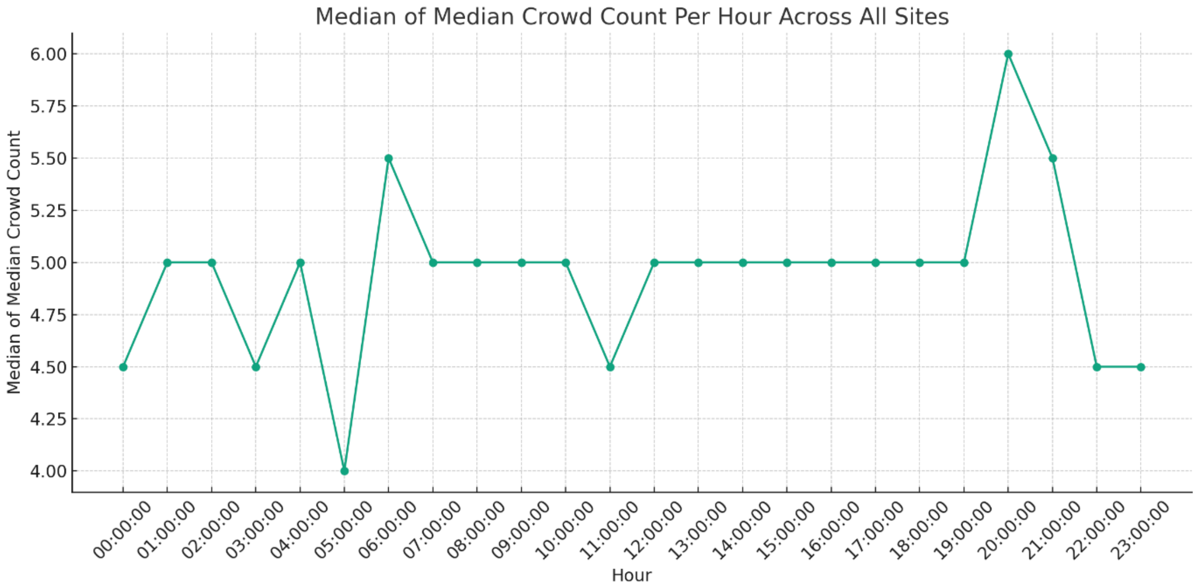


Figure 7-21: Median of Median Crowd Count Per Hour

7.7.2 Crowd Count Distribution

The Figure 7-22 shows variations in the number of people among various MOBILE tower installations over one day. These variations show different rhythms of movement and concentration in certain parts of the area depending on time. These include instances where SiteID 1 and SiteID 10 have noticeable peaks in crowd count, implying places of heavy movement and high traffic activities, respectively.

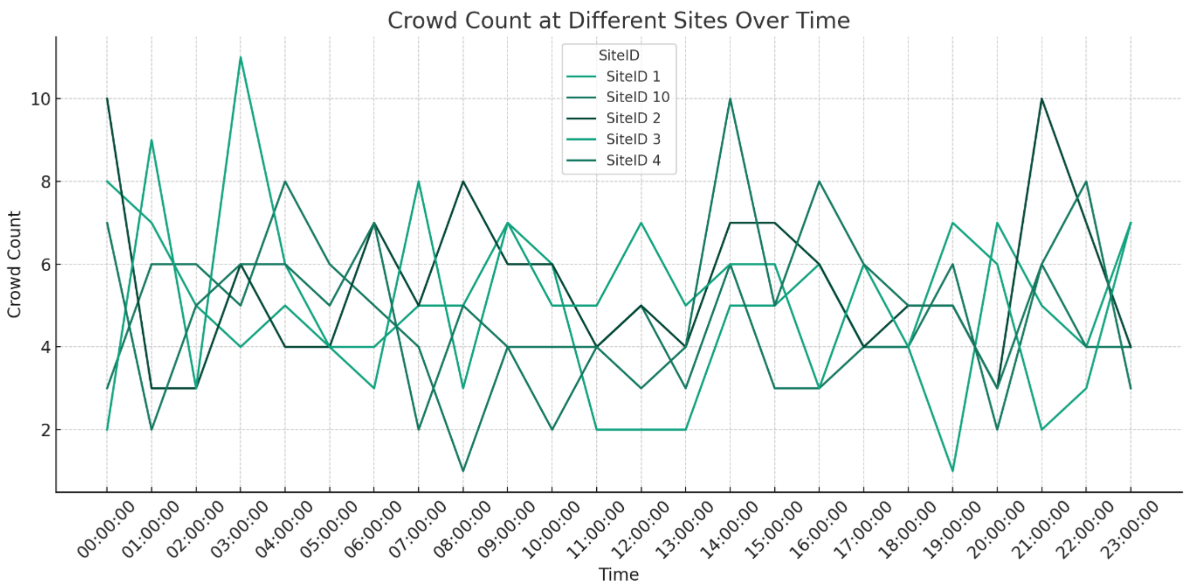


Figure 7-22: Crowd Count Distribution

7.7.3 Social Interaction

The bar chart in figure 7-23represents the number of social visits in different locations with significantly high social visitation counts for Site ID 5 and Site ID 9. This may suggest that these sites play an important role as either social meeting points or the reflection of interpersonal relations among residents of those communities served by these sites.

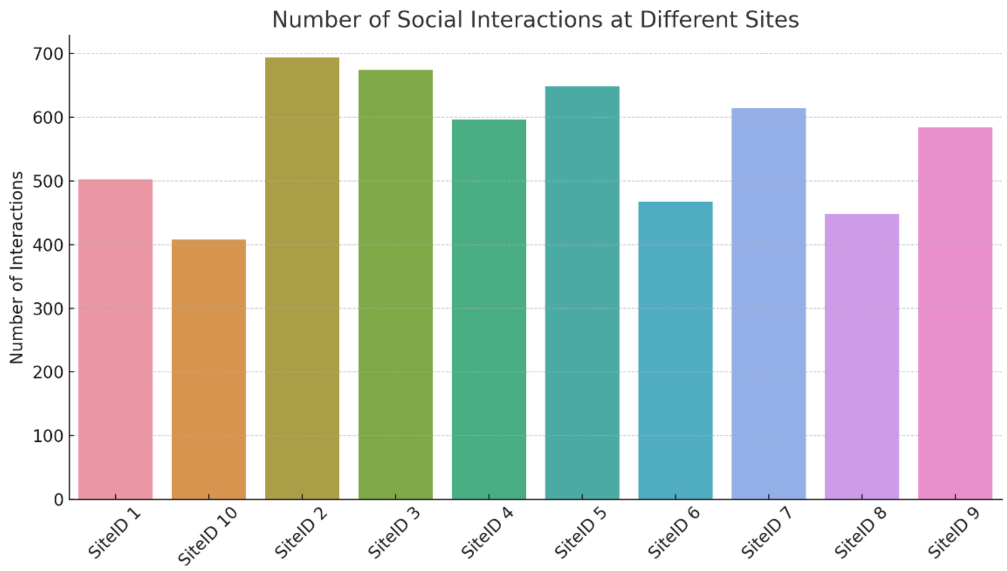


Figure 7-23: Number of Social Interaction

7.7.4 Individual Strong Connections with Notations

The network diagram shows in figure 7-24 presents the complex network of strong social links. Line thickness denotes the intensity of communications. This shows a highly networked community with some very important nodes acting as communication centers for the people of that environment. Prominent clusters and example individuals such as P3 P17 are connected strongly over the spatiotemporal opportunistic environment.

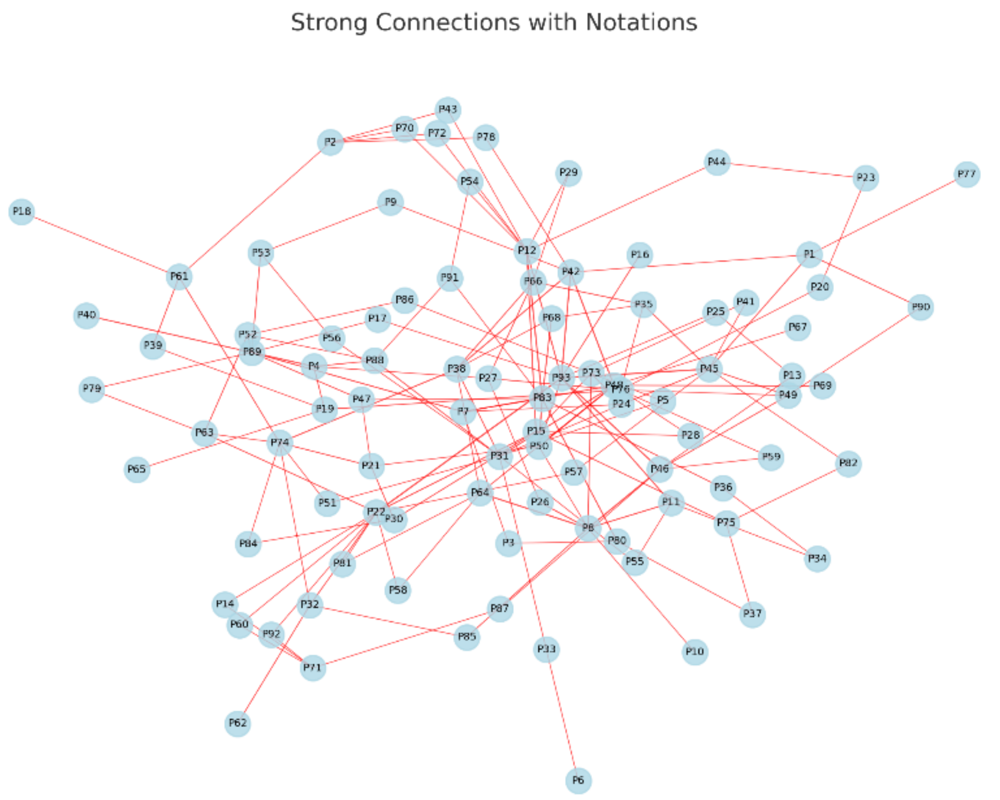


Figure 7-24: Individual Strong Connections with Notations

The plot analysis helps gain information on movement and social interaction patterns observed in MOBILE tower-based human mobilities.

7.7.5 Prediction for the Next Hour

The figure 7-25 describes an expected crowd density depending on time. These predictions can provide a basis for efficient crowd management, deployment of emergency services, and load-balancing networks for a telecommunication firm. In particular, the key hours that change predicted crowd count may serve as rush hours or even dismissal, which can help urban planning and management strategies.

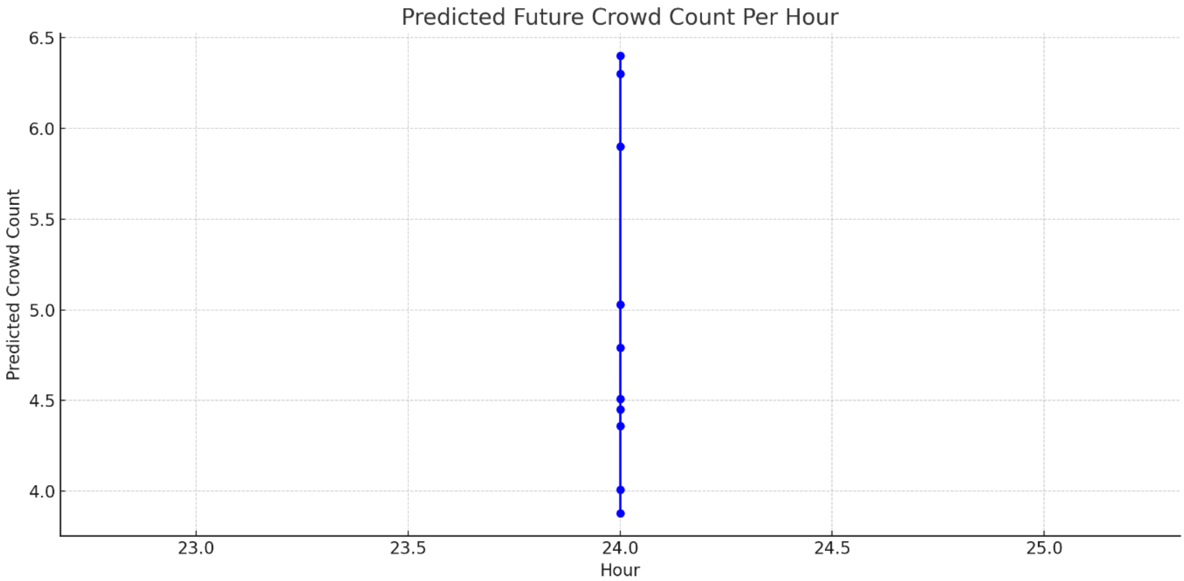


Figure 7-25: Prediction for the Next Hour

7.8 Conclusion

Analysis of MOBILE tower data gave a thorough understanding of crowding activity and personal movement trajectories. The research applied a structured manner to look at macroscopic Crowd Density Analysis, compute the raw count of crowds by days and per week based on the method of M-o-Ments, and classify crowds' density by quartiles. Thus, a mobility matrix was built for towers in Cumulative Crowd Mobility Analysis, indicating movements towards and away from them, giving net flows and resulting in net density estimation. The study of Microscopic Individual Mobility and Social Interaction Analysis shed some light on the movement patterns at the micro level to build up a Social Interaction Matrix. Self-learning models with dynamic elements that updated the model parameters to adjust to newly emerged data structures were developed. Lastly, Custom Predictive Modelling was used, which utilized previous information to predict how the crowd will behave in terms of both group movement and social interaction for the next few hours/days. This complex approach not only captured existing crowd dynamics but also served as the foundation for predicting future patterns and events.

Work presnted in this chapter has been published as patent no.[1 and 2]. A conference paper is published given in my publication list [1].

CHAPTER 8: HETEROGENEOUS OPPORTUNISTIC ENVIRONMENT RESULT

The algorithm's activity diagram in Figure 8-1 provides an all-encompassing simulation of the urban dynamics, combining data from traffic, ride-sharing, WiFi systems, and social interactions—a multi-dimensional approach to estimate crowd density, mobilities, and social interactions in an urban area. The algorithm uses sophisticated predictive modelling to forecast urban dynamics in the coming years, which can be beneficial for formulating urban planning and management. This provides a holistic insight into the city's dynamics, which is critical in supporting smart city development and governance.

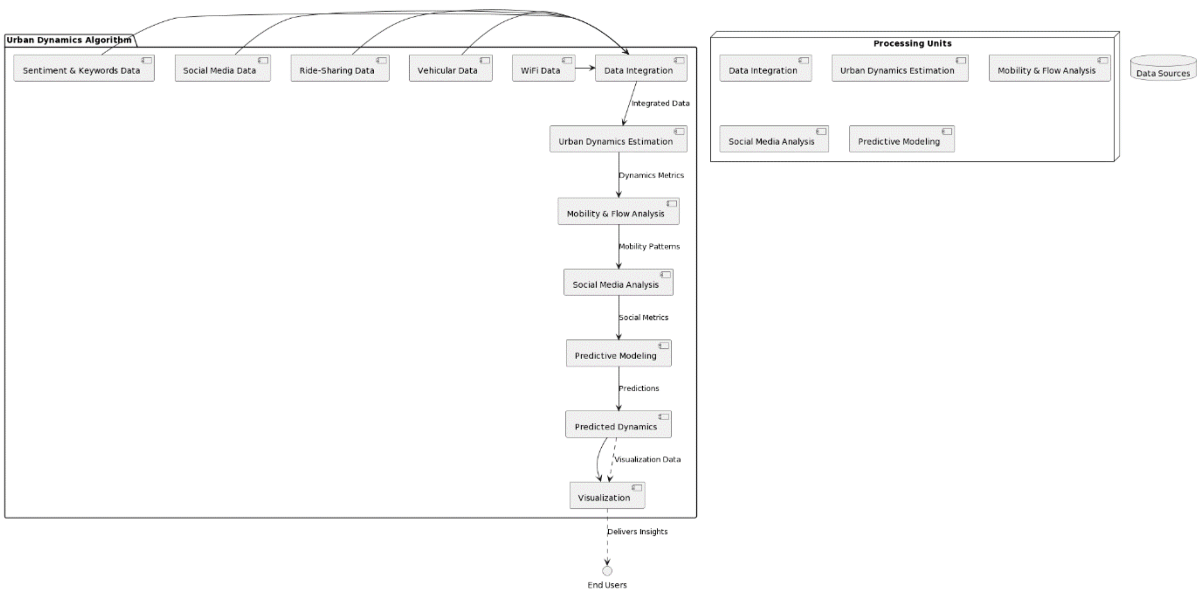


Figure 8-1: Activity Diagram

8.1 Input / Simulation Parameters

Simulations were done using a broad set of sources of urban dynamics data for various places throughout a certain time frame. Key parameters include:

- Number of Locations: Simulation of data in different locations.

- Time Frame: Each simulation stage reflected how changes in urban dynamics would occur over time.
- Data Sources: The experiment utilized information from various sources such as WiFi, vehicle movements, taxi data sharing, tweet messages, sentiment analysis, and keyword counts, among many others.

The table 8-1 presents the Input and Simulation parameters those were used for implementation of the algorithm.

Table 8-1: Input / Simulation Parameters - Heterogeneous Networks

Parameter	**Value**	**Unit**
Number of Locations	5	Locations
Time Frame Start	2023-11-21 22:50:39.328073	Date/Time
Time Frame End	2023-11-22 21:50:39.328073	Date/Time
Number of Time Points	24	Time Points
Location	3.0	Average Count/Score
WiFi	50.558333	Average Count/Score
Vehicular	25.783333	Average Count/Score
RideSharing	16.391667	Average Count/Score
Tweets	101.783333	Average Count/Score
Sentiment	0.043373	Average Count/Score
traffic	5.116667	Average Count/Score
event	5.15	Average Count/Score
weather	4.825	Average Count/Score

8.2 Crowd Density

The figure 8-2 shows urban area density levels of five different cities, each bar labelled according to its particular density. Downtown, Suburbs, Industrial area, Residential, and Commercial; apparently, none of the locations show any noticeable difference in crowd densities. This means that activity or population was spread evenly throughout these urban areas. While the actual specific values for the “crowd density” do not appear on the chart, they usually appear along the y-axis, sometimes representing people by numbers, and hence, a place with a large number of people reflected by the x-coordinate is either busy or a place to spend in groups.

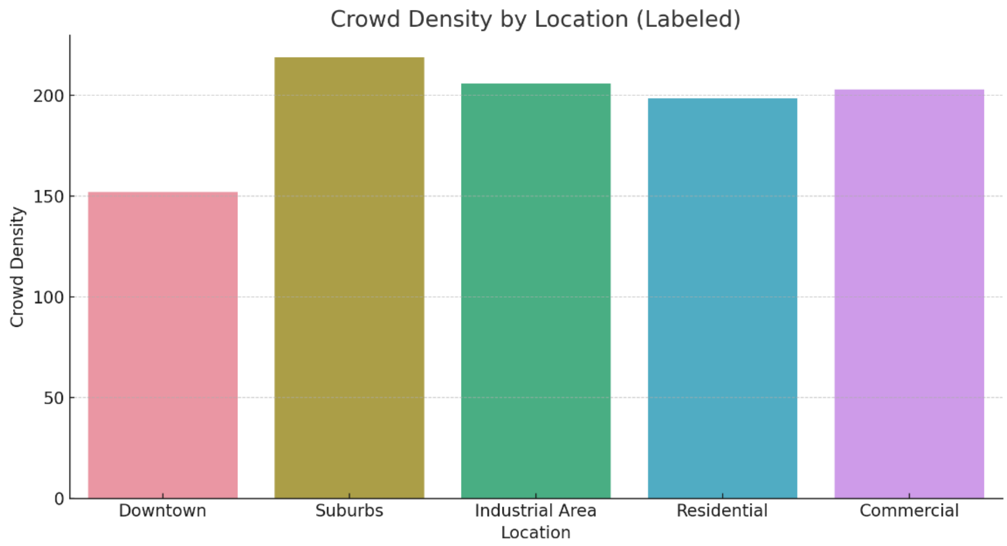


Figure 8-2: Heterogeneous Crowd Density Distribution

8.3 Density Flow Metrics

Figure 8-3 represented in a heatmap as their average flows between these five locations. Darker colors represent high flows, while light colors show low or negligible flows in each cell. The flow from Location 3 to Location 2 equals 57.2 and represents an important movement or interaction between Location 3 and Location 2. On the contrary, the flow from Position 4 to Position 3 is among the low rates (at 41.4), illustrating insignificant movement in the opposite direction. The heatmap depicts the volume of flows (traffic, commuter movements, or other flows) in the given areas.

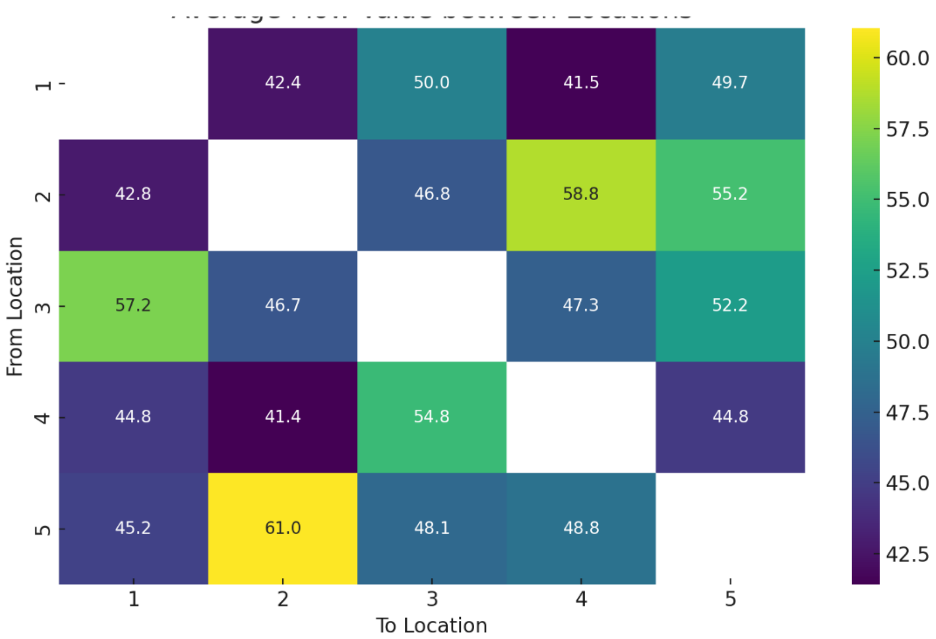


Figure 8-3: Density Flow Matrics

8.4 Rideshare Interactions

The figure 8-4 illustrates the median of 402 interactions for each location. There are differences in riding-sharing activities among locations, and Location 5 shows relatively more active participation than average.

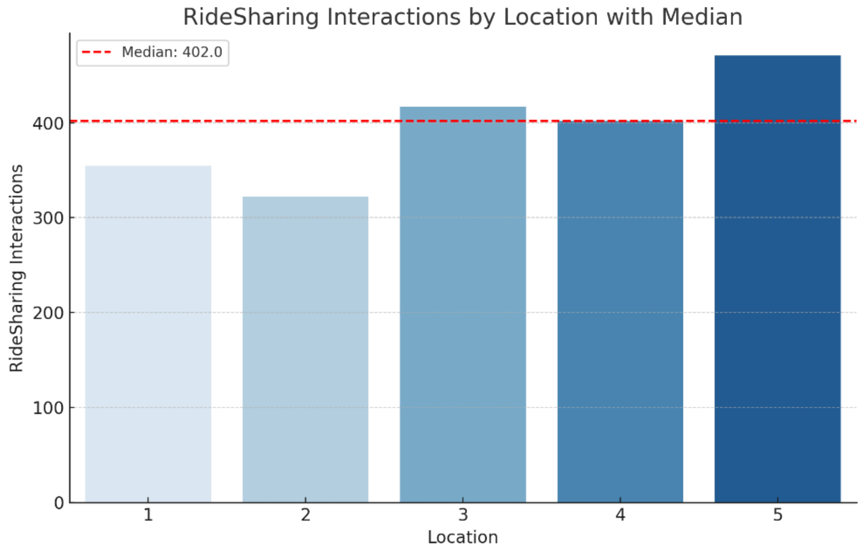


Figure 8-4: Rideshare Interaction Distribution vs Location

8.5 Twitter Interaction

Figure 8-5 shows the interaction counts on the map in the median of 2531 tweets. Interaction levels between locations show fairly consistent, while all locational sites indicate high tweet interactions almost above the median.

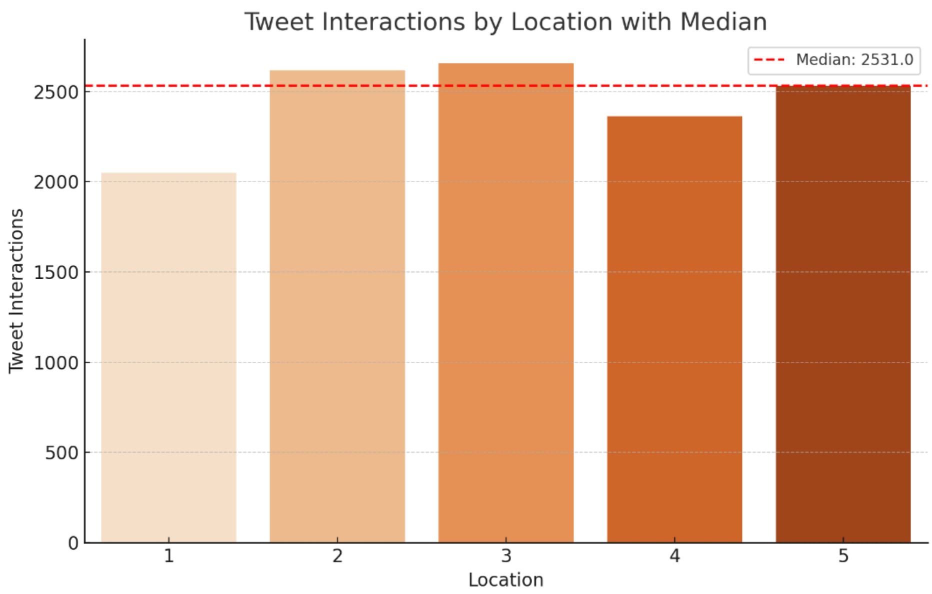


Figure 8-5: Tweet Interaction Distribution vs Location

8.6 Sentiment Distribution

The following figure 8-6 shows the sentiment distribution by place, with a median sentiment of 0.4. Generally speaking, most locations have positive sentiment, except Location 4, whose average sentiment line is placed marginally below the median line.

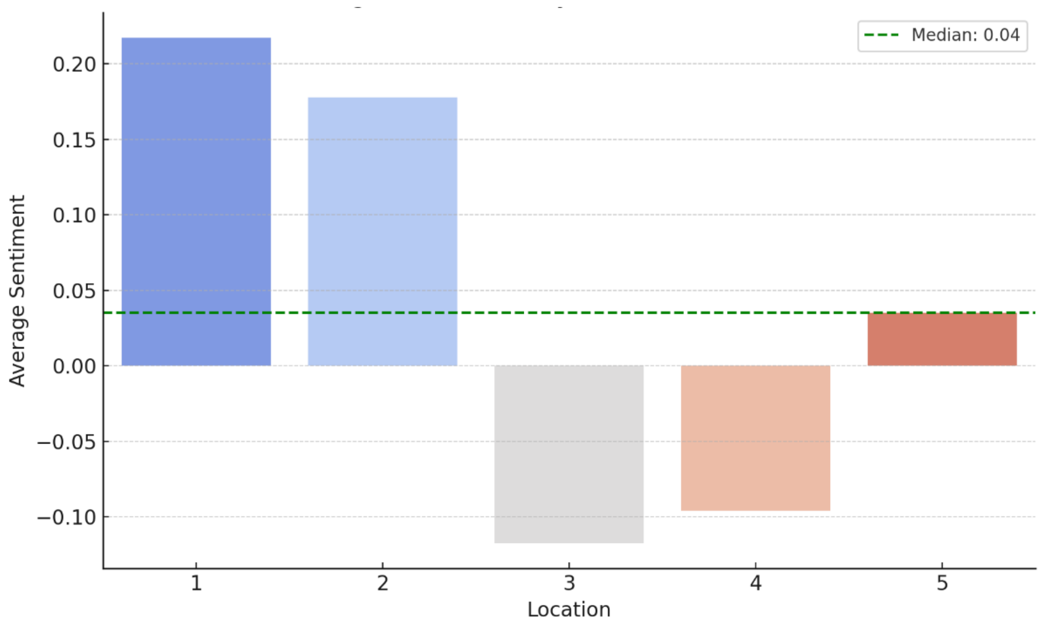


Figure 8-6: Social Media Sentiment Distribution vs Location

8.7 Prediction Distribution

Differences in predicted plots illustrate in figure 8-7 various forecasts for an urban activity that is dynamic and thus unstable on its predicted plots. This is because the time series presentation has been there for years; it enables us to observe fluctuations like peaks, troughs, and trends over time. Such changes in migrations, occurrences, and other processes could be anticipated from future displacements which are to happen in relevant destinations.

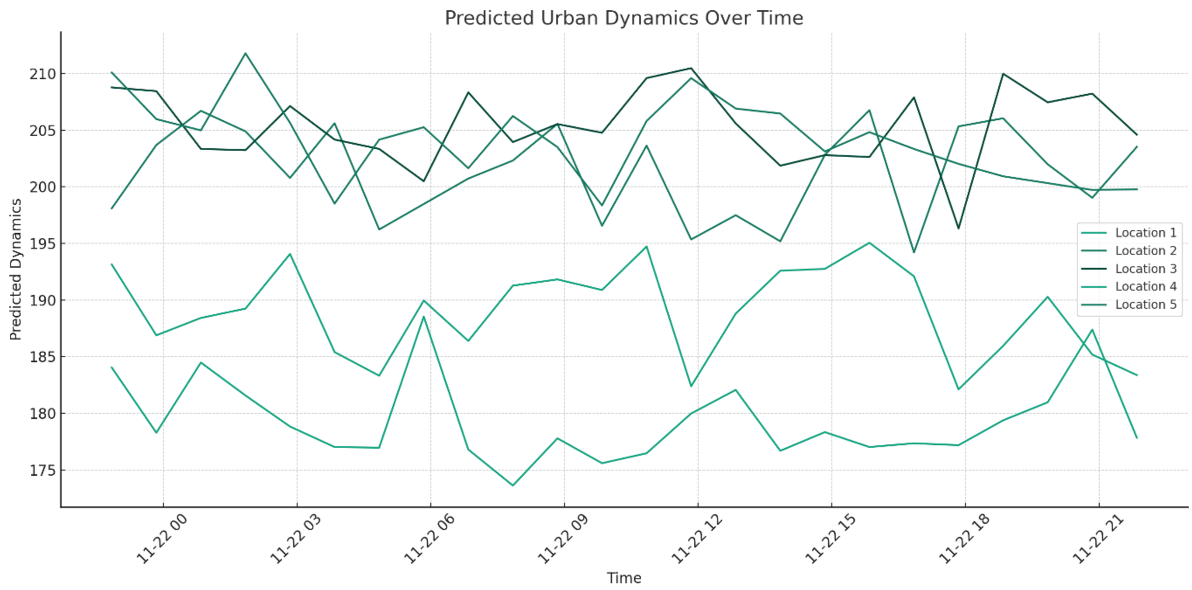


Figure 8-7: Urban Dynamics Prediction Over Time

8.8 Conclusion

The simulation collects data from various urban dynamics, representing and forecasting complex city patterns. The analyzed crowds were found in different sections of towns, and these groups exhibited a considerable rate of carpooling, among other activities on social networks. There seemed to be generally favourable social media sentiment, although there was some variation according to location. Ebbs and flows of urban activity were predicted over time using predictive modelling, suggesting that every site has its distinct rhythm, perhaps due to some local events, traffic, or population density. The sum totality of these insights leads to a delicate grasp of urban behaviour needed for the competent formulation of urban planning.

Work presented in this chapter is been published as patent [3 and 4]

CHAPTER 9: ALGORITHM 6: APPLICATION: SMART CITY CROWD MANAGEMENT USING MOBILE DATA ANALYSIS

9.1 Introduction

Due to rapid urbanization, smart cities have come to portray the future ideal of smart city governance in contemporary society. However, managing crowd densities remains a crucial problem these sprawling new urban centers pose. Crowds (characterized by number, density, and behaviour) significantly impact urban resources and infrastructure [238]. Crowd densities may result in congestion of traffic, strain on the public transport system, raise accident rates, and even hamper emergency response services. On the contrary, idle grounds signify avenues that could have been developed into opportunities by businesses and other community engagements [239-242]. This worsens, especially during major events like public holidays, peaks, or busy city areas. Poor crowd management also results in discomfort, reduces security, and lowers the quality of urban life. Hence, knowledge of how the crowds behave and prediction of these are required in having proper city governance, safety, and better urban experiences, among others.

9.2 Objective

Crowd management uses Mobile data, which is focused on providing live data for decision-making in urban areas. MOBILE data is highly pervasive and constantly produced, and therefore, it presents an extensive repository for identifying the movement and pattern of settlements in various city sectors. From this analysis, city planners and other authorities will see why and where many people gather, allowing them to know which places are densely populated and how to control such crowds. This approach can be important because it may help change reactive city management approaches into proactive ones. This ensures that planners can predict potential crowd-related matters that may arise, thus streamlining the transport system and guaranteeing public safety measures. With that said real-time crowd information can be pivotal during emergencies like natural disasters or public health disasters in directing resources toward making effective containment and evacuating strategies.

9.3 Overview of Approach

Managing crowd density in smart cities entails an algorithmically controlled system that leverages geographic information systems and MOBILE data. The initiation phase starts by setting up the initial number of MOBILE towers and data arrays for crowd counts, daily and weekly threshold levels, and quartile classifications. Crowd counts obtained from each MOBILE tower yield fine-grained information regarding people distributed in various parts of the city according to the daytime. After that, the algorithm computes the limit values for the day and week and uses them as benchmarks for the population density in the area. The thresholds, however, play a crucial role in differentiating between peak and normal crowds in terms of time or spatial contexts. Quartile classification of crowd density is the subsequent significant stage. We can label them low, medium, high, and very high through quartile categorization of the crowd data. This categorization gives defined stratification regarding crowd density and directs efficient crowd control measures in the different categories. Models based on historical information about crowds and a classification made by quartiles. Such models allow city officers to predict and plan for changing the density of crowds so that there are enough resources available in case there is a need to respond to any possible crowd-specific issue. Applying a structured approach utilizing MOBILE data in smart cities could greatly boost their ability to control crowds, thus translating into more secure, effective, and liveable metropolitan regions.

9.4 Methodology

9.4.1 Data Simulation

A simulation of MOBILE data was performed to develop a realistically functional model for controlling crowds in smart cities. The simulated environment provides a safe place to understand how the design can be applied when analyzing crowd density. The key parameters of the simulation included:

- Number of MOBILE Towers: In this particular case, certain MOBILE towers were set up at different points in the urban environment. The number of mobiles in each section was an estimate of the crowds in the surrounding area.

- Days and Hours: A simulated week was used in this exercise, with crowd density fluctuating on each day and hour. It involved all working days (including weekends), which accounted for the usual upsurges of urban traffic and population density distributions.
- Crowd Counts: Each MOBILE tower showed hours-by-hour crowd counts on a simulated data set. These counts occurred in a realistic manner of an urban population for any specific day.

The figure 9-1 is created to illustrate comprehensive overview of urban mobility patterns implementation, giving the basis for future analysis and population control policies.

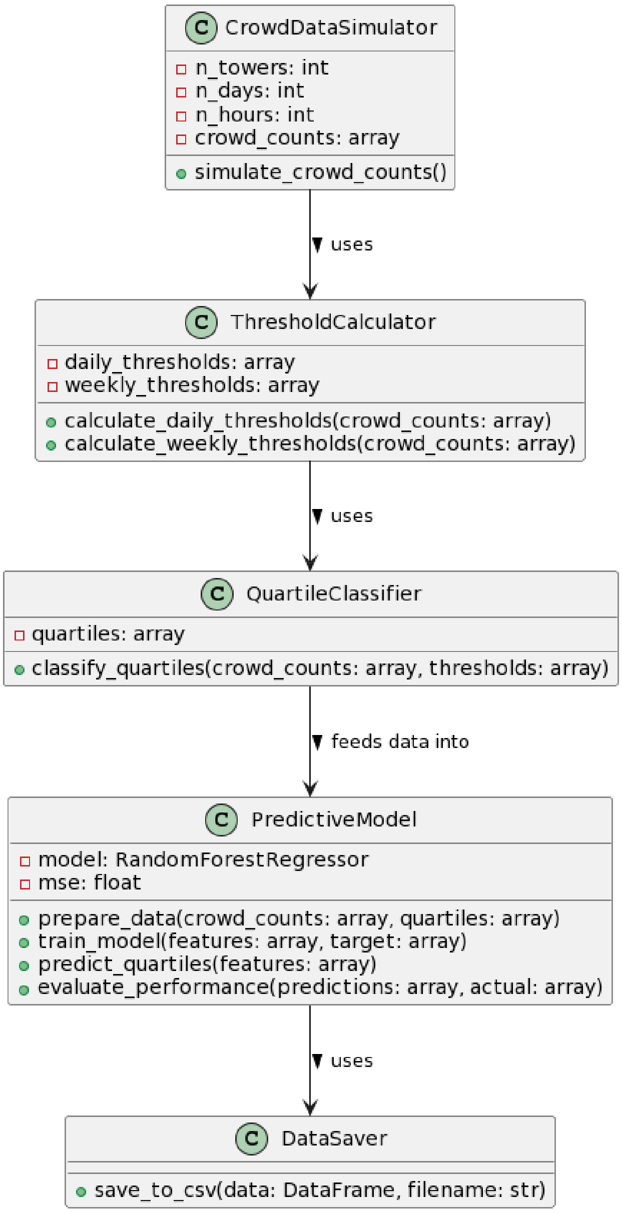


Figure 9-1: Flow of Application - Crowd Density Estimation

9.4.2 Results and Discussion

This section delves into the empirical findings from the smart city crowd management system. By meticulously analyzing real-world MOBILE data, we uncover the intricate patterns of urban crowd movement. The ensuing results encapsulate the ebb and flow of city life, translating raw data into actionable insights.

A. Daily Threshold Trends for All Towers

Figure 9-2 line graph maps the thresholds of crowd density for one week. The observation shows an interesting trend regarding these thresholds that separate the various crowd density grades. For instance, the first-hour threshold on day zero moves up as high as hour eighteen on day 6. This could be due to a weekly occurrence or just the local.

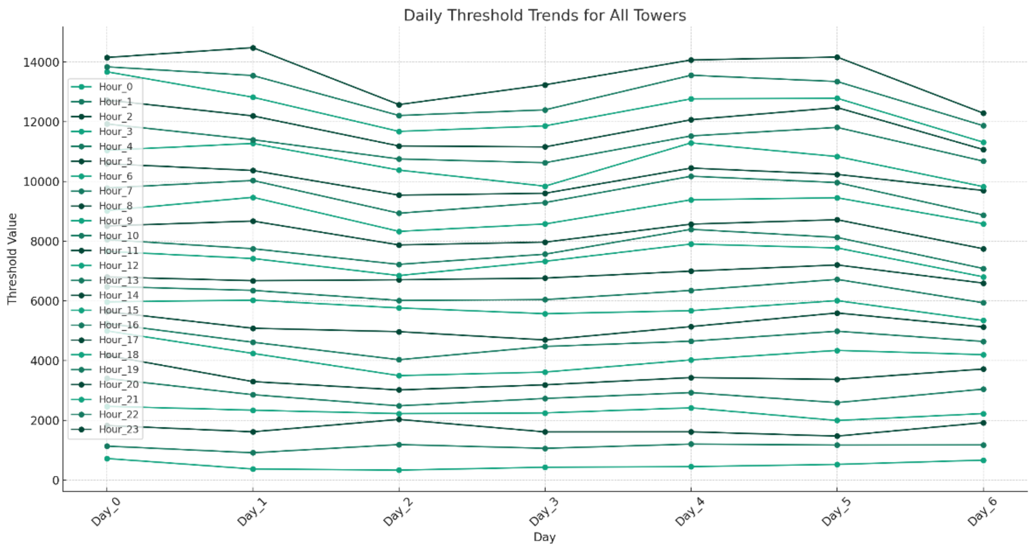


Figure 9-2: Daily Threshold Trends for All Towers

B. Quartile Classifications for Day 0

Figure 9-3 illustrates on a closer look at a particular day, the heatmap indicates varying crowds of people at different towers. For example, Tower 0 has elevated density (indicated by a vibrant color) at about 8 am - the usual traffic jam between jobs. On the other hand, Tower 4 has lower density values during the day, probably showing a residential or less populated zone.

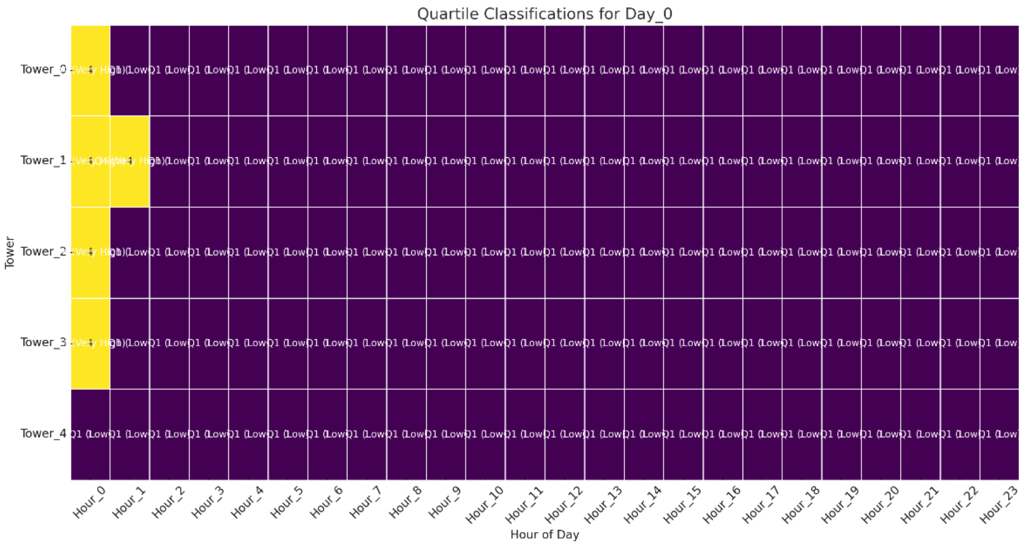


Figure 9-3: Quartile Classifications for Day 0

C. **Average Quartile Classifications for the Week**

In Figure 9-4 the Heatmap shows the overall picture of crowd distributions for 1 week period and shows urban dynamics on a global scale. A recurring trend with high median values in the afternoon (Mid-day hours; 12 pm – 2 pm) might be attributed to lunchtime habits. The observed pattern shows that lower quartiles recorded in the early hours increase towards the end of the day, which is characteristic of many cities where populations move to and from work.

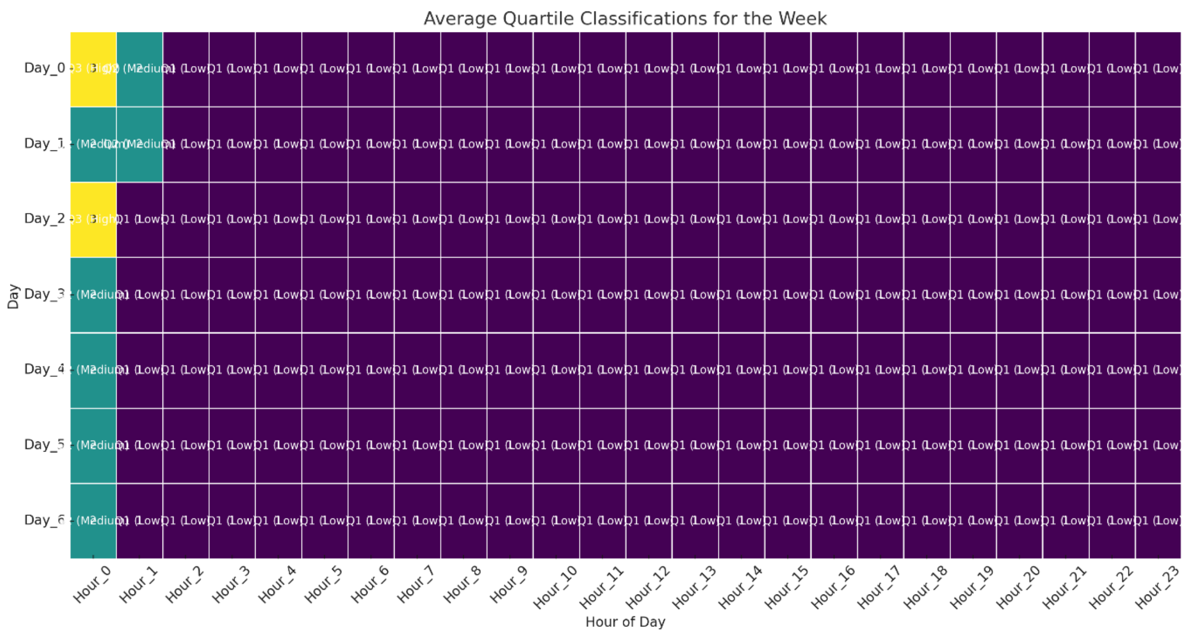


Figure 9-4: Average Quartile Classifications for the Week

D. **Predicted Quartile Classifications for Day 7**

Figure 9-5 presents the model predicting the crowd densities for future days, which is useful for proactive planning. As stipulated by the model, Tower 2 may be highly dense with people from as early as 5 p.m. The likely scenario may involve evening social activities or the typical ending of a workday. Such a prediction can help develop traffic control policies and schedule public transport.

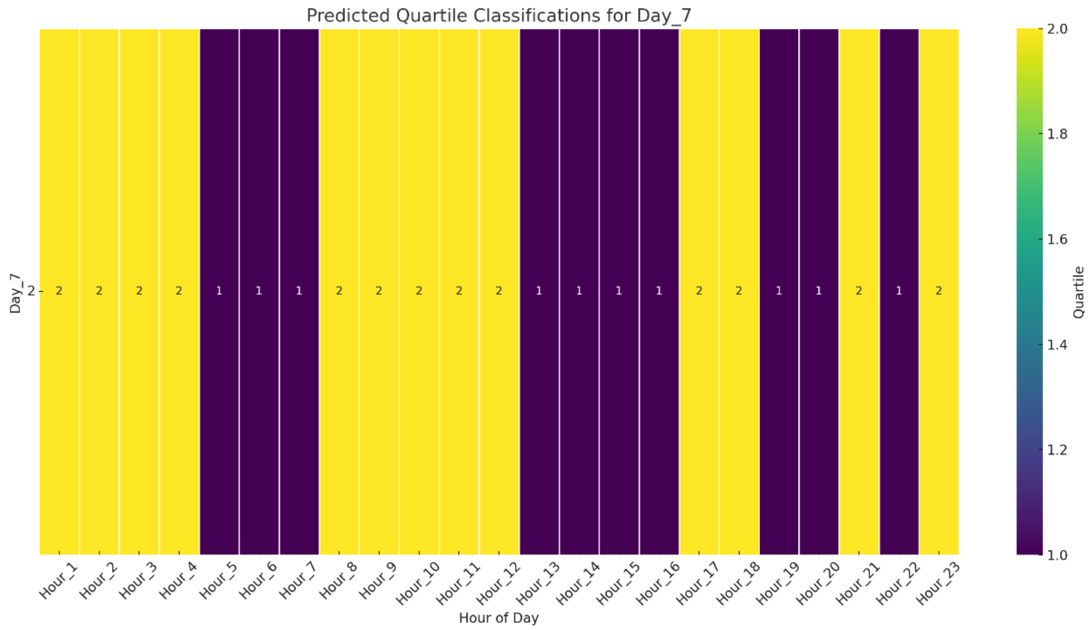


Figure 9-5: Predicted Quartile Classifications for Day 7

The urban dynamics synthesized from MOBILE towers provide a bird's-eye and granular view of urban dynamics. Rather than being retrograde and retrospective, these analyses, with predictive character, contribute to the modernization and forward-looking vision of the smart city plans.

9.5 Applications and Implications

The analysis provided by this smart city crowd management system can be seamlessly integrated into various facets of urban planning and management, offering substantial benefits in several key areas:

- Event Planning: Its use is greatly beneficial in large-scale events. Organizers should look for crowd patterns and restructure their schedules, layout, and logistics to guarantee easy movement and high security. For example, concerning festival or concert arrangements, providing services such as security and first aid can be deployed based on forecasted crowd densities.

- Traffic Regulation: This is useful information that traffic authorities can use in directing and controlling vehicle flows. The analysis of crowd predictive behaviour can provide vital input in choosing traffic light patterns for the traffic system, closure of selected roads, and scheduling the public transport system. Adopting this proactive stance can help minimize traffic jams, crash possibilities, and total road security.
- Emergency Response: Real-time crowd data is needed during emergencies like natural disasters and security threats. This information can assist emergency services in planning evacuation, allocation of resources, and finding relatively less congested routes for faster response time.

9.6 Future Prospects

The potential for integrating this system with other smart city technologies presents exciting prospects for urban development:

- Smart Infrastructure Integration: The system will be linked with the smart infrastructure of intelligent lighting and IoT-enabled devices that will enable dynamism of resource adjustment according to information about crowd density. In this case, a few examples of how streetlights can be optimized concerning energy efficiency and safety, such as streetlight dimming dependent on pedestrian traffic.
- Data-Driven Urban Design: Over time, information about crowds can influence the creation of facilities for public use, such as shopping centers and traffic lines. The approach creates cities that operate efficiently and continuously adapt to people's changing needs and behavioural patterns.
- Enhanced Citizen Engagement: Integrating such technology into mobile apps and public portals lets citizens be equipped with live alerts to decide their movements within these cities.

9.7 Conclusion

This crowd management system is one of the major steps towards improving the smart city model. It goes beyond reactive models to deliver a proactive data-centric model for effective urban administration. The system will help optimize daily commutes and improve emergency safety for the city's operations. Thus, it makes smart cities more reliable, efficient, and centered on citizens, which is what a smart city should be about, preparing for future smart technology integration.

CHAPTER 10: CONCLUSION

The development and application of Enhanced DUCSIM is a significant leap in urban dynamics studies. This powerful instrument uses different communication Networks like (Mobile, WiFi. Li-Fi networks), environments like (cars and ride-sharing) Social Media Networks like (Facebook, Instagram and Twitter among others) to fully understand how the city functions. The urban dynamic analysis offers a key advantage in supporting urban planning, sociological studies, and smart city initiatives.

10.1 Summary of Findings

DUCSIM, is an improved predictive algorithm to understand urban dynamics at more detail. The system of crowd density projection and movement provides great accuracy by using up-to-date data of urban activity, combined with the pattern of historical events. The workflow of the algorithm is divided into several key phases: data collection and cleaning, threshold analysis, model application to the data and rigorously evaluated. At first, DUCSIM gets updated crowd data from the city in terms of numbers (A_raw(gt)) and past crowns' population density (D_hist(gt)). Normalization and structured of this data takes place during the Preprocessing phase is vital to ensure consistency and quality in further analyses.

Median-of-Medians (M-o-M) threshold, $T_{M\text{-}o\text{-}M(gt)}$, forms the basis for DUCSIM's threshold analysis, comparing current activity levels against priorly set parameters, which marks the first phase in recognising variations in crowd population. Once threshold analysis becomes complete and necessary, DUCSIM enters into its application phase in which it applies a predictive model that enables forecasting of future crowding level, i.e., D_{pred} (gt+1), incorporating the actual activities, historical data, as per threshold analysis It entails a smart algorithm which gives a reliable weight to each piece of information for it to produce the best forecasting.

The DUCSIM is an efficient, sophisticated instrument which can estimate how the urban crowd flows around or through streets as well as density of crowd in such street sections at certain time range intervals. The algorithm of this highly complex system is based on actual urban activity during various periods, pre-processed beforehand for data normalization and structuring. This M-o-M threshold analysis is crucial in determining if recent activities are higher than what was set as the baselines to forecast the future change of density for crowds. This provides its predictive accuracy which happens to be done through systematic balancing of current and historical data plus thresholds results. DUCSIM improves with iterative fine-tuning, hence its evolution is always applicable to different city scenarios. Despite the lack of specific numeric values and exact accuracy measurements mentioned in the cited paragraphs, the design of DUCSIM was intended to overcome data diversity and large volumes by using scalable computing architectures and techniques for merging data. DUCSIM assumes the role of a complex computer model that is indispensable in smart city administration, reemphasizes its flexibility, and indicates its suitability for evolving smart city dynamics.

10.2 Contributions of the Research

Based on the initial DUCSIM through to EDPAF, this suite of algorithms has made a vital contribution to crowd dynamics and urban management research. The first steps in developing modern algorithms used elementary approaches that served as a basis. The first two stages provided parameter settings for properly and numerically understanding crowd dynamics. The ensuing algorithms became more sophisticated and detailed, involving MOBILE tower-based recording, quartile classifications, and the inclusion of historical details. Such advancements were critical in improving the precision of the models and assisting in discerning crowd behaviour patterns in greater detail. Notably, Algorithm 6 introduced extensive threshold analysis per day or week, including quartile classification, for more holistic insight into crowd flows. This is a big move ahead in this area as it facilitated a more precise and thorough assessment of the evolution of crowd motions over time. Algorithm 7 symbolized this transition to active crowd management, confirming the algorithm's feasibility for real-time scenes. These algorithms enabled a preemptive crowd control strategy instrumental in urban planning and community security. Later algorithms included integrating individual mobility patterns, the sense of group identity, and social tie strength measurement.

After a long evolution, EDPAF is the ultimate of all these developments, integrating dynamic and self-learning features with up-to-date system parameters in real-time operation. This latest form of predictive modelling technology has shown exceptional versatility in adapting to the continuously shifting dynamics of urban masses and offers highly credible forecasts. Thus, this suite of progressing algorithms is crucial for crowd dynamics research, providing a full package of tools to handle crowds within cities. Every algorithm in this suite is a step towards increased technical ability and greater insight into the intricate dynamics of what influences crowd behaviour. However, the work forms useful groundwork for town planners, security police, and scientists and provides novel ways of coping with urban chaos in a globalizing environment.

10.3 Reflection on Enhanced DUCSIM's Journey

Regarding urban computational analysis, the Enhanced DUCSIM algorithm is a milestone because it integrates multiple data sources, revealing the intricacies of cities. This process started with observing the existence of a reliable tool to perceive and foresee urban changes. Over time, EDUCSIM has evolved into an exemplary prototype of urban complexity.

10.3.1 Comprehensive Data Integration

Enhanced DUCSIM is strong in terms of data integration. This algorithm vividly describes urban dynamics using information from Wi-Fi systems, traffic flow, ridesharing, and social media. Integrating this perspective into other facets of the city will reveal more details about how these parts intermingle and inform one another in new ways.

10.3.2 Dual Analysis: Macroscopic and Microscopic

The dual nature of Enhanced DDUCSIM in analysing macroscopic trends such as city movements and the density of crowds, among others, and microscopic behaviours like individual social interaction and sentiment analysis are revolutionary. This strategy offers a complete perspective critical in city planning and administration. This method provides small trends and patterns that may not be obvious using traditional analysis focusing only on one factor.

10.3.3 Predictive Modelling: A Leap into the Future

Enhanced DUCSIM is unique in its predicting capacities. Through sophisticated mathematical models, the algorithm reads present data and predicts new urban conditions in the future. The predictive analysis is necessary for proactive urban planning by city planners and others to anticipate and make preparations earlier for coming challenges and changes in urban dynamics.

10.3.4 Practical Applications and Broader Implications

The practical applications of Enhanced DUCSIM are very broad and diverse. The algorithm can, among other things, help with city planning and design, inform emergency response strategies, and contribute to sociological studies. This improves efficiency, safety, and responsiveness for the cities' inhabitants and helps reduce costs related to emergencies/disasters. When applied in a smart city project, it contributes to more sustainable and livable cities.

10.3.5 Navigating Challenges and Ethical Considerations

Despite its strong sides, Enhanced DUCSIM has its weaknesses to be overcome. Focus should be taken on the ethical implications of data use, especially users' data from social media or ride-sharing. Striving for balance, assuring privacy, and the right use of data by keeping the algorithm's effectiveness intact.

10.4 Future Directions: Scalability, Adaptability, and Technological Integration

The future development of Enhanced DUCSIM needs to focus on scalability and adaptability. Adapting the algorithm for various urban environments, dimensions, and cultures is critical for the method to be widely useful. By utilizing new technologies such as artificial intelligence and machine learning, efficiency can improve predictability, but there is a need to customize the future algorithm based on the real world's ever-changing dynamics. The future scope are further explained as follows:

Advanced Integration of Diverse Data Sources: Further research might consider a more seamless integration of different data sources, such as environmental sensors and public transport systems or extending to other social media platforms. It would also add to the data pool, so that our understanding of urban dynamics could be more nuanced.

Enhancement of Predictive Accuracy through AI and ML: Further integrating the most advanced Artificial Intelligence (AI) and Machine Learning-based techniques can further refine Enhanced DUCSIM's predictive ability. Thus research in this area can seek out more advanced algorithms which learn from a wide range of data inputs, yielding even greater accuracy for forecasts about the dynamics of cities.

Customization for Different Urban Environments: Similarly, the potential for varied urban settings--diverse in cultural contexts and geography as well as socio-economic backgrounds--to yield different scenarios is a key feature deserving continued investigation. In a later specific strategy, the algorithm can be modified according to different cities 'characteristics and problems around the world.

Ethical and Privacy Considerations in Data Use: Since Enhanced DUCSIM depends largely on such data provided by users, further research will need to explore the ethical aspects. This includes establishing rigorous privacy protection policies and data handling procedures that are as open to public scrutiny.

Application in Disaster Management and Emergency Response: For emergency response and disaster relief uses, Enhanced DUCSIM has a considerable amount of upside. In future studies, there is scope to explore how this tool can help make real-time decisions in response to crises. This has obvious safety and efficiency advantages.

Integration with Smart City Infrastructure: Integrating an Enhanced DUCSIM with existing and future smart city infrastructures is another area opened up. This would make for a more coordinated and overall approach to urban affairs, so that the people of our cities can live better.

Longitudinal Studies on Urban Dynamics: Long-term studies with Enhanced DUCSIM would give us a good idea of how cities change through time. This could help with long-term urban planning.

User Behaviour Analysis and Sociological Implications: Discussing sociological aspects of urban motion, whereby the group sets limits on individual behaviour or how community life affects various types of people, helps us get beyond numerical data to understand more clearly about human elements in urban planning.

Scalability and Performance Optimization: Increasing complex cities will be the test for Scalability of Enhanced DUCSIM. Going forward, the algorithm must be optimized for performance and scalability so that it is effective even with massive amounts of data.

Interdisciplinary Collaboration: By cooperating among different levels of expertise, such as in urban planning, computer science and sociology or environmental sciences, Enhanced DUCSIM can develop into a more complete solution covering almost every aspect of Urban Mobility.

The Enhanced DUCSIM is a critical advancement in urban analysis across these areas, i.e., technology, urban planning, and social science. With the growth of city tools like ENHANCED DUCSIM, people will need them as they navigate the intricacies of urban life. Therefore, developing this software will aid in building intelligent, flexible, and humane environments.

10.5 List of Publication

10.5.1 Patents

[1]. [202142038429 – Indian Patent: Publication Date: 03/09/2021]

[2021107444 – Australia Innovation Patent, Publication Date: 08/12/2021. Granted]

Title: Medical-IoT System for Estimating Hospital Beds Vacancy, Re-Routing Of Emergency Human Logistic Vehicle & Thereof.

https://patentscope.wipo.int/search/en/detail.jsf?docId=AU342862121

Innovators: Addepalli Lavanya Murali, Vidyasagar S.D., Ashutosh Verma, **Dr. Jaime Lloret Mauri**, Dr. Darsha Panwar, Dr. Navandar Yogeshwar, Dr. Prabhakar, C. J., Shubhangi Kachhawa

[2]. [IN202341052661} - SMART SURVEILLANCE FOR PUBLIC HEALTH: BIKER TRACKING AND CONTAGIOUS DISEASE MAPPING VIA ON-ROAD CAMERAS AND GEO-FENCED BLOCKCHAIN

https://patentscope.wipo.int/search/en/detail.jsf?docId=IN414847304

Inventors: Addepalli Lavanya, Darsha Panwar, Vidya Sagar S D, Jaime Lloret Mauri, Haritha Dasari, Dr. Vamsi Krishna Uppalapati, Maloth Bhavsingh

[3]. **[IN202341086155]: System and Method for Estimating Crowd Density and Assessing Social Networks Using Geo-Fenced Blockchain Environment and Wireless Data**

Inventors: Addepalli Lavanya Murali, Jaime Lloret Mauri

[4]. **[IN202341074214] – A system of Quantum-Enhanced Secure Lightwave Data Hyper-fusion for Li-Fi Communications and there of**

Inventors: Dr. Suman Rani, Dr. M. Sri Lakshmi, Addepalli Lavanya Murali, Jaime Lloret Mauri, Vidya Sagar S D, Maloth Bhavsingh

10.5.2 Other Network Patents

[5]. [IN202341074335] - UAV-Supported IoT Network Health Surveillance via Video Steganography Techniques

Inventors: Dr. Jaibir Singh, Dr. M. Kalpana Devi, K. Samunnisa, Addepalli Lavanya, Murali, Jaime Lloret Mauri, Vidya Sagar S D, Maloth Bhavsingh

[6]. [IN 202341066982] - A blockchain-hybrid Network System with Physiological Signal Integration of XAI Enhanced stress Reduction in Gaming

Inventors: Addepalli Lavanya Murali, Dr. Banoth Samya, Maloth Bhavsingh, Dr. Prasadu Peddi, Vidya Sagar S D , Jaime Lloret Mauri

[7]. [IN202341066991] Farm-to-Shelf Blockchain Network System for Unstructured Retailers Trading Among Small-Scale Producers

Inventors: Addepalli Lavanya Murali, Dr. Vishal Dattana, Joydeep Mookerjee, Sri Pooja Chavali, Vidya Sagar S D, Jaime Lloret Mauri, Maloth Bhavsingh

[8]. [IN202341074182] - Secured Airway Assessment System: A Hybrid Network-Blockchain Approach for Medical Image Evaluation

Inventors: Dr. Vamsi Krishna Uppalapati, B. Swarna Jyothi, Dr, N.V. Muthu Lakshmi, Addepalli Lavanya Murali, Vidya Sagar S D, Jaime Lloret Mauri, Maloth Bhavsingh

[9]. **[IN202241004223 - A CLOUD-IOT SYSTEM WITH GEO-FENCED BLOCKCHAIN TO IDENTIFY WAREHOUSE EMPTY SPACE & AUTOMATIC BILLING**

https://patentscope.wipo.int/search/en/detail.jsf?docId=IN350376787

Inventors: Addepalli Lavanya, **Dr. Jaime Lloret Mauri**, Vidya Sagar, S, D., Hemanth Shinde Dileep Naigapula, Dr. J. Bhavani, Dr. Navandar Yogeshwar, Dr. Alexis Bañón Gomis, Dr. Pablo Ruiz-Palomino

[10]. **[IN-202341050538] AUTONOMOUS TRAFFICGUARD - AI-ENABLED TRAFFIC ANALYSIS ECOSYSTEM FOR ENHANCED ROAD SAFETY**

https://patentscope.wipo.int/search/en/detail.jsf?docId=IN414862822

Inventors: JAIME LLORET MAURI, Addepalli Lavanya, Vidya Sagar S

[11]. **[IN202341053117] - A DECENTRALIZED PARKING NETWORK WITH LICENSE PLATE RECOGNITION AND AUTOMATED PAYMENTS ON THE BLOCKCHAIN**

https://patentscope.wipo.int/search/en/detail.jsf?docId=IN414845295

Inventors: Addepalli Lavanya, Jitendra Pandey, Vidya Sagar S D, Jaime Lloret Mauri, Maloth Bhavsingh, Vrinda Santosh Bhalerao, Shubhangi Kachhawa

10.5.3 Papers

[1]. Lavanya, A, Waqas Ali, Dr. Jaime Lloret, Vidya Sagar, S. D, and Chivukula Bharadwaj, "A Real-time Visualization of Global Sentiment Analysis on Declaration of Pandemic", Int. J. Comput. Eng. Res. Trends, vol. 9, no. 6, pp. 104–113, Jun. 2022.

10.5.4 Conference

[1]. Lavanya A., P. Darsha, P. Akhil, J. Lloret and N. Yogeshwar, "A Real-Time Human Mobility Visualization of Covid-19 Spread from East Asian Countries," 2021 Eighth International Conference on Social Network Analysis, Management and Security (SNAMS), 2021, pp. 1-8,
doi: 10.1109/SNAMS53716.2021.9732103.

10.5.5 Book chapter

[1]. **Lavanya. A**, Darsha Panwar, **Jaime Lloret,** Ali Waqas, Digvijay Pandey, Sagar Pathare, Anik Biswas (2022) Event Based Multi Model Classification to assess the User Participation Levels on Twitter – multi model classification of twitter user activity frequencies. Thakur, N., & Parameshachari, B. D. (Eds.). Human-Computer Interaction and Beyond: Advances Towards Smart and Interconnected Environments (Part II). Bentham Science Publishers.

REFERENCES

[1] A. Smith and B. Johnson, "Advances in Crowd Dynamics Modeling," *Urban Planning Journal*, vol. 15, no. 4, pp. 100–115, 2022.

[2] S. Di Martino and L. L. L. Starace, "Towards uniform urban map coverage in vehicular crowd-sensing: A decentralized incentivization solution," *IEEE Open J. Intell. Transp. Syst.*, vol. 3, pp. 695–708, 2022.

[3] M. Wu, B. Yan, Y. Huang, and M. N. I. Sarker, "Big data-driven urban management: Potential for urban sustainability," *Land (Basel)*, vol. 11, no. 5, pp. 680–689, 2022.

[4] E. Barka, C. A. Kerrache, N. Lagraa, and A. Lakas, "Behavior-aware UAV-assisted crowd sensing technique for urban vehicular environments," in *2018 15th IEEE Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA*, 2018, pp. 1–7.

[5] Y. Zhang, R. Zong, L. Shang, Z. Kou, H. Zeng, and D. Wang, "CrowdOptim: A crowd-driven neural network hyperparameter optimization approach to AI-based smart urban sensing," in *Proceedings of the ACM on Human-Computer Interaction*, *New York, NY, USA*, 2022, vol. 6, pp. 1–27.

[6] D. Stojanovic, B. Predic, and N. Stojanovic, "Mobile crowd sensing for smart urban mobility," in *European Handbook of Crowdsourced Geographic Information*, London, United Kingdom: Ubiquity Press, 2016, pp. 371–382.

[7] F. R. de la Rocha and M. Wangham, "Using a crowd-sensing strategy to support public transport tracking," in *Precision Positioning with Commercial Smartphones in Urban Environments*, S. Paiva, Ed. Switzerland : Springer Cham, 2021, pp. 29–51.

[8] J. Jeong, H. Jeong, J. Lim, J. Choi, S. Yun, and J. Y. Choi, "Selective ensemble network for accurate crowd density estimation," in *24th International Conference on Pattern Recognition (ICPR), Beijing, China*, 2018, pp. 1–6.

[9] K.-W. Yang *et al.*, "Improving crowd density estimation by fusing aerial images and radio signals," *ACM Trans. Multimed. Comput. Commun. Appl.*, vol. 18, no. 3, pp. 1–23, 2022.

[10] M.-N. Nguyen, V.-H. Tran, and T.-N. Huynh, "Depth embedded and dense dilated convolutional network for crowd density estimation," in *2021 International Conference on System Science and Engineering (ICSSE), Ho Chi Minh City, Vietnam*, 2021, pp. 221–225.

[11] S. Wang, Z. Pu, Q. Li, Y. Guo, and M. Li, "Estimating crowd density with edge intelligence based on lightweight convolutional neural networks," *Expert Systems with Applications: An International Journal*, vol. 206, no. C, pp. 1–7, Nov. 2022.

[12] H. Gao, M. Deng, D. Zhang, and W. Zhao, "MSFANet: Crowd density estimation based on multi-scale feature adaptation," *Academic Journal of Computing & Information Science*, vol. 4, no. 7, pp. 1–19, 2021.

[13] N. Boumaaz, H. Semlali, A. Maali, A. Soulmani, J. F. Diouris, and A. Ghammaz, "Low complexity spectrum sensing approach applying random sampling in Cognitive Radio networks," *IAENG International Journal of Computer Science*, vol. 49, no. 2, pp. 1–6, Jun.

2022.

[14] S. Rocke, “On random sampling for compliance monitoring in opportunistic spectrum access networks,” Ph.D. Thesis, Worcester Polytechnic Institute, Massachusetts, USA, 2013.

[15] T.-H. Yu, “Energy-efficient VLSI signal processing for wideband spectrum sensing in cognitive radios,” Ph.D. Thesis, University of California, Los Angeles, CA, USA, 2013.

[16] F. Delmastro, V. Arnaboldi, and M. Conti, “People-centric computing and communications in Smart Cities,” *Communications Magazine*, vol. 54, no. 7, pp. 122–128, 2016.

[17] H. Yoshikawa, A. Uchiyama, and T. Higashino, “TSVNet: Combining time-series and opportunistic sensing by transfer learning for dynamic thermal sensation estimation,” *IEEE Access*, vol. 9, pp. 102835–102846, 2021.

[18] L. Xiong, K. Tian, Y. Li, W. Ning, X. Gao, and Q. C. Zhang, “Online single-cell data integration through projecting heterogeneous datasets into a common cell-embedding space,” *Nat. Commun.*, vol. 13, no. 1, pp. 28–32, 2022.

[19] Y. Tan, W. Liu, J. Su, and X. Bai, “Generative adversarial networks based heterogeneous data integration and its application for intelligent power distribution and utilization,” *Appl. Sci. (Basel)*, vol. 8, no. 1, pp. 93–98, 2018.

[20] R. Reda, F. Piccinini, G. Martinelli, and A. Carbonaro, “Heterogeneous self-tracked health and fitness data integration and sharing according to a linked open data approach,” *Computing*, vol. 104, no. 4, pp. 835–857, 2022.

[21] B. Lee, S. Zhang, A. Poleksic, and L. Xie, “Heterogeneous multi-layered network model for omics data integration and analysis,” *Front. Genet.*, vol. 10, pp. 1–6, 2020.

[22] T. Reynolds *et al.*, “Interpretation of psychiatric genome-wide association studies with multispecies heterogeneous functional genomic data integration,” *Neuropsychopharmacology*, vol. 46, no. 1, pp. 86–97, 2021.

[23] S. Liu, Y.-Y. Fan, A. Samal, and Z. Guo, “Advances in computational facial attractiveness methods,” *Multimed. Tools Appl.*, vol. 75, no. 23, pp. 16633–16663, 2016.

[24] X. Chen, “Social media mental health analysis framework through applied computational approaches,” Ph.D. Thesis, Loughborough University, Leicestershire, England, 2020.

[25] B. Hambly, R. Xu, and H. Yang, “Recent advances in reinforcement learning in finance,” *Mathematical Finance*, vol. 33, no. 3, pp. 437–503, 2023.

[26] M. Varma *et al.*, “Identification of social engagement indicators associated with Autism spectrum disorder using a game-based mobile app: Comparative study of gaze fixation and visual scanning methods,” *J. Med. Internet Res.*, vol. 24, no. 2, p. e31830, 2022.

[27] G. Nwansi and A. Ikechi, “Brand choice influencing factors and Customer Preference Level of mobile network service providers,” *International Journal of Innovative Research in Social Sciences and Strategic Management Techniques*, vol. 10, no. 1, pp. 102–117, 2023.

[28] F. D. Priscoli and F. Muratore, “Radio and network comparisons for msbn and gsm systems in a geostationary satellite environment,” *Int. J. Satellite Commun.*, vol. 14, no. 6, pp. 439–

454, 1996.

[29] R. Alotaibi, B. Alzahrani, R. Wang, T. Alafif, A. Barnawi, and L. Hu, "Performance comparison and analysis for large-scale crowd counting based on convolutional neural networks," *IEEE Access*, vol. 8, pp. 204425–204432, 2020.

[30] J. Ferryman and A. Shahrokni, "An overview of the PETS 2009 challenge," in *Eleventh IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, Miami, Florida, USA.*, 2009, pp. 1–4.

[31] M. Abdelaal, N. D. Schiele, K. Angerbauer, K. Kurzhals, M. Sedlmair, and D. Weiskopf, "Comparative evaluation of bipartite, node-link, and matrix-based network representations," *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, vol. 29, no. 1, pp. 896–9076, 2022.

[32] Y. Wang *et al.*, "The HI/OH/Recombination line survey of the inner Milky Way (THOR): Data release 2 and HI overview," *Astronomy &*, vol. 595, no. A32, pp. 21–26, 2016.

[33] X. Jiang *et al.*, "Crowd counting and density estimation by trellis encoder-decoder network," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA*, 2019, pp. 6126–6135.

[34] B. Li, H. Huang, A. Zhang, P. Liu, and C. Liu, "Approaches on crowd counting and density estimation: a review," *Pattern Anal. Appl.*, vol. 24, no. 3, pp. 853–874, 2021.

[35] O. Elharrouss, N. Almaadeed, S. Al-Maadeed, K. Abualsaud, A. Mohamed, and T. Khattab, "Crowd counting Using DRL-based segmentation and RL-based density estimation," in *18th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Las Vegas, NV, USA*, 2022, pp. 1–7.

[36] V. A. Sindagi and V. M. Patel, "CNN-based cascaded multi-task learning of High-level prior and density estimation for crowd counting," in *Conference on Advanced Video and Signal Based Surveillance (AVSS), Lecce, Italy*, 2017, pp. 1–6.

[37] B. Ptak, D. Pieczyński, M. Piechocki, and M. Kraft, "On-board crowd counting and density estimation using low altitude unmanned aerial vehicles—looking beyond beating the benchmark," *Remote Sens. (Basel)*, vol. 14, no. 10, pp. 2288–2298, 2022.

[38] L. Yang *et al.*, "A crowd counting method via density map and counting residual estimation," *Multimed. Tools Appl.*, vol. 81, no. 30, pp. 43503–43512, 2022.

[39] Z. Fan, H. Zhang, Z. Zhang, G. Lu, Y. Zhang, and Y. Wang, "A survey of crowd counting and density estimation based on convolutional neural network," *Neurocomputing*, vol. 472, pp. 224–251, 2022.

[40] M. Tersek, M. Kljun, P. Peer, and Z. Emersic, "Re-evaluation of the CNN-based state-of-the-art crowd-counting methods with enhancements," *Comput. Sci. Inf. Syst.*, vol. 19, no. 3, pp. 1177–1198, 2022.

[41] Lijia Deng, Qinghua Zhou, Shuihua Wang, Juan Manuel Górriz, Yudong Zhang, "Deep learning in crowd counting: a survey," *CAAI Trans. Intell. Technol*, vol. 1, no. 35, pp. 1–12, 2023.

[42] Y. Meng *et al.*, "Spatial uncertainty-aware semi-supervised crowd counting," in *IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada*, 2021, pp.

15529–15539.

[43] T. Golda, F. Kruger, and J. Beyerer, "Temporal extension for encoder-decoder-based crowd counting approaches," in *2021 17th International Conference on Machine Vision and Applications (MVA), Aichi, Japan*, 2021, pp. 1–5.

[44] A. Kumar and M. Kumari, "Design and analysis of IoT-based system for crowd density estimation techniques," in *Advances in Data and Information Sciences*, Singapore: Springer, 2020, pp. 307–315.

[45] O. Kurnaz and C. Hanilci, "Features and regression techniques for crowd density estimation: A comparison," in *11th International Conference on Electrical and Electronics Engineering (ELECO), Bursa, Turkey*, 2019.

[46] K. Akhouri, S. Jain, T. Chhikara, R. Rani, P. Bansal, and A. Sharma, "Auto-encoders for crowd density estimation," in *IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI), Delhi, India*, 2022, pp. 12–17.

[47] X. Ma, W. Xi, X. Zhao, Z. Chen, H. Zhang, and J. Zhao, "Wisual: Indoor crowd density estimation and distribution visualization using WI-fi," *IEEE Internet Things J.*, vol. 9, no. 12, pp. 10077–10092, 2022.

[48] A. F. M. S. Saif and Z. R. Mahayuddin, "Crowd density estimation from autonomous Drones using deep learning: Challenges and applications," *J. Eng. Sci. Res.*, vol. 5, no. 6, pp. 1–6, 2021.

[49] C. Julien, "Opportunistic crowds: A place for device-to-device collaboration in pervasive crowd applications," in *IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), Kyoto, Japan*, 2019, pp. 21–26.

[50] W. Alasmary and S. Valaee, "Crowd sensing in vehicular networks using uncertain mobility information," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 11227–11238, 2019.

[51] A. Capponi, C. Fiandrino, D. Kliazovich, P. Bouvry, and S. Giordano, "A cost-effective distributed framework for data collection in cloud-based mobile crowd sensing architectures," *IEEE Trans. Sustain. Comput.*, vol. 2, no. 1, pp. 3–16, 2017.

[52] N. Thepvilojanapong, S. Konomi, Y. Tobe, Y. Ohta, M. Iwai, and K. Sezaki, "Opportunistic collaboration in participatory sensing environments," in *Proceedings of the fifth ACM international workshop on Mobility in the evolving internet architecture, New York, NY, USA*, 2010.

[53] M. A. Razzaque, "Editorial for special issue on 'Participatory and opportunistic mobile sensing: Applications and technologies,'" *Int. J. Distrib. Sens. Netw.*, vol. 13, no. 8, p. 155014771772766, 2017.

[54] C. A. Lowery *et al.*, "Defining the mode of action of tetramic acid antibacterials derived from Pseudomonas aeruginosa quorum sensing signals," *J. Am. Chem. Soc.*, vol. 131, no. 40, pp. 14473–14479, 2009.

[55] S. Kianoush, S. Savazzi, M. Beschi, S. Sigg, and V. Rampa, "A multisensory edge-cloud platform for opportunistic radio sensing in cobot environments," *IEEE Internet of Things Journal*, vol. 8, no. 2, pp. 1154–1168, 2021.

[56] M. Fiore, A. Nordio, and C. F. Chiasserini, “Driving factors toward accurate mobile opportunistic sensing in urban environments,” *IEEE Trans. Mob. Comput.*, vol. 15, no. 10, pp. 2480–2493, 2016.

[57] A. Ghasemi and E. S. Sousa, “Collaborative spectrum sensing for opportunistic access in fading environments,” in *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, (DySPAN 2005), Baltimore, MD, USA,* 2005.

[58] S. Abdul Rahman, A. Mourad, and M. El Barachi, “An infrastructure-assisted crowdsensing approach for on-demand traffic condition estimation,” *IEEE Access*, vol. 7, pp. 163323–163340, 2019.

[59] W. Gis, J. Gensel, and M. Tomko, “Web and wireless geographical information systems,” in *14th International symposium (W2GIS 2015), Grenoble, France,* 2015.

[60] P. Kortoci, A. Mehrabi, C. Joe-Wong, and M. Di Francesco, “Incentivizing Opportunistic Data Collection for Time-Sensitive IoT Applications,” in *18th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), Virtual Mode,* 2021.

[61] D. Zhao, H. Ma, Q. Li, and S. Tang, “A unified delay analysis framework for opportunistic data collection,” *Wirel. Netw.*, vol. 24, no. 4, pp. 1313–1325, 2018.

[62] Q. T. Minh, P. N. Huu, and T. Tsuchiya, “Mobile crowd-sourced data fusion and urban traffic estimation,” *J. Mob. Multimed.*, vol. 18, no. 4, pp. 1035–1062, 2022.

[63] A. Gupta and R. K. Bathla, “Energy efficient opportunistic sensing management in fog cloud environment,” *International Journal of Computer Science and Mobile Computing*, vol. 10, no. 10, pp. 20–26, 2021.

[64] S. Sigg and X. Fu, “Social opportunistic sensing and social centric networking: Enabling technology for smart cities,” in *Proceedings of the 2014 ACM international workshop on Wireless and mobile technologies for smart cities, Philadelphia, Pennsylvania, USA*, 2014, pp. 83–90.

[65] D. Basu, U. Ghosh, and R. Datta, “6G for industry 5.0 and smart CPS: A journey from challenging hindrance to opportunistic future,” in *IEEE Silchar Subsection Conference (SILCON), Silchar, India,* 2022, pp. 1–6.

[66] S. Gaito, E. Pagani, G. P. Rossi, and M. Zignani, “Sensing multi-dimensional human behavior in opportunistic networks,” in *Proceedings of the third ACM international workshop on Mobile Opportunistic Networks, Zürich, Switzerland*, 2012, pp. 89–90.

[67] P. B. E. Prabhakar and K. T. Sigamani, “Reliability enhancement using opportunistic routing in mobile ad-hoc networks,” *Indian J. Public Health Res. Dev.*, vol. 9, no. 3, p. 880, 2018.

[68] P. Soroye, N. Ahmed, and J. T. Kerr, “Opportunistic citizen science data transform understanding of species distributions, phenology, and diversity gradients for global change research,” *Glob. Chang. Biol.*, vol. 24, no. 11, pp. 5281–5291, 2018.

[69] R. Zhu, Q. Jiang, X. Huang, D. Li, and Q. Yang, “A reinforcement-learning-based opportunistic routing protocol for energy-efficient and void-avoided UASNs,” *IEEE Sens. J.*, vol. 22, no. 13, pp. 13589–13601, 2022.

[70] J. Xian, H. Wu, X. Mei, X. Chen, and Y. Yang, "Low-delay and energy-efficient opportunistic routing for maritime search and rescue wireless sensor networks," *Remote Sens. (Basel)*, vol. 14, no. 20, p. 5178, 2022.

[71] R. Shanmugam and B. Kaliaperumal, "An energy-efficient clustering and cross-layer-based opportunistic routing protocol (CORP) for wireless sensor network," *Int. J. Commun. Syst.*, vol. 34, no. 7, 2021.

[72] L. Pei, M. Zhang, D. Zou, R. Chen, and Y. Chen, "A survey of crowd sensing opportunistic signals for indoor localization," *Mob. Inf. Syst.*, vol. 2016, pp. 1–16, 2016.

[73] N. Thomas and S. Patil, "Trustworthiness of data in IoT crowd sensing environments," in *2022 International Conference on Futuristic Technologies (INCOFT), Belgaum, India*, 2022, pp. 966–970.

[74] S. Mehanna, Z. Kedad, and M. Chachoua, "Completeness assessment and improvement in mobile crowd-sensing environments," *SN Comput. Sci.*, vol. 3, no. 3, 2022.

[75] R. Gordon, D. K. Grimm, and F. Bai, "Crowd-sensing driving environments using headway dynamics," in *Proceedings of the 30th International Conference on Advances in Geographic Information Systems, Seattle, Washington, USA*, 2022, pp. 35–40.

[76] X. Jiang *et al.*, "Transferring priors from virtual data for crowd counting in real world," *Front. Comput. Sci.*, vol. 16, no. 3, pp. 1–4, 2022.

[77] S. Thalhammer, T. Patten, and M. Vincze, "SyDPose: Object detection and pose estimation in cluttered real-world depth images trained using only synthetic data," in *International Conference on 3D Vision (3DV), Québec City, Canada*, 2019.

[78] Z. Dang, L. Wang, Y. Guo, and M. Salzmann, "MatchNorm: Learning-based point cloud registration for 6D object pose estimation in the real world," *IEEE transactions on pattern analysis and machine intelligence*, vol. 10, no. 1, pp. 10–16, 2022.

[79] Z. Chu, S. Rathbun, and S. Li, "Learning Infomax and domain-Independent Representations for causal effect inference with real-world data," in *Proceedings of the 2022 SIAM International Conference on Data Mining (SDM), Virtual Mode*, 2022, pp. 433–441.

[80] K. Swami, A. Muduli, U. Gurram, and P. Bajpai, "Do what you can, with what you have: Scale-aware and high quality monocular depth estimation without real world labels," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), New Orleans, Louisiana, USA*, 2022, pp. 1–12.

[81] J. Moore, F. M. Rossi, M. A. Welsh, K. E. Nyffeler, and H. E. Blackwell, "A Comparative Analysis of Synthetic Quorum Sensing Modulators in Pseudomonas aeruginosa: New Insights into Mechanism, Active Efflux Susceptibility, Phenotypic Response, and Next-Generation Ligand Design," *J. Am. Chem. Soc.*, vol. 137, no. 46, pp. 14626–14639, 2015.

[82] R. Leonardi, F. Ragusa, A. Furnari, and G. M. Farinella, "Egocentric human-object interaction detection exploiting synthetic data," in *Image Analysis and Processing – ICIAP, Lecce, Italy*, 2022, pp. 24–30.

[83] J. Hu, O. Akande, and Q. Wang, "Multiple imputation and synthetic data generation with NPBayesImputeCat," *R J.*, vol. 13, no. 2, pp. 90–110, 2021.

[84] M. Cinquini, F. Giannotti, and R. Guidotti, “Boosting synthetic data generation with effective nonlinear causal discovery,” in *IEEE Third International Conference on Cognitive Machine Intelligence (CogMI), Atlanta, GA, USA*, 2023, pp. 54–63.

[85] M. Hernandez *et al.*, “Incorporation of synthetic data generation techniques within a controlled data processing workflow in the health and wellbeing domain,” *Electronics (Basel)*, vol. 11, no. 5, p. 812, 2022.

[86] Y. Zhou, S. De, W. Wang, K. Moessner, and M. Palaniswami, “Spatial indexing for data searching in mobile sensing environments,” *Sensors (Basel)*, vol. 17, no. 6, p. 1427, 2017.

[87] S. V. M. A. dos Reis D. M. Maletzke A. G. et al., “Challenges in benchmarking stream learning algorithms with real-world data,” *Data Min Knowl Disc*, vol. 34, pp. 1805–1858, 2020.

[88] K. Hiramatsu, A. Barrett, Y. Miyata, and PhRMA Japan Medical Affairs Committee Working Group 1, “Current status, challenges, and future perspectives of real-world data and real-world evidence in japan,” *Drugs Real World Outcomes*, vol. 8, no. 4, pp. 459–480, 2021.

[89] K. Togo and N. Yonemoto, “Real world data and data science in medical research: present and future,” *Jpn. J. Stat. Data Sci.*, vol. 5, no. 2, pp. 769–781, 2022.

[90] K. S. Rallis, D. Makrakis, I. A. Ziogas, and G. Tsoulfas, “Immunotherapy for advanced hepatocellular carcinoma: From clinical trials to real-world data and future advances,” *World J. Clin. Oncol.*, vol. 13, no. 6, pp. 448–472, 2022.

[91] O. Maqbool and J. Roßmann, “Formal scenario-driven logical spaces for randomized synthetic data generation,” in *Proceedings of the 10th International Conference on Model-Driven Engineering and Software Development, Vienna, Austria*, 2022.

[92] J. P. Scott and P. J. Werdell, “Comparing level-2 and level-3 satellite ocean color retrieval validation methodologies,” *Opt. Express*, vol. 27, no. 21, p. 30140, 2019.

[93] T. Oliveira, P. Carvalho, P. Ribeiro, and B. Bonatto, “PV hosting capacity dependence on harmonic voltage distortion in low-voltage grids: Model validation with experimental data,” *Energies*, vol. 11, no. 2, p. 465, 2018.

[94] J. F. Hair Jr, “Next-generation prediction metrics for composite-based PLS-SEM,” *Ind. Manag. Data Syst.*, vol. 121, no. 1, pp. 5–11, 2020.

[95] V. Ahlgren *et al.*, “Cray System Monitoring: Successes Requirements and Priorities,” in *USDOE National Nuclear Security Administration (NNSA), Albuquerque, NM, USA*, 2018.

[96] G. B. J. Mancini, “Digital coronary angiography: advantages and limitations,” in *Developments in Cardiovascular Medicine*, Dordrecht, Netherlands: Springer, 1991, pp. 23–42.

[97] L. Keohane and L. Sharma, “Improving care for individuals dually eligible for medicare and medicaid: Preliminary findings from recent evaluations of the financial alignment initiative,” *Commonwealth Fund Reports*, 2019.

[98] A. P. Klein, J. L. Ulmer, W. M. Mueller, F. D. Raslau, W. Gaggl, and M. Maheshwari, “DTI for Presurgical Mapping,” in *Functional Brain Tumor Imaging*, New York, NY, USA:

Springer, 2014, pp. 95–109.

[99] M. Shubik, “Game theory and operations research: Some musings 50 years later,” *Oper. Res.*, vol. 50, no. 1, pp. 192–196, 2002.

[100] D. Alzaleq, S. Alkhushayni, and A. FitzGerald, “Health tracker: data acquisition and analysis for monitoring health trends and assessing disease risk,” *Int. J. Eng. Technol.*, vol. 10, no. 1, p. 72, 2021.

[101] E. T. Subramaniam, B. Kumar, and R. Bhowmik, “Data Acquisition Systems-Current and future trends,” in *DAE Symposium on Nuclear Physics, New Delhi, India*, 2010, vol. 55, p. 1108.

[102] X. Ma and Korean Association For Learner-Centered Curriculum And Instruction, “An analysis of study trends on Korean tense,” *Korean Assoc. Learn.-Centered Curric. Instr.*, vol. 22, no. 13, pp. 147–169, 2022.

[103] W. P. Fendler, C. Bluemel, J. Czernin, and K. Herrmann, “PET imaging in prostate cancer, future trends: PSMA ligands,” *Clin. Transl. Imaging*, vol. 4, no. 6, pp. 467–472, 2016.

[104] J. G. Lu, “A social network perspective on the Bamboo Ceiling: Ethnic homophily explains why East Asians but not South Asians are underrepresented in leadership in multiethnic environments,” *J. Pers. Soc. Psychol.*, vol. 122, no. 6, pp. 959–982, 2022.

[105] E. Milani, E. Weitkamp, and P. Webb, “The visual vaccine debate on Twitter: A social network analysis,” *Media Commun.*, vol. 8, no. 2, pp. 364–375, 2020.

[106] K. L. Cela, M. Á. Sicilia, and S. Sánchez, “Social network analysis in E-learning environments: A preliminary systematic review,” *Educ. Psychol. Rev.*, vol. 27, no. 1, pp. 219–246, 2015.

[107] R. C. Perkins, M. Ouellet, C. J. Howell, and D. Maimon, “The illicit ecosystem of hacking: A longitudinal network analysis of website defacement groups,” *Soc. Sci. Comput. Rev.*, p. 089443932210978, 2022.

[108] A. Bobic, I. Jakovljevic, C. Gütl, J.-M. Le Goff, and A. Wagner, “Implicit user network analysis of communication platform open data for channel recommendation,” in *Ninth International Conference on Social Networks Analysis, Management and Security (SNAMS), Milan, Italy*, 2022, pp. 1–8.

[109] E. M. Daly and M. Haahr., “Social network analysis for information flow in disconnected delay-tolerant MANETs,” *IEEE Transactions on Mobile Computing*, vol. 8, no. 5, pp. 606–621, 2008.

[110] A. Galati, K. Djemame, and C. Greenhalgh, “Analysis of human mobility patterns for opportunistic forwarding in shopping mall environments,” *Soc. Netw. Anal. Min.*, vol. 5, no. 1, 2015.

[111] M. K. Mahbub Khan and M. Sajjadur Rahim, “Performance analysis of social-aware routing protocols in delay tolerant networks,” in *2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), Rajshahi, Bangladesh*, 2018, pp. 1–4.

[112] B. Kantarci, K. G. Carr, and C. D. Pearsall, “SONATA: Social Network Assisted Trustworthiness Assurance in smart city crowdsensing,” *Int. J. Distrib. Syst. Technol.*, vol.

7, no. 1, pp. 59–78, 2016.

[113] X. Wang *et al.*, “A city-wide real-time traffic management system: Enabling crowdsensing in social internet of vehicles,” *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 19–25, 2018.

[114] M. H. Cheung, F. Hou, and J. Huang, “Make a difference: Diversity-driven social mobile crowdsensing,” in *INFOCOM 2017 - IEEE Conference on Computer Communications, Atlanta, GA, USA*, 2017, p. 345.

[115] P. Bellavista, A. Corradi, L. Foschini, A. Noor, and A. Zanni, “Participact for smart and connected communities: Exploiting social networks with profile extension in crowdsensing systems,” in *Proceedings of the Workshop Program of the 19th International Conference on Distributed Computing and Networking, Varanasi India*, 2018, pp. 1–8.

[116] V. Grasso, A. Crisci, M. Morabito, P. Nesi, and G. Pantaleo, “Public crowdsensing of heat waves by social media data,” *Adv. Sci. Res.*, vol. 14, pp. 217–226, 2017.

[117] A. Bodaghi and B. Teimourpour, “Automobile Insurance Fraud Detection Using Social Network Analysis,” in *Applications of Data Management and Analysis*, Moshirpour, M., Far, B., Alhajj, R, Ed. Springer, Cham., 2018, pp. 11–16.

[118] S. Singh, P. Singh, and A. K. Tiwari, “Comprehensive analysis of social-based opportunistic routing protocol: A study,” in *Lecture Notes in Networks and Systems*, Singapore: Springer Singapore, 2020, pp. 69–79.

[119] A. Socievole, F. De Rango, and A. Caputo, “Wireless contacts, Facebook friendships and interests: Analysis of a multi-layer social network in an academic environment,” in *2014 IFIP Wireless Days (WD), Rio de Janeiro, Brazil*, 2014, pp. 1–7.

[120] G. Kang, J. Ning, and G. Kang, “Modeling social Gauss-Markov mobility for opportunistic network,” in *The Seventh International Conference on Wireless & Mobile Network (WiMo 2015) Sydney, Australia*, 2015, pp. 13–25.

[121] B. Czerkawski, “When learning analytics meets E-learning,” *Online Journal of Distance Learning Administration*, vol. 18, no. 2, pp. 1–5, 2015.

[122] Y. Wang, Z. Su, J. Ni, N. Zhang, and X. Shen, “Blockchain-empowered space-air-ground integrated networks: Opportunities, challenges, and solutions,” *IEEE Commun. Surv. Tutor.*, vol. 24, no. 1, pp. 160–209, 2022.

[123] B. Al Homssi *et al.*, “Next generation mega satellite networks for access equality: Opportunities, challenges, and performance,” *IEEE Commun. Mag.*, vol. 60, no. 4, pp. 18–24, 2022.

[124] M. M. Azari, S. Solanki, S. Chatzinotas, and M. Bennis, “THz-empowered UAVs in 6G: Opportunities, challenges, and trade-offs,” *IEEE Communications Magazine*, vol. 60, no. 5, pp. 24–30, 2022.

[125] R. Bajracharya, R. Shrestha, S. Kim, and H. Jung, “6G NR-U based wireless infrastructure UAV: Standardization, opportunities, challenges and future scopes,” *IEEE Access*, vol. 10, pp. 30536–30555, 2022.

[126] I. Yaqoob, K. Salah, R. Jayaraman, and Y. Al-Hammadi, “Blockchain for healthcare data management: opportunities, challenges, and future recommendations,” *Neural Comput.*

Appl., vol. 34, no. 14, pp. 11475–11490, 2022.

[127] Y. Chon, N. D. Lane, F. Li, H. Cha, and F. Zhao, “Automatically characterizing places with opportunistic crowdsensing using smartphones,” in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing, Pittsburgh Pennsylvania, USA*, 2012, pp. 481–490.

[128] S. Roy, D. Sarkar, S. Hati, and D. De, “Internet of Music Things: an edge computing paradigm for opportunistic crowdsensing,” *J. Supercomput.*, vol. 74, no. 11, pp. 6069–6101, 2018.

[129] S. Chen *et al.*, “A blockchain-based creditable and distributed incentive mechanism for participant mobile crowdsensing in edge computing,” *Math. Biosci. Eng.*, vol. 19, no. 4, pp. 3285–3312, 2022.

[130] S. S. Nagesh, N. Fernando, S. W. Loke, A. Neiat, and P. N. Pathirana, “Opportunistic mobile crowd computing: Task-dependency based work-stealing,” in *MobiCom ’22: Proceedings of the 28th Annual International Conference on Mobile Computing And Networking, Sydney NSW Australia*, 2022, pp. 775–777.

[131] V. Mahendran, R. Gunasekaran, and C. Siva Ram Murthy, “Performance modeling of delay-tolerant network routing via queueing Petri nets,” *IEEE Trans. Mob. Comput.*, vol. 13, no. 8, pp. 1816–1828, 2014.

[132] T. D. Adjin-Tettey, D. Selormey, and H. A. Nkansah, “Ubiquitous technologies and learning: Exploring perceived academic benefits of social media among undergraduate students,” *Int. J. Inf. Commun. Technol. Educ.*, vol. 18, no. 1, pp. 1–16, 2021.

[133] V. Das Swain, K. Saha, G. D. Abowd, and M. De Choudhury, “Social media and ubiquitous technologies for remote worker wellbeing and productivity in a post-pandemic world,” in *IEEE Second International Conference on Cognitive Machine Intelligence (CogMI), Atlanta, GA, USA*, 2020, pp. 121–130.

[134] D. Branco *et al.*, “Wild by design: Workshop on designing ubiquitous health monitoring technologies for challenging environments,” in *UbiComp-ISWC ’21 Adjunct: Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers, Virtual, USA*, 2021, pp. 508–510.

[135] G. Poizat, M. Fréjus, and Y. Haradji, “Analysis of activity in domestic settings for the design ubiquitous technologies,” in *European Conference on Cognitive Ergonomics (ECCE), Helsinki, Finland*, 2009, pp. 1–2.

[136] V. Petrov *et al.*, “Vehicle-based relay assistance for opportunistic crowdsensing over narrowband IoT (NB-IoT),” *IEEE Internet Things J.*, vol. 5, no. 5, pp. 3710–3723, 2018.

[137] A. Hassani, P. D. Haghighi, and P. P. Jayaraman, “Context-aware recruitment scheme for opportunistic mobile crowdsensing,” in *IEEE 21st International Conference on Parallel and Distributed Systems (ICPADS), Melbourne, VIC, Australia*, 2015, pp. 266–273.

[138] J. Xiong, M. Zhao, M. Z. A. Bhuiyan, L. Chen, and Y. Tian, “An AI-enabled three-party game framework for guaranteed data privacy in mobile edge crowdsensing of IoT,” *IEEE Trans. Industr. Inform.*, vol. 17, no. 2, pp. 922–933, 2021.

[139] V. K. Singh, A. S. Jasti, S. K. Singh, and S. Mishra, “QUAD: A quality aware multi-unit

double auction framework for IoT-based mobile crowdsensing in strategic setting," *IEEE Access*, vol. 10, pp. 1–24, 2022.

[140] M. Wang, "Optimizing multitask assignment of Internet of Things devices by reinforcement learning in mobile crowdsensing scenes," *Secur. Commun. Netw.*, vol. 2022, pp. 1–10, 2022.

[141] L. Perlepes, C. Smailis, D. Dres, D. Drakoulis, and J. Soldatos, "PLATO - intelligent middleware platform for the collection, analysis, processing of data from multiple heterogeneous sensor systems and application development for business intelligence," in *16th Panhellenic Conference on Informatics, Piraeus, Greece*, 2012, pp. 266–270.

[142] S. Palkar and E. Brunskill, "Analysis of the impact of errors made during health data collection using mobile phones: Exploring error modeling and automatic diagnosis," in *Proceedings of the 3rd ACM Symposium on Computing for Development, Bangalore India*, 2013, pp. 1–2.

[143] S. Fallmann, I. Psychoula, L. Chen, F. Chen, J. Doyle, and D. Triboan, "Reality and perception: Activity monitoring and data collection within a real-world smart home," in *IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), San Francisco, CA, USA*, 2017, pp. 1–6.

[144] D. M. Kebede, "Federalism and intergovernmental relations in Ethiopia: A critical analysis of the political challenges of federal-afar interaction (1991-2020)," *Global Journal of Human-Social Science*, vol. 22, no. F3, pp. 51–65, 2022.

[145] A. Muaremi, J. Seiter, G. Tröster, and A. Bexheti, "Monitor and understand pilgrims: Data collection using smartphones and wearable devices," in *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication, Zurich Switzerland*, 2013, pp. 679–688.

[146] J. Gama, "Issues and challenges in learning from data streams extended abstract," in *ICT Innovations 2014: Advances in Intelligent Systems and Computing, Ohrid, Macedonia*, 2014, pp. 1–5.

[147] Q. Duan, J. Huang, S. Hu, R. Deng, Z. Lu, and S. Yu, "Combining federated learning and edge computing toward ubiquitous intelligence in 6G network: Challenges, recent advances, and future directions," *IEEE Commun. Surv. Tutor.*, vol. 25, no. 4, pp. 2892–2950, 2023.

[148] S. Shu and B. K. P. Woo, "Use of technology and social media in dementia care: Current and future directions," *World J. Psychiatry*, vol. 11, no. 4, pp. 109–123, 2021.

[149] S. Tanwar, A. Popat, P. Bhattacharya, R. Gupta, and N. Kumar, "A taxonomy of energy optimization techniques for smart cities: Architecture and future directions," *Expert Syst.*, vol. 39, no. 5, 2022.

[150] A. Leemann, "Apps for capturing language variation and change in German-speaking Europe: Opportunities, challenges, findings, and future directions," *Linguist. Vanguard*, vol. 7, no. s1, 2021.

[151] S. Tuli, S. Ilager, K. Ramamohanarao, and R. Buyya, "Dynamic scheduling for stochastic edge-cloud computing environments using A3C learning and residual recurrent neural

networks," *IEEE Transactions on Mobile Computing*, vol. 21, no. 3, pp. 940–954, 2020.

[152] N. Thakur and C. Y. Han, "An ambient intelligence-based human behavior monitoring framework for ubiquitous environments," *Information*, vol. 12, no. 2, pp. 81–90, 2021.

[153] C.-T. Li, C.-C. Lee, and C.-Y. Weng, "An extended chaotic maps based user authentication and privacy preserving scheme against DoS attacks in pervasive and ubiquitous computing environments," *Nonlinear Dyn.*, vol. 74, no. 4, pp. 1133–1143, 2013.

[154] J. Sarivougioukas, A. Vagelatos, and I. Lagaris, "Administration of medical contexts with Denotational Mathematics in ubiquitous computing home environments," *Int. J. Softw. Sci. Comput. Intell.*, vol. 7, no. 2, pp. 1–30, 2015.

[155] M. De Lourdes Martinez-Villasenor and M. Gonzalez-Mendoza, "Sharing and reusing context information in ubiquitous computing environments," in *13th Mexican International Conference on Artificial Intelligence, Tuxtla Gutierrez, Mexico*, 2014, pp. 227–230.

[156] A. A. Assefa, W. Tian, N. W. Hundera, and M. U. Aftab, "Crowd density estimation in spatial and temporal distortion environment using parallel multi-size receptive fields and stack ensemble meta-learning," *Symmetry (Basel)*, vol. 14, no. 10, p. 2159, 2022.

[157] M. A. Khan, H. Menouar, and R. Hamila, "DroneNet: Crowd Density Estimation using Self-ONNs for Drones," in *IEEE 20th Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA*, 2023, pp. 455–460.

[158] M. A. Khan, H. Menouar, and R. Hamila, "Crowd Density Estimation using Imperfect Labels," in *IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA*, 2023, pp. 1–6.

[159] Y.-C. Li, R.-S. Jia, Y.-X. Hu, and H.-M. Sun, "A lightweight dense crowd density estimation network for efficient compression models," *Expert Syst. Appl.*, vol. 238, no. 122069, p. 122069, 2024.

[160] M. A. Khan, H. Menouar, and R. Hamila, "LCDnet: A Lightweight Crowd Density estimation model for real-time video surveillance," *Journal of Real-Time Image Processing*, vol. 20, no. 29, pp. 1–20, 2023.

[161] A. Subramaniam, V. G. Shah, P. D. Tibrewala, and R. Kale, "Crowd Count Estimator Application using CSRNet," in *2021 International Conference on Intelligent Technologies (CONIT), Mumbai India*, 2021, pp. 1–4.

[162] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, "Single-image crowd counting via multi-column convolutional neural network," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA*, 2016, pp. 589–597.

[163] S. Pan, Y. Zhao, F. Su, and Z. Zhao, "SANet++: Enhanced scale aggregation with densely connected feature fusion for crowd counting," in *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2021, Toronto, ON, Canada*, 2021, pp. 1980–1984.

[164] W. Shu, J. Wan, and A. B. Chan, "Generalized characteristic function loss for crowd analysis in the frequency domain," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. PP, 2023.

[165] Z. Ma, X. Wei, X. Hong, H. Lin, Y. Qiu, and Y. Gong, "Learning to count via unbalanced optimal transport," in *Proceedings of the AAAI Conference on Artificial Intelligence, Virtual*

Mode, 2021, vol. 35, pp. 2319–2327.

[166] Z. Zheng, N. Ni, G. Xie, A. Zhu, Y. Wu, and T. Yang, “HARNet: Hierarchical adaptive regression with location recovery for crowd counting,” *Neurocomputing*, vol. 513, pp. 329–340, 2022.

[167] Y. E. Kara, G. Genc, O. Aran, and L. Akarun, “Modeling annotator behaviors for crowd labeling,” *Neurocomputing*, vol. 160, pp. 141–156, 2015.

[168] W. R. Huang, T. N. Sainath, C. Peyser, S. Kumar, D. Rybach, and T. Strohman, “Lookup-table recurrent language models for long tail speech recognition,” in *Interspeech 2021*, 2021.

[169] D. Prasad, A. Gadpal, K. Kapadni, M. Visave, and K. Sultanpure, “CascadeTabNet: An approach for end to end table detection and structure recognition from image-based documents,” in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA*, 2020, pp. 2439–2447.

[170] J. Gao, H. Jie, B. Xu, L. Liu, J. Hu, and W. Liu, “LCDeT: LiDAR curb detection network with transformer,” in *2023 International Joint Conference on Neural Networks (IJCNN), Gold Coast, Australia*, 2023, pp. 1–9.

[171] X. Wang, C. Li, and D. Song, “CrowdNet: Identifying large-scale malicious attacks over android kernel structures,” *IEEE Access*, vol. 8, pp. 15823–15837, 2020.

[172] E. Hodgson, “Human Environments,” in *Progress in Molecular Biology and Translational Science*, vol. 112, Elsevier, 2012, pp. 1–10.

[173] J. Liu *et al.*, “Device-free crowd density estimation with off-the-shelf WiFi traffic,” in *Tenth International Conference on Advanced Cloud and Big Data (CBD), Guilin, China*, 2022, pp. 306–311.

[174] K. Shibata and H. Yamamoto, “People crowd density estimation system using deep learning for radio wave sensing of cellular communication,” in *International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Okinawa, Japan*, 2019, pp. 143–148.

[175] X. Zhang, Y. Sun, Q. Li, X. Li, and X. Shi, “Crowd density estimation and mapping method based on surveillance video and GIS,” *ISPRS Int. J. Geoinf.*, vol. 12, no. 2, p. 56, 2023.

[176] S. Kyosuke and Y. Hiroshi, “Development and evaluation of people crowd density estimation system using broadband radio wave sensing of cellular communication,” in *International conference on artificial intelligence in information and communication (ICAIIC), Biwako-Kusatsu, Japan*, 2018, pp. 143–148.

[177] M. De Sanctis, T. Rossi, S. Di Domenico, E. Cianca, G. Ligresti, and M. Ruggieri, “LTE signals for device-free crowd density estimation through CSI secant set and SVD,” *IEEE Access*, vol. 7, pp. 159943–159951, 2019.

[178] D. Sammel, M. Boehler, A. J. Gamel, and M. Schumacher, “Lightweight integration of a data cache for opportunistic usage of HPC resources in HEP workflows,” *Comput. Softw. Big Sci.*, vol. 7, no. 1, 2023.

[179] M. Grabow *et al.*, “Data-integration of opportunistic species observations into hierarchical modeling frameworks improves spatial predictions for urban red squirrels,” *Front. Ecol.*

Evol., vol. 10, 2022.

[180] E. C. Hector and R. Martin, "Turning the information-sharing dial: efficient inference from different data sources," *arXiv [stat.ME]*, 2022.

[181] V. K. Dimitriadis *et al.*, "An open-source platform integrating emerging data sources to support multi-modal active pharmacovigilance," *Front. Drug Saf. Regul.*, vol. 2, 2023.

[182] A. Bernasconi, A. Canakoglu, M. Masseroli, P. Pinoli, and S. Ceri, "A review on viral data sources and search systems for perspective mitigation of COVID-19," *Brief. Bioinform.*, vol. 22, no. 2, pp. 664–675, 2021.

[183] S. Panacek, F. Rademakers, and R. Brun, "ROOT, an object oriented data analysis framework," *Linux Journal*, vol. 51, no. 1, p. 27, 1998.

[184] W. Saber, R. Eisa, and R. Attia, "Efficient geospatial data analysis framework in fog environment," *IEEE Access*, vol. 10, pp. 133591–133600, 2022.

[185] S. Singh Kumar, J. Cha, T. Kim Woo, and J. Park Hyuk, "Machine learning based distributed big data analysis framework for next generation web in IoT," *Comput. Sci. Inf. Syst.*, vol. 18, no. 2, pp. 597–618, 2021.

[186] E. B. Varghese, S. M. Thampi, and S. Berretti, "A psychologically inspired fuzzy cognitive deep learning framework to predict crowd behavior," *IEEE Trans. Affect. Comput.*, vol. 13, no. 2, pp. 1005–1022, 2022.

[187] X. Yan and Z. Fang, "Application of psychological hint and crowd behavior in risk management of public-private partnership project," *REV. ARGENT. DE CLIN. PSICOL.*, 2020.

[188] T. Simpson, "Real-time drone surveillance system for violent crowd behavior unmanned aircraft system (UAS) – human autonomy teaming (HAT)," in *IEEE/AIAA 40th Digital Avionics Systems Conference (DASC), San Antonio, TX, USA*, 2021, pp. 1–9.

[189] Y. Li, K. Li, and L. Wang, "Crowd behavior simulation model based on information-gap theory," in *6th International Conference on Intelligent Computing and Signal Processing (ICSP), Xi'an, China*, 2021, pp. 590–594.

[190] X. Zhang, Q. Zhang, S. Hu, C. Guo, and H. Yu, "Energy level-based abnormal crowd behavior detection," *Sensors (Basel)*, vol. 18, no. 2, p. 423, 2018.

[191] X. Gong, M. Herty, B. Piccoli, and G. Visconti, "Crowd dynamics: Modeling and control of multiagent systems," *Annu. Rev. Control Robot. Auton. Syst.*, vol. 6, no. 1, 2023.

[192] R. M. Colombo, M. Gokieli, and M. D. Rosini, "Modeling crowd dynamics through hyperbolic - elliptic equations," in *Partial Differential Equations, Mathematical Physics, and Stochastic Analysis A Volume in Honor of Helge Holden's 60th Birthday*, European Mathematical Society, 2018, pp. 111–128.

[193] D. Kim and A. Quaini, "A 2D kinetic model for crowd dynamics with disease contagion," in *Modeling and Simulation in Science, Engineering and Technology. Birkhäuser*, Bellomo, N., Chaplain, M.A.J., Ed. Cham, Switzerland: Springer International Publishing, 2021, pp. 265–284.

[194] Z. Li, P. Zhao, C. Jiang, W. Huang, and H. Liang, “A learning-based model predictive trajectory planning controller for automated driving in unstructured dynamic environments,” *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 5944–5959, 2022.

[195] G.-N. Wang, T. Chen, J.-W. Chen, K. Deng, and R.-D. Wang, “Simulation of crowd dynamics in pedestrian evacuation concerning panic contagion: A cellular automaton approach,” *Chin. Physics B*, vol. 31, no. 6, p. 060402, 2022.

[196] P. Yuan, X. Pang, and X. Zhao, “Influence of crowd participation features on mobile edge computing,” *Future Internet*, vol. 10, no. 10, p. 94, 2018.

[197] A. Ahmad, M. A. Rahman, B. Sadiq, S. Mohammed, S. Basalamah, and M. R. Wahiddin, “Visualization of a scale free network in a smartphone-based multimedia big data environment,” in *2015 IEEE International Conference on Multimedia Big Data, Beijing, China*, 2015, pp. 286–287.

[198] C. Cheng and J. Li, “ODEs learn to walk: ODE-Net based data-driven modeling for crowd dynamics,” *arXiv [cs.LG]*, 2022.

[199] Z. Jiang, Y. Liu, X. Fan, C. Wang, J. Li, and L. Chen, “Understanding urban structures and crowd dynamics leveraging large-scale vehicle mobility data,” *Front. Comput. Sci.*, vol. 14, no. 5, 2020.

[200] T. Mao, J. Wang, R. Meng, Q. Yan, S. Liu, and Z. Wang, “A fusion crowd simulation method: Integrating data with dynamics, personality with common,” *Comput. Graph. Forum*, vol. 41, no. 8, pp. 131–142, 2022.

[201] Z. Zhao, G. Xu, N. Zhang, and Q. Zhang, “Performance analysis of the hybrid satellite-terrestrial relay network with opportunistic scheduling over generalized fading channels,” *IEEE Trans. Veh. Technol.*, vol. 71, no. 3, pp. 2914–2924, 2022.

[202] C. Malesios, “Some variations on the standard theoretical models for the h-index: A comparative analysis,” *Journal of the Association for Information Science and Technology*, vol. 66, no. 11, pp. 2384–2388, 2015.

[203] C. Ziółkowski and J. M. Kelner, “Empirical models of the azimuthal reception angle—part I: Comparative analysis of empirical models for different propagation environments,” *Wirel. Pers. Commun.*, vol. 91, no. 2, pp. 771–791, 2016.

[204] M. Garcia, C. Martinez, and J. Lloret, *Wireless Sensors self-location in an Indoor WLAN environment, International Conference on Sensor Technologies and Applications (SENSORCOMM 2007)*. Valencia (España), 2007.

[205] M. Garcia, “The Development of Two Systems for Indoor Wireless Sensors Self-location,” *Ad Hoc Sensor Wireless Networks*, vol. 8, no. 3–4, pp. 235–258, 2009.

[206] Z. Liu, L. Li, F. Tian, and J. Jia, “Lightweight Web3D crowd rendering for online massive conferencing,” in *IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), Singapore*, 2022, pp. 536–541.

[207] A. Panayiotou, T. Kyriakou, M. Lemonari, Y. Chrysanthou, and P. Charalambous, “CCP: Configurable Crowd Profiles,” in *Special Interest Group on Computer Graphics and Interactive Techniques Conference Proceedings, Vancouver BC Canada*, 2022, pp. 1–10.

[208] C. Li, R. Zhao, C. Han, R. Arifur, Y. Ma, and Q. Liu, “Dynamic disturbance propagation model of pedestrian panic behaviors and Lyapunov-based crowd stability analysis,” *Appl. Sci. (Basel)*, vol. 13, no. 21, p. 11762, 2023.

[209] L. Arp, D. van Vreumingen, D. Gawehns, and M. Baratchi, “Dynamic macro scale traffic flow optimisation using crowd-sourced urban movement data,” in *21st IEEE International Conference on Mobile Data Management (MDM), Versailles, France*, 2020, pp. 168–177.

[210] Z. Cai, F. Ren, Y. Chi, X. Jia, L. Duan, and Z. Ding, “Multi-vehicles dynamic navigating method for large-scale event crowd evacuations,” *Geoinformatica*, vol. 22, no. 2, pp. 435–462, 2018.

[211] I. Joshi, M. Grimmer, C. Rathgeb, C. Busch, F. Bremond, and A. Dantcheva, “Synthetic data in human analysis: A survey,” *arXiv [cs.CV]*, 2022.

[212] W. Hahn *et al.*, “Contribution of synthetic data generation towards an improved patient stratification in palliative care,” *J. Pers. Med.*, vol. 12, no. 8, p. 1278, 2022.

[213] K. Binici, N. T. Pham, T. Mitra, and K. Leman, “Preventing catastrophic forgetting and distribution mismatch in knowledge distillation via synthetic data,” in *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA*, 2021, pp. 3625–3633.

[214] G. I. Dzhardimalieva and I. E. Uflyand, “Synthetic methodologies and spatial organization of metal chelate dendrimers and star and hyperbranched polymers,” *Dalton Trans.*, vol. 46, no. 31, pp. 10139–10176, 2017.

[215] K. Kishida, “Empirical Comparison of External Evaluation Measures for Document Clustering by Using Synthetic Data,” *IPSJ SIG Technical Report*, vol. 113, no. 1, pp. 1–7, 2014.

[216] K. El Emam, L. Mosquera, X. Fang, and A. El-Hussuna, “Utility metrics for evaluating synthetic health data generation methods: Validation study,” *JMIR Med. Inform.*, vol. 10, no. 4, p. e35734, 2022.

[217] S. Bambach and M.-L. Ho, “Deep learning for synthetic CT from bone MRI in the head and neck,” *AJNR Am. J. Neuroradiol.*, vol. 43, no. 8, pp. 1172–1179, 2022.

[218] A. Thummerer *et al.*, “SynthRAD2023 Grand Challenge dataset: generating synthetic CT for radiotherapy,” *Med Phys*, vol. 50, no. 1, pp. 4664–4674, 2023.

[219] Z. Car, B. Šegota, Sandi, and B. Dobraš, “Synthetic data generation for biomedical deep learning: Methods, challenges, and opportunities,” in *2nd International Conference on Chemo and BioInformatics, Serbia*, 2023, pp. 26–33.

[220] C. Maxey, J. Choi, H. Lee, D. Manocha, and H. Kwon, “UAV-Sim: NeRF-based synthetic data generation for UAV-based perception,” *arXiv [cs.CV]*, 2023.

[221] S. Asre and A. Anwar, “Synthetic energy data generation using time variant generative adversarial network,” *Electronics (Basel)*, vol. 11, no. 3, p. 355, 2022.

[222] S. Imtiaz, M. Arsalan, V. Vlassov, and R. Sadre, “Synthetic and private smart health care data generation using GANs,” in *International Conference on Computer Communications and Networks (ICCCN), Athens, Greece*, 2021, pp. 1–7.

[223] M. Dimopoulou and M. Antonini, “Data and image storage on synthetic DNA: existing solutions and challenges,” *EURASIP J. Image Video Process.*, vol. 2022, no. 1, 2022.

[224] A. Gonzales, G. Guruswamy, and S. R. Smith, “Synthetic data in health care: A narrative review,” *PLOS Digit. Health*, vol. 2, no. 1, p. e0000082, 2023.

[225] S. R. Richter, “Visual perception with synthetic data,” Ph.D. Thesis, Technical University, Darmstadt, 2020.

[226] A. M. Kamashev, A. A. Duchkov, and G. S. Chernyshov, “Development of an algorithm for the suppression of multiple based on the decomposition of seismic data into wave packets,” *Russian Journal of geophysical technologies*, no. 3, pp. 28–40, 2022.

[227] D. Kurniadi, F. Nuraeni, N. Abania, L. Fitriani, A. Mulyani, and Y. H. Agustin, “Scholarship recipients prediction model using k-nearest neighbor algorithm and synthetic minority over-sampling technique,” in *12th International Conference on System Engineering and Technology (ICSET), Bandung, Indonesia*, 2022, pp. 89–94.

[228] J. Gandhi and Z. Narmawala, “Fair comparative analysis of opportunistic routing protocols: An empirical study,” in *Advances in Intelligent Systems and Computing*, Singapore: Springer Singapore, 2020, pp. 285–294.

[229] E. H. Waller, A. Keil, and F. Friederich, “Quantum range-migration-algorithm for synthetic aperture radar applications,” *Sci. Rep.*, vol. 13, no. 1, 2023.

[230] W. Wang, S. L. Capitaneanu, D. Marinca, and E.-S. Lohan, “Comparative analysis of channel models for industrial IoT wireless communication,” *IEEE Access*, vol. 7, pp. 91627–91640, 2019.

[231] J. Lloret, J. Tomas, A. Canovas, and I. Bellver, “Geowifi: A geopositioning system based on wifi networks,” in *International Conference on Networking and Services, Venice/Mestre, Italy*, 2011, pp. 38–43.

[232] Y. He, X. Zu, Y. Xia, Y. Wang, and B. Yin, “Ghost attention efficient network for crowd counting,” in *7th International Conference on Computer and Communication Systems (ICCCS), Wuhan, China*, 2022, pp. 289–293.

[233] A. Caporuscio, M. C. Pietronudo, F. Schiavone, and D. Leone, “Unlocking value with a crowdsourcing configuration of smart city: a system dynamic simulation,” *TQM J.*, vol. 20, no. 1, pp. 1–19, 2022.

[234] M. Garcia, D. Bri, J. Tomas, and J. Lloret, “A cooperative decision making algorithm for wireless location systems using interlinking data,” in *Lecture Notes in Computer Science*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 85–92.

[235] S. Gong, E.-B. Bourennane, and X. Yang, “MFNet: Multi-feature convolutional neural network for high-density crowd counting,” in *11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada*, 2020, pp. 0384–0390.

[236] J. Doe, “Telecommunications Data in Urban Crowd Analysis,” *Journal of Urban Technology*, vol. 10, no. 2, pp. 35–50, 2021.

[237] Y. Shitara, T. Yamazaki, K. Yamori, and T. Miyoshi, “Congestion field detection for Service Quality improvement using Kernel density estimation,” in *18th Asia-Pacific Network*

Operations and Management Symposium (APNOMS), Kanazawa, Japan, 2016, pp. 1–3.

[238] Â. Frimane and J. M. Bright, “Validation of synthetic solar irradiance data,” in *Synthetic Solar Irradiance*, AIP Publishing LLCMelville, New York, 2021, pp. 4-1-4–44.

[239] L. Tonetto, M. Untersperger, and J. Ott, “Towards exploiting WI-fi signals from low density infrastructure for crowd estimation,” in *Proceedings of the 14th Workshop on Challenged Networks, Los Cabos Mexico*, 2019, pp. 27–32.

[240] Y. Kawamoto, N. Yamada, H. Nishiyama, N. Kato, Y. Shimizu, and Y. Zheng, “A feedback control-based crowd dynamics management in IoT system,” *IEEE Internet Things J.*, vol. 4, no. 5, pp. 1466–1476, 2017.

[241] S. Denis, R. Berkvens, B. Bellekens, and M. Weyn, “Large scale crowd density estimation using a sub-GHz wireless sensor network,” in *IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Bologna, Italy*, 2018, pp. 849–855.

APPENDIX – CODE

1. Code: Algorithm 6

Code 1: Algorithm 6

```
import pandas as pd

# Example DataFrame structure
data = pd.DataFrame({
    'Time': ['00:00:00', '01:00:00', '02:00:00', '00:00:00', '01:00:00'],
    'Ref_Count': [100, 150, 200, 110, 160],
    'Current_Count': [120, 180, 210, 105, 165]
})

# Convert Time to datetime and extract hour
data['Time'] = pd.to_datetime(data['Time']).dt.hour

# Define the initial classification function
def classify_initial(current_count, median_ref_count):
    return 'Above Median' if current_count >= median_ref_count else 'Below Median'

# Define the quartile-based classification function
def classify_quartiles(current_count, q1, q3):
    if current_count < q1:
        return 'Low'
    elif current_count < q3:
        return 'Medium'
    else:
        return 'High'

# Process data
results = []
for hour in data['Time'].unique():
    # Filter data for the current hour
    hour_data = data[data['Time'] == hour]

    # Step 1: Calculate the median Ref_Count
    median_ref_count = hour_data['Ref_Count'].median()

    # Step 2: Initial Classification
    hour_data['Initial_Classification'] = hour_data['Current_Count'].apply(classify_initial,
args=(median_ref_count,))

    # Step 3: Further Classification
    for initial_class in ['Above Median', 'Below Median']:
        class_data = hour_data[hour_data['Initial_Classification'] == initial_class]
        q1 = class_data['Current_Count'].quantile(0.25)
        q3 = class_data['Current_Count'].quantile(0.75)
```

```
        class_data['Final_Classification'] =
class_data['Current_Count'].apply(classify_quartiles, args=(q1, q3))
        results.append(class_data)

# Compile final results
final_results = pd.concat(results)

print(final_results.head())
```

2. Code: Algorithm 7

Code 2: Algorithm 7

```
import pandas as pd
import numpy as np

# Sample input data
data = pd.DataFrame({
    'Time': ['00:00:00', '01:00:00'],
    'Ref_Count': [4921, 5356],
    'Current_Count': [6607, 7777]
})

# Define a function to classify records as "Above Median" or "Below Median"
def classify_initial(record, median):
    if record['Ref_Count'] >= median:
        return "Above Median"
    else:
        return "Below Median"

# Define a function to classify records based on quartiles
def classify_quartiles(record, q1, q3):
    if record['Ref_Count'] < q1:
        return "Low"
    elif q1 <= record['Ref_Count'] < q3:
        return "Medium"
    else:
        return "High"

# Extract distinct hours
hours = data['Time'].str.split(':').str[0].unique()

# Create an empty DataFrame for the final classification
final_classification = pd.DataFrame()

# Create an empty DataFrame to store threshold values
threshold_values = pd.DataFrame(columns=['Hour', 'Median', 'Q1', 'Q3'])

for hour in hours:
    # Extract records for the current hour
    hour_data = data[data['Time'].str.startswith(hour)]

    # Calculate the median for this hour
    median_t = hour_data['Ref_Count'].median()

    # Perform initial classification and add it to the DataFrame
    hour_data['InitialClass'] = hour_data.apply(lambda x: classify_initial(x, median_t),
axis=1)
```

```
    # Calculate quartiles for the classified group
    q1_t_class = hour_data[hour_data['InitialClass'] == 'Below
Median']['Ref_Count'].quantile(0.25)
    q3_t_class = hour_data[hour_data['InitialClass'] == 'Below
Median']['Ref_Count'].quantile(0.75)

    # Store threshold values in the threshold_values DataFrame
    threshold_values = threshold_values.append({'Hour': hour, 'Median': median_t, 'Q1':
q1_t_class, 'Q3': q3_t_class}, ignore_index=True)

    # Perform further classification and add it to the DataFrame
    hour_data['FinalClass'] = hour_data.apply(lambda x: classify_quartiles(x, q1_t_class,
q3_t_class), axis=1)

    # Append the hour's data to the final classification DataFrame
    final_classification = final_classification.append(hour_data)

# Print the final classification
print(final_classification)

# Save threshold values to CSV
threshold_values.to_csv('threshold_values.csv', index=False)

# Save final classification to CSV
final_classification.to_csv('final_classification.csv', index=False)
```

3. Code: Algorithm 8

Code 3: Algorithm 8

```
import pandas as pd
import numpy as np

# Load the CSV file into a DataFrame
input_file = 'Input_Synthetic.csv'
df = pd.read_csv(input_file)

# Initialization
D_grouped = df.groupby('Time')
T = {}

# Compute Hourly Median Thresholds
for hour, group in D_grouped:
    median_h = np.median(group['Current_Count'])
    T[hour] = median_h

# Categorization of Counts into Above and Below Threshold
def categorize_count(row):
    hour = row['Time']
    if row['Current_Count'] > T[hour]:
        return 'Above Threshold'
    else:
        return 'Below Threshold'

df['Threshold_Category'] = df.apply(categorize_count, axis=1)

# Quartile Classification based on Threshold
Q1 = df['Current_Count'].quantile(0.25)
Q2 = df['Current_Count'].quantile(0.5)
Q3 = df['Current_Count'].quantile(0.75)

def classify_quartile(row):
    if row['Current_Count'] <= Q1:
        return 'Low Density'
    elif Q1 < row['Current_Count'] <= Q2:
        return 'Medium Density'
    elif Q2 < row['Current_Count'] <= Q3:
        return 'High Density'
    else:
        return 'Very High Density'

df['Density_Range'] = df.apply(classify_quartile, axis=1)

# Compilation of Results
R = df.copy()
```

```
# Export to CSV
result_file = 'result.csv'
threshold_file = 'thresholds.csv'
R.to_csv(result_file, index=False)
threshold_df = pd.DataFrame({'Hour': T.keys(), 'Median': T.values()})
threshold_df.to_csv(threshold_file, index=False)
```

4. Code: Algorithm 9

Code 4: Algorithm 9

```
import pandas as pd
import numpy as np
from collections import defaultdict

# Step 1: Load the CSV data and calculate Hourly Median Thresholds
df = pd.read_csv('synthetic_individual_movements.csv')

# Extract hours from the 'Time' column
try:
    hours = df['Time'].str.extract('(\d+:\d+:\d+)')[0]
    df['Hour'] = pd.to_datetime(hours, format='%H:%M:%S').dt.hour
except Exception as e:
    print("Error:", e)
    exit()

# Calculate Hourly Median Thresholds
median_thresholds =
df.groupby('Hour')['SiteID'].value_counts().groupby('Hour').median().reset_index()
median_thresholds.columns = ['Hour', 'Median_Threshold']

# Save Hourly Median Thresholds to a CSV file
median_thresholds.to_csv('hourly_median_thresholds.csv', index=False)

# Step 2: Simulate data for subsequent steps (simplified example, replace with historical
data)
# Generate random group data and social tie strengths for illustration purposes (replace
with actual data)
n = len(df['IndividualID'].unique())
H = len(df['Hour'].unique())
group_data = defaultdict(list)
social_tie_strengths = defaultdict(float)

for h in range(0, H):
    groups = {}
    for i in range(1, n + 1):
        tower_id = df[(df['IndividualID'] == f'Individual {i}') & (df['Hour'] ==
h)]['SiteID'].values[0]
        if tower_id not in groups:
            groups[tower_id] = []
        groups[tower_id].append(i)

    for k, v in groups.items():
        group_data['Hour'].append(h)
        group_data['SiteID'].append(k)
        group_data['Individuals'].append(v)
        social_tie_strengths[(h, k)] = len(v)
```

```
group_dynamics = pd.DataFrame(group_data)

# Step 3: Simulate predictive models based on historical data (replace with actual
predictive models)
# Generate random predictive models for illustration purposes (replace with actual
models)
# Load historical data (replace with your actual historical data)
# historical_data = pd.read_csv('historical_data.csv')

# Define the number of future days to predict
n_days_to_predict = 7  # Example: Predict for the next 7 days

# Create an empty DataFrame to store predictions
future_predictions = pd.DataFrame(columns=['Hour', 'PredictiveModel'])

# Loop through the next n_days_to_predict days and make predictions
for day in range(1, n_days_to_predict + 1):
    # Replace this with your actual predictive model
    predictive_model = pd.DataFrame({'Hour': range(0, H), 'PredictiveModel':
np.random.rand(H)})

    # Append the predictions for the current day to the future_predictions DataFrame
    future_predictions = pd.concat([future_predictions, predictive_model],
ignore_index=True)

# Save future predictions to a CSV file
future_predictions.to_csv('future_predictions.csv', index=False)

# Step 4: Save Mobility Profiles (as before)

# Create mobility profiles based on the simplified data
mobility_profiles = df.groupby('IndividualID')['SiteID'].apply(list).reset_index()
mobility_profiles.columns = ['IndividualID', 'MobilityProfile']
mobility_profiles.to_csv('mobility_profiles.csv', index=False)

print("Execution completed. Results saved to CSV files.")
```

5. Code: Algorithm 10

Code 5: Algorithm 10

```
import pandas as pd
from datetime import datetime
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error

# Function to calculate similarity between individuals
def calculate_similarity(df):
    similarity_matrix = pd.DataFrame(index=df.index, columns=df.index, data=0)
    for i in df.index:
        for j in df.index:
            if i != j:
                similarity_matrix.loc[i, j] = sum(df.loc[i] == df.loc[j])
    return similarity_matrix

# Load the dataset
file_path = 'path_to_your_file.csv'  # Update this with your file path
data = pd.read_csv(file_path)

# Convert 'Time' column to datetime and extract the hour
data['Time'] = pd.to_datetime(data['Time'], format='%H:%M:%S')
data['Hour'] = data['Time'].dt.hour

# Generate crowd count data
crowd_count_data = data.groupby(['SiteID',
'Hour']).size().reset_index(name='CrowdCount')

# Macroscopic Analysis
# Calculate hourly thresholds and perform quartile classification
hourly_thresholds =
crowd_count_data.groupby('Hour')['CrowdCount'].median().reset_index(name='HourlyT
hreshold')
crowd_count_data_merged = pd.merge(crowd_count_data, hourly_thresholds, on='Hour')
crowd_count_data_merged['Quartile'] =
pd.qcut(crowd_count_data_merged['CrowdCount'], 4, labels=['Q1', 'Q2', 'Q3', 'Q4'])

# Microscopic Analysis
# Individual mobility patterns
individual_mobility =
data.groupby('IndividualID')['SiteID'].nunique().reset_index(name='UniqueSitesVisited')

# Social ties inference
pivot_data = data.pivot_table(index='IndividualID', columns='Hour', values='SiteID',
aggfunc=lambda x: x)
social_ties = calculate_similarity(pivot_data)

# Predictive Modelling
```

```
# Preparing data for the model
X = crowd_count_data[['SiteID', 'Hour']]
X['SiteID'] = X['SiteID'].apply(lambda x: int(x.split()[1]))
y = crowd_count_data['CrowdCount']

# Splitting the data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# RandomForestRegressor model
model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

# Predicting and evaluating
y_pred = model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)

# Output results
print("Mean Squared Error of the model:", mse)

# Save results to CSV
crowd_count_data_merged.to_csv('crowd_count_data_with_quartiles.csv', index=False)
individual_mobility.to_csv('individual_mobility_patterns.csv', index=False)
social_ties.to_csv('social_ties_similarity_matrix.csv')
```

6. Code: Algorithm 11

Code 6: Algorithm 11

```
import pandas as pd
import numpy as np

def load_data(file_path):
    data = pd.read_csv(file_path)
    data['Hour'] = pd.to_datetime(data['Time']).dt.hour
    return data

def macroscopic_density_analysis(data):
    crowd_count_raw = data.groupby(['SiteID',
'Hour']).size().reset_index(name='RawCrowdCount')
    daily_thresholds =
crowd_count_raw.groupby(['Hour'])['RawCrowdCount'].median().reset_index(name='Dai
lyThreshold')
    merged_data = pd.merge(crowd_count_raw, daily_thresholds, on='Hour')
    quartiles = pd.qcut(merged_data['RawCrowdCount'], 4, labels=['Q1', 'Q2', 'Q3', 'Q4'])
    merged_data['Quartile'] = quartiles
    return merged_data

def cumulative_mobility_analysis(data):
    site_ids = data['SiteID'].unique()
    hours = data['Hour'].unique()
    mobility_matrix = pd.DataFrame(columns=['FromSite', 'ToSite', 'Hour',
'CrowdMovement'])
    for hour in hours:
        current_hour_data = data[data['Hour'] == hour]
        next_hour_data = data[data['Hour'] == (hour + 1) % 24]
        for from_site in site_ids:
            for to_site in site_ids:
                if from_site != to_site:
                    moved_individuals = len(set(current_hour_data[current_hour_data['SiteID']
== from_site]['IndividualID']) &
                                            set(next_hour_data[next_hour_data['SiteID'] ==
to_site]['IndividualID']))
                    mobility_matrix = mobility_matrix.append({'FromSite': from_site, 'ToSite':
to_site, 'Hour': hour, 'CrowdMovement': moved_individuals}, ignore_index=True)
    density_estimation = mobility_matrix.groupby(['ToSite',
'Hour'])['CrowdMovement'].sum().reset_index(name='Incoming')
    density_estimation = pd.merge(density_estimation,
mobility_matrix.groupby(['FromSite',
'Hour'])['CrowdMovement'].sum().reset_index(name='Outgoing'), left_on=['ToSite',
'Hour'], right_on=['FromSite', 'Hour'])
    density_estimation['DensityEstimation'] = density_estimation['Incoming'] -
density_estimation['Outgoing']
    density_estimation = density_estimation[['ToSite', 'Hour', 'DensityEstimation']]
    return density_estimation
```

```
def save_results(density_analysis, mobility_analysis, density_analysis_file,
mobility_analysis_file):
    density_analysis.to_csv(density_analysis_file, index=False)
    mobility_analysis.to_csv(mobility_analysis_file, index=False)

# Paths for data file and output files
file_path = 'Input_synthetic_individual_movements.csv'
density_analysis_file = 'macroscopic_crowd_density_analysis.csv'
mobility_analysis_file = 'cumulative_crowd_mobility_analysis.csv'

# Load and process data
data = load_data(file_path)
density_analysis = macroscopic_density_analysis(data)
mobility_analysis = cumulative_mobility_analysis(data)

# Save results to CSV files
save_results(density_analysis, mobility_analysis, density_analysis_file,
mobility_analysis_file)
```

7. **Code: Algorithm 12**

Code 7: Algorithm 12

```
import pandas as pd
from itertools import combinations

# Load the dataset
file_path = 'path_to_your_file.csv'  # Replace with your file path
data = pd.read_csv(file_path)

# Macroscopic Crowd Density Analysis
# Compute Raw Crowd Count
crowd_count_raw = data.groupby(['SiteID',
'Time']).size().reset_index(name='CrowdCount')

# Calculate Daily Threshold (Method of M-o-Ments)
daily_threshold = crowd_count_raw.groupby('SiteID')['CrowdCount'].agg(['mean',
'std']).reset_index()
daily_threshold['Threshold'] = daily_threshold['mean'] + daily_threshold['std']

# Classify crowd density into quartiles
quartiles = crowd_count_raw['CrowdCount'].quantile([0.25, 0.5, 0.75]).to_dict()
def classify_quartiles(count):
    if count <= quartiles[0.25]:
        return 'Q1'
    elif count <= quartiles[0.5]:
        return 'Q2'
    elif count <= quartiles[0.75]:
        return 'Q3'
    else:
        return 'Q4'
crowd_count_raw['Quartile'] =
crowd_count_raw['CrowdCount'].apply(classify_quartiles)

# Cumulative Crowd Mobility Analysis (Adjusted)
# Note: The original method is adjusted due to data limitations
# We use the raw crowd count as an estimate for crowd density

# Microscopic Individual Mobility and Social Interaction Analysis
# Construct a Social Interaction Matrix
social_interaction_data = data.groupby(['Time',
'SiteID'])['IndividualID'].apply(list).reset_index()
social_interaction_data['Interactions'] =
social_interaction_data['IndividualID'].apply(lambda x: list(combinations(x, 2)))
exploded_social_interaction_data = social_interaction_data.explode('Interactions')
exploded_social_interaction_data =
exploded_social_interaction_data.dropna(subset=['Interactions'])

# Save results to CSV files
crowd_count_raw.to_csv('crowd_count_raw.csv', index=False)
```

```
daily_threshold.to_csv('daily_threshold.csv', index=False)
exploded_social_interaction_data.to_csv('social_interaction_data.csv', index=False)

print("Data analysis complete. Files saved.")
```

8. Code: Algorithm 13

Code 8: Algorithm 13

```
import pandas as pd
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error
import numpy as np

# Load the dataset
file_path = 'path_to_your_file.csv'  # Replace with your file path
data = pd.read_csv(file_path)

# Process the data for modelling
# Assume 'SiteID' and 'Time' are relevant columns
data['Hour'] = pd.to_datetime(data['Time']).dt.hour
data['SiteID_Encoded'] = data['SiteID'].astype('category').cat.codes

# Compute Raw Crowd Count
crowd_count_raw = data.groupby(['SiteID_Encoded',
'Hour']).size().reset_index(name='CrowdCount')

# Features and target variable
X = crowd_count_raw[['Hour', 'SiteID_Encoded']]
y = crowd_count_raw['CrowdCount']

# Splitting the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# Building the RandomForestRegressor model
model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

# Making predictions on the test set and evaluating the model
y_pred = model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
rmse = np.sqrt(mse)

# Predicting future crowd density
# Assuming the 'next hour' is the hour following the last hour in the dataset
next_hour = data['Hour'].max() + 1
future_predictions = pd.DataFrame({'Hour': next_hour, 'SiteID_Encoded':
data['SiteID_Encoded'].unique()})
future_predictions['PredictedCrowdCount'] = model.predict(future_predictions)

# Mapping encoded SiteID back to original SiteID
site_id_mapping = data[['SiteID', 'SiteID_Encoded']].drop_duplicates()
future_predictions = future_predictions.merge(site_id_mapping, on='SiteID_Encoded')
future_predictions = future_predictions[['Hour', 'SiteID', 'PredictedCrowdCount']]
```

```
# Save the future predictions to a CSV file
future_predictions.to_csv('future_predictions.csv', index=False)

print(f"Model RMSE: {rmse}")
print("Future predictions saved to 'future_predictions.csv'")
```

9. Code: Algorithm 14

Code 9: Algorithm 14

```
import pandas as pd

# Load the dataset
file_path = 'Input_Synthetic.csv'  # Replace with your actual file path
df = pd.read_csv(file_path)

# Extract the hour from the Time field
df['Hour'] = pd.to_datetime(df['Time']).dt.hour

# Group by the 'Hour' field and compute medians
grouped = df.groupby('Hour')
medians = grouped['Current_Count'].median()

# Compute the Median-of-Medians (M-o-M) Threshold
M-o-M_threshold = medians.median()

# Categorization Based on M-o-M Threshold
df['Category_Threshold'] = df['Current_Count'].apply(lambda x: 'Above Threshold' if x >
M-o-M_threshold else 'Below Threshold')

# Quartile Classification
Q1, Q2, Q3 = df['Current_Count'].quantile([0.25, 0.5, 0.75])
def classify_quartile(value):
    if value <= Q1:
        return 'Q1'
    elif value <= Q2:
        return 'Q2'
    elif value <= Q3:
        return 'Q3'
    else:
        return 'Q4'
df['Quartile'] = df['Current_Count'].apply(classify_quartile)

# Prediction Model
# Define weights (assuming equal weights for simplicity)
w_act = 1
w_hist = 1
w_thresh = 1
def prediction_model(current_count, ref_count, threshold, M-o-M_threshold):
    f_thresh = 1 if current_count > M-o-M_threshold else 0
    return w_act * current_count + w_hist * ref_count + w_thresh * f_thresh
df['Prediction'] = df.apply(lambda x: prediction_model(x['Current_Count'],
x['Ref_Count'], x['Category_Threshold'], M-o-M_threshold), axis=1)

# Resultant dataset R and hourly medians T
R = df.copy()
T = medians.reset_index()
```

```
T.columns = ['Hour', 'Median_Current_Count']

# File paths for the CSV exports
resultant_file_path = 'Resultant_Dataset.csv'  # Replace with your desired file path
medians_file_path = 'Hourly_Medians.csv'      # Replace with your desired file path

# Export to CSV
R.to_csv(resultant_file_path, index=False)
T.to_csv(medians_file_path, index=False)
```

10. Code: Algorithm 15

Code 10: Algo 15 Crowd Density

```
import pandas as pd

# Load the data
simulated_data = pd.read_csv('simulated_data.csv')

# Crowd Density Estimation
simulated_data['TotalActivity'] = simulated_data[['WiFi', 'Vehicular', 'RideSharing', 'Tweets']].sum(axis=1)
crowd_density = simulated_data.groupby('Location')['TotalActivity'].median().reset_index()
crowd_density.rename(columns={'TotalActivity': 'CrowdDensity'}, inplace=True)
quartiles = pd.qcut(crowd_density['CrowdDensity'], 4, labels=['Q1', 'Q2', 'Q3', 'Q4'])
crowd_density['Quartile'] = quartiles
crowd_density.to_csv('crowd_density.csv', index=False)
```

Code 11: Algorithm 15 Mobility Density

```
import pandas as pd
import random

def simulate_mobility_flow(num_locations, num_timepoints):
    flow_data = []
    for t in range(num_timepoints):
        for i in range(1, num_locations + 1):
            for j in range(1, num_locations + 1):
                if i != j:
                    flow_value = random.randint(0, 100)  # Simulating flow value
                    flow_data.append([i, j, t, flow_value])
    columns = ['FromLocation', 'ToLocation', 'Timepoint', 'FlowValue']
    return pd.DataFrame(flow_data, columns=columns)

# Parameters
num_locations = 5
num_timepoints = 24

# Generate flow matrix
flow_matrix = simulate_mobility_flow(num_locations, num_timepoints)
flow_matrix.to_csv('flow_matrix.csv', index=False)
```

Code 12: Algorithm Social Media Interaction

```
import pandas as pd

# Load the data
simulated_data = pd.read_csv('simulated_data.csv')

# Interaction Matrix
interaction_matrix = simulated_data.groupby('Location')[['RideSharing',
'Tweets']].sum().reset_index()
interaction_matrix.rename(columns={'RideSharing': 'RideSharingInteractions', 'Tweets':
'TweetInteractions'}, inplace=True)

# Sentiment Matrix
sentiment_matrix = simulated_data.groupby('Location')['Sentiment'].mean().reset_index()
sentiment_matrix.rename(columns={'Sentiment': 'AverageSentiment'}, inplace=True)

# Save to CSV
interaction_matrix.to_csv('interaction_matrix.csv', index=False)
sentiment_matrix.to_csv('sentiment_matrix.csv', index=False)
```