smart cities

Article

LEAF: A Lifestyle Approximation Framework Based on Analysis of Mobile Network Data in Smart Cities

Somaye Moghari [1,*], Mohammad K. Fallah [2], Saeid Gorgin [2] and Seokjoo Shin [2,*]

1 Faculty of Mathematical Sciences, Shahrood University of Technology, Shahrood 3619995161, Iran
2 Department of Computer Engineering, Chosun University, Gwangju 61453, Republic of Korea; mkfallah@chosun.ac.kr (M.K.F.); gorgin@chosun.ac.kr (S.G.)
* Correspondence: s.moghari@shahroodut.ac.ir (S.M.); sjshin@chosun.ac.kr (S.S.)

Highlights:

What are the main findings?

- The proposed framework offers a robust and accurate method for modeling and understanding urban residents' lifestyles by analyzing anonymized mobile network data.
- We have defined a set of analytical patterns designed to extract key insights and valuable knowledge from the input data. By fusing this information with ontological data and geographical maps, our approach uncovers significant and straightforward perspectives that enhance understanding of urban lifestyles.

What is the implication of the main finding?

- The framework allows for identifying lifestyle patterns and epidemic behaviors enabling more informed decision-making and strategic planning, ensuring that services and infrastructure can be tailored to meet the specific needs of different urban areas.
- This approach provides valuable insights into daily routines, preferences, and behaviors, which can be crucial for urban planners, policymakers, and businesses.

Abstract: The increasing use of mobile networks is an opportunity to collect and model users' movement data for extracting knowledge about life and health while considering privacy leakage risk. This study aims to approximate the lifestyles of urban residents, employing statistical information derived from their movements among various Points of Interest (PoI). Our investigations comprehend a multidimensional analysis of key urban factors to provide insights into the population's daily routines, preferences, and characteristics. To this end, we developed a framework called LEAF that models lifestyles by interpreting anonymized cell phone mobility data and integrating it with information from other sources, such as geographical layers of land use and sets of PoI. LEAF presents the information in a vector space model capable of responding to spatial queries about lifestyle. We also developed a consolidated lifestyle pattern framework to systematically identify and analyze the dominant activity patterns in different urban areas. To evaluate the effectiveness of the proposed framework, we tested it on movement data from individuals in a medium-sized city and compared the results with information collected through surveys. The RMSE of 5.167 between the proposed framework's results and survey-based data indicates that the framework provides a reliable estimation of lifestyle patterns across diverse urban areas. Additionally, summarized patterns of criteria ordering were created, offering a concise and intuitive representation of lifestyles. The analysis revealed high consistency between the two methods in the derived patterns, underscoring the framework's robustness and accuracy in modeling urban lifestyle dynamics.

Keywords: smart city; lifestyle approximation; mobility data analysis; vector space model

Citation: Moghari, S.; Fallah, M.K.; Gorgin, S.; Shin, S. LEAF: A Lifestyle Approximation Framework Based on Analysis of Mobile Network Data in Smart Cities. *Smart Cities* **2024**, *7*, 3315–3333. https://doi.org/10.3390/smartcities7060128

Academic Editor: Pierluigi Siano

Received: 30 July 2024
Revised: 11 October 2024
Accepted: 29 October 2024
Published: 2 November 2024

1. Introduction

Lifestyle is the distinctive and recognizable mode of living and consists of observable and deducible expressive behaviors [1]. Understanding the lifestyle of individuals in a specific city area is motivated by various factors. For instance, selecting a residence in a community with a compatible lifestyle is crucial for personal satisfaction and social harmony [2,3]. Businesses benefit from this understanding by strategically placing product or service centers to meet the needs and preferences of the local population [4]. Additionally, location-based advertising becomes more effective when tailored to the lifestyle patterns of the target audience [5,6]. Moreover, urban planning and development can be optimized by facilitating conscious facility relocation, ensuring that public services and infrastructure are accessible and aligned with residents' behaviors, income levels, and preferences [7,8]. Another critical motivation is the modeling and analysis of epidemics; understanding movement patterns and lifestyle behaviors aids in predicting and controlling the spread of infectious diseases [9,10]. These motivations collectively highlight the importance of modeling and analyzing urban residents' lifestyles to enhance city life and public health.

Human mobility data are a precious source of information about lifestyle [11,12]. It is shaped by structural patterns influenced by geography and social norms [13,14] and facilitates the extraction of contextual information and analysis of the periodic nature of visitation behaviors [15]. This information includes the footprint of individual or group lifestyles, incorporating a contextual framework linked to real-world locations. In addition, an anthology of the associated places facilitates lifestyle structure analysis according to visitation frequencies. On the other hand, utilizing cell phone location data to analyze urban-scale and larger-scale human mobility patterns is a preferred approach [16–18]. Identifying the location of a cell phone involves various technologies including GPS, WiFi, IP address, and cell triangulation [19–21]. Privacy considerations precede the granularity of the data source, a factor directly impacting accuracy. Utilizing anonymized data from network operators is a viable method to acquire location data, and multiple algorithms have been developed for cellular location calculation using such data [22–24].

In this investigation, we utilized data from BTS antennas to analyze the anonymous tracking of mobile phones. Each device was assigned a unique key, accompanied by a collection of points, each associated with specific start and end times. We also developed a framework for indexing and interpreting cell phone tracking data to provide an API for inquiring about some dimensions of urban areas' lifestyles. This framework aims to deliver a location-based lifestyle search service, applying artificial intelligence techniques to spatial data derived from individuals' cell phone movements. The presentation includes an abstraction of lifestyle, elucidating the data flows and components constituting the approximation framework. Methodologically, we delve into the processing, indexing, retrieval, query evaluation, and inferencing of lifestyle for specific areas. The proposed framework underwent evaluation using real data from a study area, demonstrating close alignment between patterns obtained through the framework and those observed in the collected information.

The remainder of this paper is structured as follows. Section 2 provides background information and discusses related research on cell phone data collection and related architectures. Section 3 details the proposed framework. Section 4 presents experimental results from a mid-sized city as a case study. Finally, Section 5 concludes the research.

2. Background and Related Work

Mobile phone data can be categorized into event-driven and network-driven types [25]. Event-driven data captures details of interactions like phone calls, text messages, or internet access. Network-driven data are aggregated at the cell tower level and include passive location updates from the Base Transceiver Station (BTS), reflecting regular intervals, device power changes, signal reception, and connection-type changes. Figure 1 illustrates a schematic of user trajectory data collection facilitated by BTS antennas using mobile phone signals. It displays a mobile phone in seven distinct positions, capturing data on coordinates, time, and the device's assigned ID for each position. The user's presence at a specific location can be determined by analyzing the tracking data and overlaying them with geographic information. Also, the probabilistic approaches process mobility data and estimate devices' mobility from cellular data [26], and techniques such as spatial interpolation and AI-based methods enhance data quality for further processing [27,28].

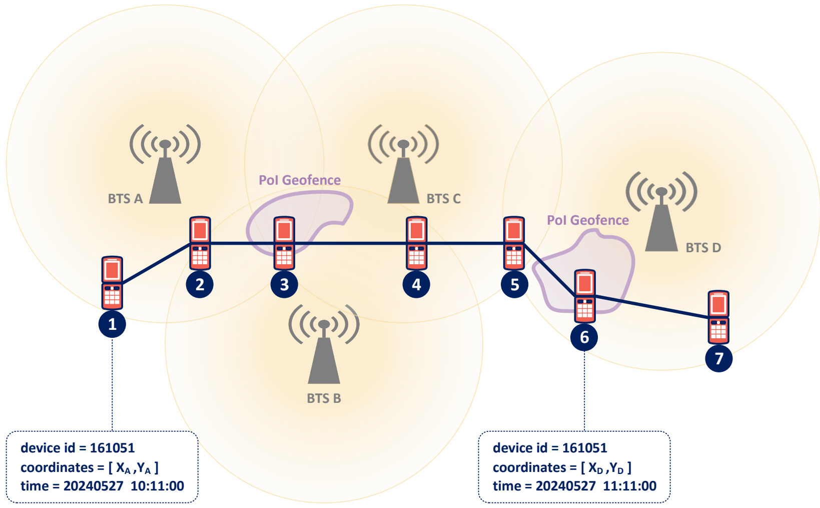


Figure 1. Collection of mobile phone signaling data and user trajectories by BTS.

Mobile phone network data have been used for various purposes, including estimating population distribution, identifying activities in different city areas, determining mobility patterns, analyzing local events, and examining the geography of social networks [29,30]. It can describe how people organize their visitation patterns and mobility around the city, uncovering that lifestyles are a continuum spectrum of the relative balance between work, shopping, transportation, or leisure time [31]. Specifically, by integrating with other data sources such as remote sensing imagery, PoI, or road networks, it enables the extraction of new insights, including land use patterns [32,33], traffic monitoring [34], analyzing the influence of environmental factors on mobile-phone-distracted driving [35], optimizing parks and promoting human well-being [36], and mitigating negative ecological effects to improve transit-oriented development in urban areas [37]. Analyzing mobile phone mobility data in healthcare is crucial in tracking the spread of infectious diseases [38], monitoring public health trends [39], optimizing healthcare resource allocation [40], and improving emergency response times [41]. Also, it aids in understanding population demographics and mobility changes during outbreaks, which can inform public health interventions and policy decisions, ultimately enhancing the effectiveness of healthcare systems [42,43]. While these studies provide valuable insights into mobility patterns, land use, and public health trends, our proposed framework builds upon these achievements by introducing a more refined approach to lifestyle pattern recognition. Specifically, we employ BTS-based mobility data and geospatial analysis to capture a more granular view of urban lifestyles. This research integrates the strengths of previous methods, such as spatial-temporal patterns and PoI visiting analysis. Still, it extends them by incorporating the duration of stay at specific locations and multi-dimensional lifestyle indicators.

Lifestyle modeling and analysis according to PoI visiting capture the collective effects of population activities and the restoration of infrastructure and business services [44]. Government agencies and planning departments can leverage these insights to ensure sustainable development and enhance the quality of life [45]. Figure 2 provides an overview of the techniques for processing and analyzing human mobility data. The raw data consist of multiple sources, including various datasets related to human mobility and the geofence of PoIs. Data analysis involves merging these diverse data sources to track visits at POIs and employing comprehensive methods for clustering the data. The metrics, which are the final outputs of these methods, include calculating recovery trajectories and other relevant indicators to assess the restoration and dynamics of population activities. Of course, different methods have been developed for characterizing lifestyle signatures, each contributing unique insights into understanding human behavior and mobility patterns. The framework proposed in [12] identifies lifestyle signatures in four steps. First, visits between PoIs are detected. Second, a network of places is generated, where nodes represent PoIs and links represent visits, with weights indicating the number of visits. Third, human visitation motifs are constructed, and fourth, these motifs are characterized by calculating distances and counting quantities for uncovering spatiotemporal disparities. Also, in LTP-Net [46], the spatial-temporal-pattern dimensions of human mobility are combined to provide more comprehensive information for individual identification. It analyzes the spatial, temporal, and pattern dimensions according to the trajectory heat map, location–time pattern sequence, and some profile features extracted from the trajectory. Another graph-based mobility profiling is also presented in [47]. They utilized this graph representation to introduce a workflow for identifying groups of individuals based on their mobility behavior. Their method contributes to a deeper understanding of lifestyle signatures and helps refine lifestyle modeling techniques by categorizing individuals through movement patterns. While previous approaches, such as network generation and spatial-temporal analysis, provide valuable techniques for identifying lifestyle signatures, our proposed framework significantly extends these methods by introducing new analytical models specifically designed to capture the dynamic nature of urban residents' lifestyles. Moreover, by introducing novel analytical methods, our framework surpasses existing models by capturing the sequence of visits to PoIs by urban subdivision residents in addition to the spatial and temporal nuances of their mobility. This enables more informed decision-making, optimizing public services, resource allocation, and even healthcare interventions based on the dynamic lifestyle patterns of city residents.

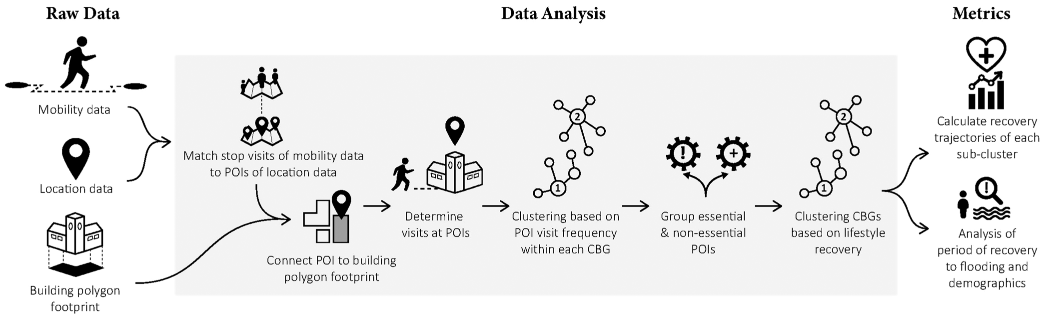


Figure 2. Specifying population visitation to POIs to characterize lifestyle patterns [44].

3. Proposed Framework

In this section, we present the proposed **Lif**e**s**tyle **A**pproximation **F**ramework (LEAF). Figure 3 illustrates the components of LEAF. We discuss these components in four categories: data providers, data storage, processing nodes, and data flows.

3.1. Data Providers

These components are the data gateways to LEAF fetching the required information from various sources and transferring them to the storage. The main functionalities of these providers are data integration, management, access, and retrieval. Integrating data with other datasets within the system provides enriched information and ensures compatibility and interoperability of data from different sources.

Mobility data collector. The mobility data collector is critical for gathering and processing anonymous cell phone tracking data. This component is responsible for understanding and analyzing human mobility patterns by capturing detailed and continuous location information from mobile devices. It ensures that all collected data are anonymized to protect user privacy.

Geodata provider. This component is designed within the framework to supply and manage geographic information essential for spatial analysis. Its primary function is to act as a centralized source of all geographical data, ensuring that other system components have seamless access to location-based information. This includes providing data such as the coordinates of PoIs, land use classifications, and spatial relationships like distances or adjacency between different geographical entities. The Geodata provider is responsible for retrieving raw geographic data and organizing, processing, and delivering it in a format that components like the Refiner, Indexer, or Query engine can immediately utilize. This allows other parts of the system to incorporate spatial context into their analyses, such as calculating proximity to landmarks or overlaying mobility data with geographic boundaries.

Ontology provider. An ontology is a formal, explicit specification of a shared conceptualization. It provides a structured framework to categorize and describe the properties and relationships of concepts within a specific domain [48]. The ontology provider is a core component of the proposed framework and is responsible for supplying, managing, and facilitating access to ontological data. It acts as a centralized repository for domain-specific knowledge, ensuring that different components within the system can consistently interpret and process these data. The ontology defines the concepts, relationships, and hierarchies relevant to the urban environment, such as types of locations, PoIs, and their connections to various lifestyle criteria. It allows the seamless integration of geospatial data with lifestyle criteria by ensuring that the system components interpret the information consistently. The ontology provider supports querying for domain-specific data and reasoning about the relationships between different location types. This allows the system to deduce patterns like a user's likelihood of visiting multiple PoIs in a category and infer broader lifestyle patterns. Also, as urban environments and data sources evolve, the ontology provider can be updated to reflect new concepts or relationship changes, ensuring that the framework remains adaptable to different cities or regions.

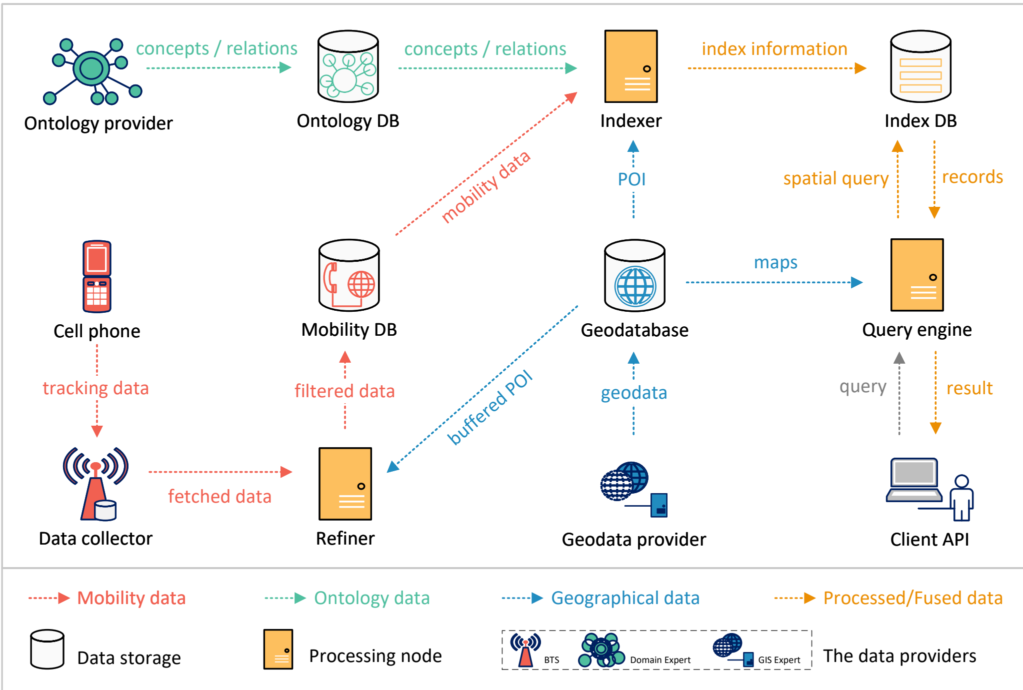


Figure 3. The proposed architecture of LEAF.

3.2. Data Storages

The proposed architecture consists of four storages for the mobility data, ontology, maps and PoIs, and indexing data provided by processing nodes or data providers.

Mobility database. This database consists of two main tables. The first table includes columns for the tracking ID, point coordinates, and the timestamp of arrival at each location. The second table, derived from the information in the first table, contains columns for the tracking ID, representative point coordinates (latitude and longitude), the timestamp of arrival, and the duration of presence at that location. The coordinates are captured at regular intervals to create a comprehensive movement trajectory. In the second table, several closely located points that occur in sequence are aggregated into a single point, allowing for the calculation of the duration spent at these locations.

Ontology database. This component serves as a centralized repository dedicated to storing structured domain-specific knowledge that is critical for interpreting and enriching the data within the framework. Unlike other data components, which primarily handle raw or processed spatial and mobility data, the ontology database is responsible for managing high-level conceptual information. This information includes location types (such as universities, hospitals, or factories) as well as higher-level categories that group multiple location types under each category. Additionally, the relationships between different location types and PoI are also stored in this database. The database ensures that this knowledge is readily accessible for the system's components, such as the Indexer, enabling them to map raw mobility and geographical data to relevant ontological categories. By doing so, it allows the framework to move beyond simple spatial analysis, providing context that helps in recognizing patterns in user behavior, activities, and urban lifestyles. The data stored in the ontology database also supports the framework's ability to infer higher-level insights, such as which categories of locations are frequently visited and how different activities are distributed across urban spaces.

Geodatabase. The geodatabase includes spatial information, such as maps, geographic features, and geofencing data. The geofences define virtual boundaries around specific areas of interest, ensuring that only mobility data within these boundaries are processed.

Index database. It is a specialized data repository within the LEAF architecture designed to store and manage indexed representations of individual mobility patterns. This database plays a critical role in capturing and organizing the lifestyle indices of individuals based on their frequency of visiting PoIs. The lifestyle index is structured based on the vector space model, where each dimension represents a PoI type. Here, a record consists of the ID and a weighted array, which serves as the individual's lifestyle index.

3.3. Processing Nodes

The LEAF architecture includes three key processing nodes: the Refiner, the Indexer, and the Query Engine. Each node plays a distinct role in ensuring the effective processing, management, and retrieval of mobility data.

Refiner. This component removes the irrelevant mobility data points. It applies a dynamic geospatial buffer around the PoIs, utilizing geofencing data and other geographic datasets. Geofencing involves defining a virtual boundary around a physical location using GPS, Wi-Fi, or cellular data. For each PoI, a specific buffer radius is determined based on the nature of the location, which could be influenced by local infrastructure, land use, or the expected range of influence of the PoI. Only data points within the specified buffered zones around PoIs are considered relevant here.

Indexer. The main role of the indexer is to present each person based on the frequency of visiting PoIs. Index information records include the ID and a weighted array presenting the individual's lifestyle index. The lifestyle is indexed based on the vector space model [48], where the determination of weights is based on fuzzy logic. In lifestyle vectors, cells correspond to PoI types. Also, each PoI with type p is associated with a fuzzy membership function μ_p. Let μ_p be the frequency of p and w_p be the value (weight) assigned to the corresponding cell in the lifestyle array. Then, $w_p = \mu_p(f_p)$. For example, let the set of PoI types include **g**ym, **r**estaurant and **u**niversity, with fuzzy membership functions $\mu_g(x) = LF(x; 0, 10)$, $\mu_r(x) = LF(x; 0, 4)$ and $\mu_u(x) = LF(x; 0, 15)$, where:

$$LF(x; a, b) = \begin{cases} 0, & x < a \\ \frac{b-x}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases}$$

Now, the index vector corresponding to frequency vector $F = [10, 1, 0]$ is $I = [1, 0.25, 0]$, and its normalized version is $\hat{I} = [0.8, 0.2, 0]$.

Query engine. This module retrieves records from the database for the specified target area upon receiving a request through the Application Programming Interface (API), delivering the results to the user. The input query may entail a polygon within the target area or a singular point. In the latter scenario, the target area is determined by intersecting the point with the spatial layer of urban subdivisions. Subsequently, records within this polygon are extracted, and the area's lifestyle is computed following Algorithm 1. The algorithm takes indices, criteria, and the relationship between criteria and place types as input. It then extracts the weights associated with each criterion's location types, clusters them into low, moderate, and high categories, and calculates the frequency of weights in each cluster. The lifestyle information encompasses the area's sample count and the percentages of low, moderate, and high populations for each criterion.

Algorithm 1: Calculating the lifestyle indicators for an urban area.

input : I , // the set of retrieved indices
C , // the set of criteria
R // the relation between criteria and place types

for *each criterion* $c \in C$ **do**
- $W \leftarrow \{w_p^i \mid i \in I, (c, p) \in R\}$ // the associated weights of place types to criterion c
- $LB \leftarrow \{0.2, 0.3, 0.4\}$ // the set of candidate lower bounds
- $UB \leftarrow \{0.6, 0.7, 0.8\}$ // the set of candidate upper bounds
- **for** *each* $b \in LB \cup UB$ **do**
  - $n_b \leftarrow |\ \{w \in W \mid w \in [b - 0.05, b + 0.05]\}\ |$
- $lb \leftarrow \bigvee \{b \in LB \mid n_b = \bigwedge_{b' \in LB} n_{b'}\}$ // the selected lower bound for clustering
- $ub \leftarrow \bigwedge \{b \in UB \mid n_b = \bigwedge_{b' \in UB} n_{b'}\}$ // the selected upper bound for clustering
- $L_c \leftarrow \left| \{W \cap [0, lb)\} \right|$ // the number of individuals in the low cluster
- $M_c \leftarrow \left| \{W \cap [lb, ub]\} \right|$ // the number of individuals in the moderate cluster
- $H_c \leftarrow \left| \{W \cap (ub, 1]\} \right|$ // the number of individuals in the high cluster

return $\vec{L}, \vec{M}, \vec{H}$

3.4. Data Flows

This section explores the four main types of data flows in the proposed framework: mobility data flow, ontology data flow, geographical data flow, and processed/fused data flow. Each one flows between specific components in the system to provide comprehensive analysis and decision-making processes.

Mobility data flow. Mobility data represents user movement information, collected from mobile sources and transmitted from the mobility data collector to the refiner. In the refiner, the raw data are processed to remove irrelevant data and noise. The refined data are then stored in the mobility DB and passed on to the indexer to generate indices that facilitate analysis.

Ontology data flow. Ontology data are generated by the ontology provider, responsible for supplying structured domain-specific knowledge such as categorizations of locations, activities, or relationships relevant to the framework. Once generated, these data are transmitted to the ontology DB, where it is securely stored and maintained for efficient access. The ontology DB serves as a centralized repository that manages the storage and retrieval of ontological knowledge. From the ontology DB, the data are forwarded to the indexer, where it is integrated with other data types like mobility and geographical data. In the indexer, the ontology data are used to contextualize and enrich the mobility data by associating locations and activities with their corresponding ontological categories. This allows the system to create more meaningful knowledge and insights, as it can interpret users' movements geographically, considering the meaning and multi-level interpretation of their activities and roles in urban lifestyles.

Geographical data flow. Geographical data are provided by the geodata provider and stored in the geodatabase. From there, it is distributed to several components: buffered PoI and geofence are sent to the refiner, PoI goes to the indexer, and maps are sent to the query engine. This type of data encompasses geographical features, PoIs, and area boundaries. Geographical data support spatial analysis within the framework. It helps define the physical environment in which mobility occurs, which is crucial for understanding how users interact with different locations. By integrating maps and geofences, the framework can refine user movements concerning these geographical elements.

Processed/Fused data flow. Processed or fused data are generated within the indexer, where multiple sources of information, such as mobility data, geographical data, and ontology data are combined to create a unified and enriched dataset. This fusion process involves associating users' movement patterns with geographical locations and categorizing them based on predefined ontological frameworks. Once the data are processed and fused in the indexer, it flows to the index DB, where it is securely stored in a structured format optimized for fast retrieval. The Index DB acts as the core repository for this processed data, ensuring that it is readily available for complex queries. The stored data in the index DB contains valuable insights, such as categorized locations, behavioral patterns, and lifestyle signatures, making it a key resource for various types of analysis. From the index DB, the processed data are forwarded to the query engine. The query engine is responsible for handling requests from external applications or services, typically through the client API. When a user or system sends a query via the client API, the query engine interprets the request and retrieves the relevant processed data from the index DB. The query engine then processes these data, performing any necessary filtering, aggregation, or analysis, and returns the most accurate and relevant information in response to the user's query.

4. Case Study

Our study area is Shahrood; a city and capital of Shahrood City, Semnan province, Iran. This city is located at latitude 36°25′ N and longitude 55°01′ E. It has a population of about 150,000 and an area of 52 km^2. An overview of Shahrood is shown in Figure 4, where areas N1,..., and N8 are selected randomly from the urban subdivision map. The city is equipped with base transceiver stations, allowing us to index the mobility data for December 2019 for 4000 cell phones in areas N1 through N8. We have calculated the lifestyle parameters for these eight areas according to Algorithm 1 based on the 4000 indexed mobility datum, five criteria, and 13 related places illustrated in Figure 5.

4.1. Materials and Methods

The proposed framework is implemented in Java to evaluate it with real data. In addition, we employed a development stack consisting of PostgreSQL for managing geospatial data, and PostGIS for spatial queries. The combination of these technologies supports seamless integration with various GIS layers and provides a flexible environment for processing mobility data. Also, we compared the output provided by the framework with the information obtained from statistical analysis of field survey data. Figure 6 shows an overview of the data and the procedures performed in this study. The proposed LEAF framework leverages cell phone mobility data and various GIS layers, including the study areas, land use layer, PoI layer, and the set of criteria illustrated in Figure 5. Each criterion corresponds to specific location types, with some location types being common across multiple criteria.

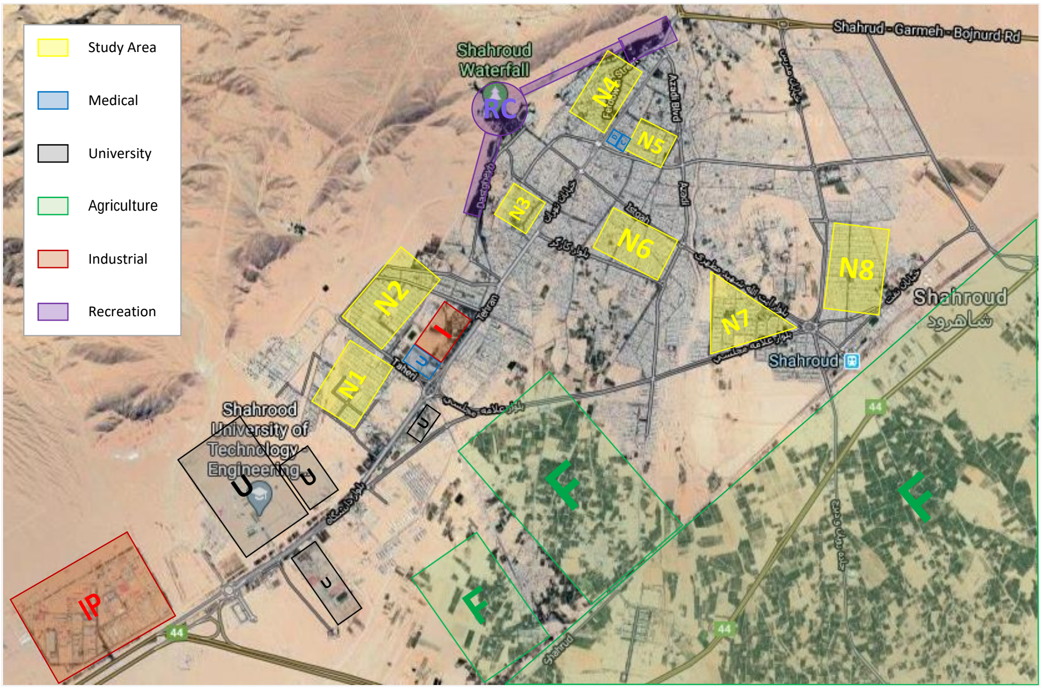


Figure 4. An overview map of Shahrood.

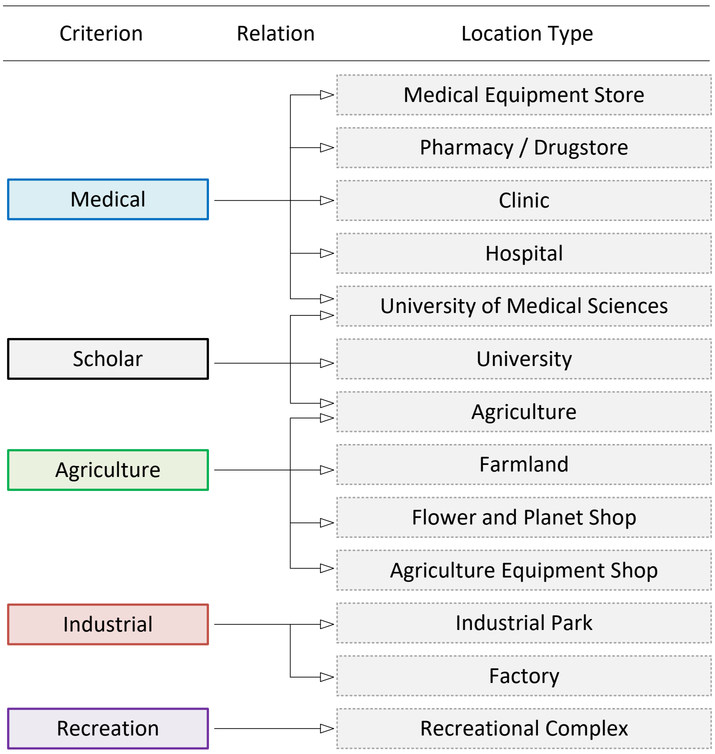


Figure 5. Criteria and location types used in the case study and their interrelations.

In addition to approximating lifestyle parameters for the candidate urban subdivisions by LEAF, we collected the field data corresponding to the list of test criteria for comparison. Therefore, we designed a tablet app with a slider of 0 to 100 for each criterion and then used it to collect the data for 8000 random people in the eight areas (about 1000 people in each area). The app was designed to collect data on participants' presence in 13 different types of locations listed in Figure 5. Users were prompted to report their level of presence at each location by interacting with a slider, which allowed them to rate their presence on a scale of 0 to 100. To ensure optimal user engagement and accurate data reporting, the app's interface was carefully designed with usability in mind. The user interface is minimalist, focusing on clarity and ease of interaction, with a simple slider for each location type. The design ensures that participants can quickly input data without confusion, enhancing the user experience by maintaining consistency and intuitive flow throughout the app. The color scheme is neutral, with visual cues to guide users through the data submission process. The data collected includes self-reported scores from participants, indicating their perceived level of presence in various locations.

After data collection, the raw scores were processed to remove inconsistencies or outliers, ensuring data integrity. Next, statistical outlier detection techniques, such as the interquartile range method, were applied to flag and exclude scores that deviated significantly from the median values. This was especially important for locations where typical presence scores followed a certain pattern. Finally, redundancy checks were conducted to ensure that duplicate entries from the same participant were either consolidated or removed, depending on the nature of the duplication. The processed data were then compared against the lifestyle parameters inferred by the proposed LEAF framework. Statistical analysis was performed to identify correlations between self-reported presence data and the framework's outputs, validating or refining the lifestyle estimates for different urban subdivisions. The output of both procedures (proposed framework and field survey analysis) for each study area is a normalized vector, each dimension of which corresponds to one of the input criteria. Next, we compared the vectors of each area obtained from two different procedures and analyzed the difference in weights obtained for the corresponding criteria to evaluate the lifestyle approximation.

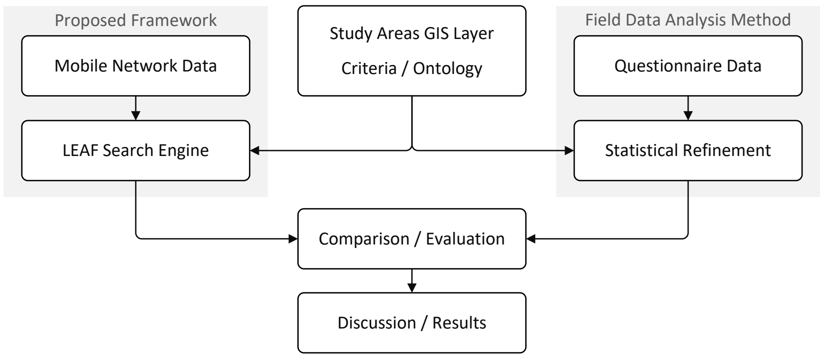
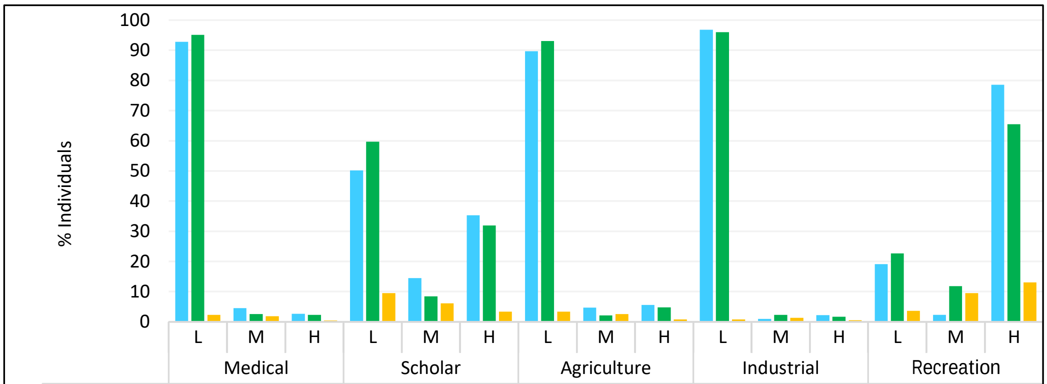


Figure 6. The data and procedures employed in this study.

4.2. Results and Discussion

In our study, we categorized each criterion Medical (M), Scholar (S), Agriculture (A), Industrial (I), and Recreation (R) into three levels: High (H), Balanced (B), and Low (L). These levels represent the frequency of visits to PoIs related to each criterion, with "High" indicating frequent visits, "Balanced" indicating moderate visits, and "Low" indicating infrequent visits. This classification allows us to model and understand lifestyle patterns within different urban subdivisions, as well as across the entire city. For each urban area (e.g., N1 to N8), we identified the dominant patterns by analyzing the order of the criteria based on the level of visit frequency. For instance, as presented in Figure 7, individuals in area N1 classified under category H displayed the pattern RSAMI, which indicates that their visit frequency is highest for Recreation (78.6%), followed by Scholar (35.6%), Agriculture (5.6%), Medical (2.7%), and Industrial (2.2%). This sequence shows the priority or preference of individuals in this area for different activities and services. Additional considerations include the percentage difference between the two methods, with a small difference suggesting close information alignment. In cases of significant differences, we delve into contextual factors such as geographical layers of land use, demographic information, and visit data to understand the causes.



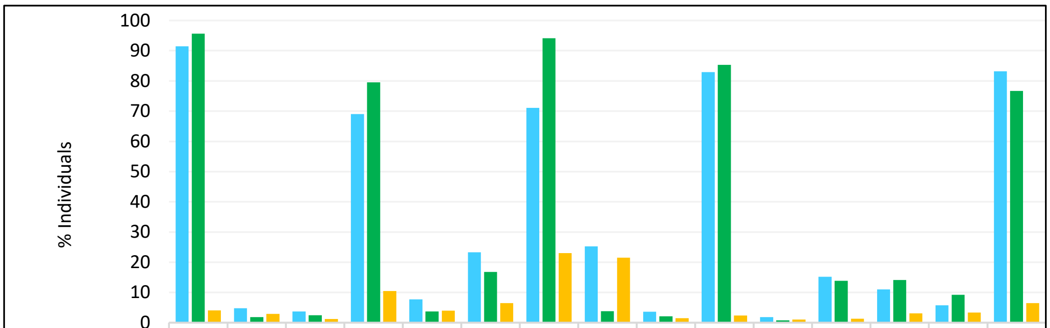
	Medical L	Medical M	Medical H	Scholar L	Scholar M	Scholar H	Agriculture L	Agriculture M	Agriculture H	Industrial L	Industrial M	Industrial H	Recreation L	Recreation M	Recreation H
Framework	92.8	4.5	2.7	50.2	14.5	35.3	89.7	4.7	5.6	96.8	1.0	2.2	19.1	2.3	78.6
Field Data	95.1	2.6	2.3	59.7	8.4	31.9	93.1	2.1	4.8	96.0	2.3	1.7	22.7	11.8	65.5
Difference	2.3	1.9	0.4	9.5	6.1	3.4	3.4	2.6	0.8	0.8	1.3	0.5	3.6	9.5	13.1

Figure 7. Comparison of the proposed framework and field data for area N1.

By classifying visit frequencies into high, balanced, and low categories, we systematically identify dominant activity patterns across different areas, providing a structured method to compare lifestyle patterns across various regions and demographics. The use of ordered patterns, such as RSAMI, simplifies complex data into understandable lifestyle signatures, facilitating straightforward comparisons between areas and highlighting the primary focus of the population in terms of activities and services. Understanding which criteria dominate in each area, such as high recreational visits in area N1, provides valuable insights into the lifestyle and needs of the residents. This approach is instrumental in urban planning and resource allocation, ensuring that the people's needs are effectively met.

Figure 7 displays the experiment results for area N1. The level pattern for all criteria in the proposed framework closely mirrors the field data, with negligible differences in three out of five criteria. Notably, the Scholar and Recreation criteria, exhibiting more divergence, share similar patterns in both methods. Criteria are further arranged in descending order for classes H and B. The results for class H and class B, except for the Recreation criterion, are consistent across both the proposed framework and field data, both demonstrating the RSAMI pattern. The four criteria related to workplaces reveal valuable insights into jobs and lifestyles, particularly highlighting the significant presence of academic professionals in an area known as the University Town, close to several universities.

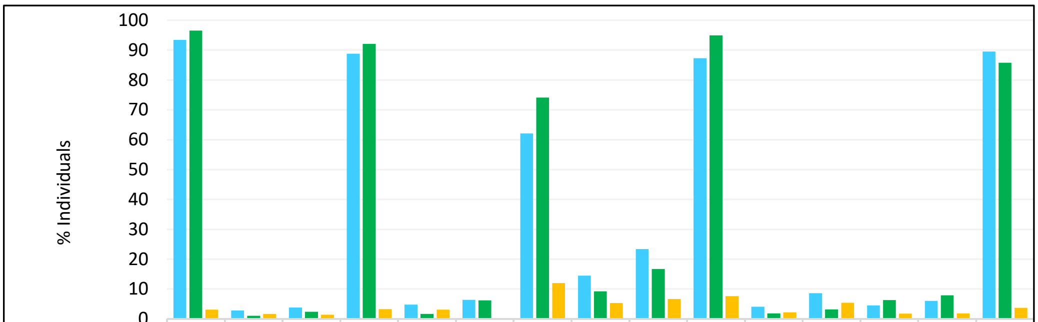
Figure 8 presents the experimental results for area N2. The level pattern of the Medical criterion exhibits slight variation between the proposed framework and field data analysis: LBH for the former and LHB for the latter. However, the values for both B and H in each method are small, with negligible differences. The combined percentages of B and H for each technique indicate that a limited number of individuals in this area frequent medical-related places. Despite the slight disparity in the level pattern for the Medical criterion, the overall inconsistency is insignificant. The level pattern for the remaining four criteria remains consistent between the two methods. In the Agriculture criterion, the percentage of individuals in class H is minimal for both methods, but class B holds significance only for the proposed method. Further investigation into geographic information and neighborhood features revealed the area's absence of fresh vegetable sales (Alborz Town). Consequently, a substantial number of residents opt to visit organic orchards for crop purchases. To substantiate this observation, a review of movement data in the area confirmed that while several individuals visit agricultural areas, their visits are numerous but of short duration. Additionally, criteria are ordered separately in descending order for classes H and B. The class H result in the proposed framework and field data analysis is RSIM/A. Although the percentages for Medical and Agriculture criteria exhibit slight differences, they interchange between the two methods, denoted by a slash in the pattern.



	Medical			Scholar			Agriculture			Industrial			Recreation		
	L	M	H	L	M	H	L	M	H	L	M	H	L	M	H
Framework	91.5	4.8	3.7	69.0	7.7	23.3	71.1	25.3	3.6	82.9	1.9	15.2	11.0	5.8	83.2
Field Data	95.6	1.9	2.5	79.5	3.7	16.8	94.1	3.8	2.1	85.3	0.8	13.9	14.1	9.2	76.7
Difference	4.1	2.9	1.2	10.5	4.0	6.5	23.0	21.5	1.5	2.4	1.1	1.3	3.1	3.4	6.5

Figure 8. Comparison of the proposed framework and field data for area N2.

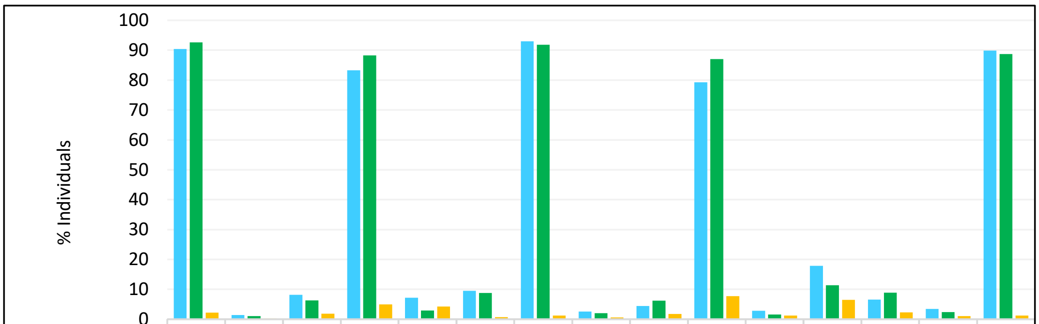
Figure 9 depicts the experimental outcomes in area N3. The level pattern for all criteria in the proposed framework aligns precisely with the field data analysis. Particularly noteworthy is the Recreation criterion, with a class H percentage of approximately 90%, a logical observation given the proximity of the area to the RC. Regarding job-related criteria, Agriculture exhibits the highest percentages in both classes H and B among the measured criteria. Consequently, the demographic structure of this area suggests a prevalence of individuals engaged in agricultural occupations, contrasting with areas N1 and N2, which have a lower proportion of academic professionals. Furthermore, criteria are arranged in descending order for classes H and B in this area. The result for class H for both methods is RAISM. Similarly, the result for class B for both methods is approximately ARI/SM. Although the percentages for Industrial and Scholar criteria exhibit slight differences and interchange between the two methods, the overall similarity of these patterns is deemed acceptable.



	Medical			Scholar			Agriculture			Industrial			Recreation		
	L	M	H	L	M	H	L	M	H	L	M	H	L	M	H
Framework	93.4	2.8	3.8	88.8	4.8	6.4	62.1	14.5	23.4	87.3	4.1	8.6	4.5	6.0	89.5
Field Data	96.5	1.1	2.4	92.1	1.7	6.2	74.1	9.2	16.7	94.9	1.9	3.2	6.3	7.9	85.8
Difference	3.1	1.7	1.4	3.3	3.1	0.2	12.0	5.3	6.7	7.6	2.2	5.4	1.8	1.9	3.7

Figure 9. Comparison of the proposed framework and field data for area N3.

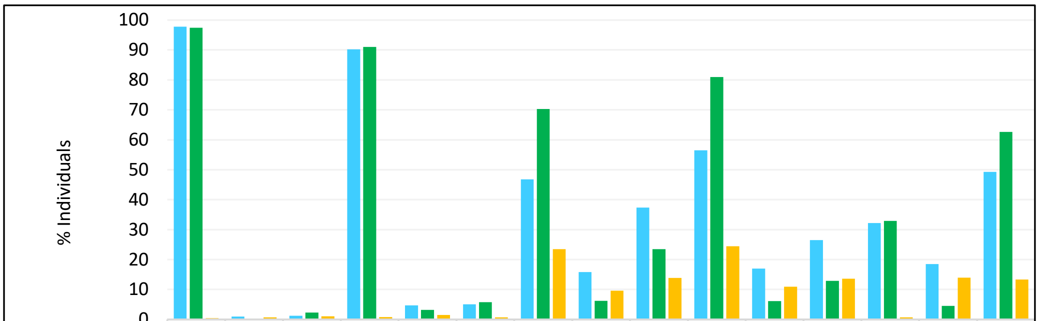
Figure 10 showcases the outcomes of our experiments in area N4. The level pattern for all criteria in the proposed framework mirrors the analyzed field data, with small differences in almost all cases. Additionally, the percentage of individuals in class H for the Recreation criterion is approximately 90%, a norm considering the proximity of this area to the RC. Another comparative analysis involves the arrangement of criteria in descending order for class H, resulting in RISMA for both methods. Similarly, the pattern for class B in both methods is SRI/AM. Although the positions of the Industrial and Agriculture criteria are switched, it is crucial to note that these two criteria exhibit very close percentages, indicating that the patterns in both methods are nearly identical. In terms of job-related criteria, the chart highlights that among the measured criteria, Industrial stands out significantly. This observation suggests a notable presence of individuals in this area who regularly visit industrial environments such as factories and industrial parks. This could be indicative of a strong industrial sector in the region, potentially contributing significantly to local employment and economic activity.



	L	M	H	L	M	H	L	M	H	L	M	H	L	M	H
	Medical			Scholar			Agriculture			Industrial			Recreation		
Framework	90.4	1.4	8.2	83.3	7.2	9.5	93.0	2.6	4.4	79.3	2.8	17.9	6.6	3.5	89.9
Field Data	92.6	1.1	6.3	88.3	2.9	8.8	91.8	2.0	6.2	87.0	1.6	11.4	8.9	2.4	88.7
Difference	2.2	0.3	1.9	5.0	4.3	0.7	1.2	0.6	1.8	7.7	1.2	6.5	2.3	1.1	1.2

Figure 10. Comparison of the proposed framework and field data for area N4.

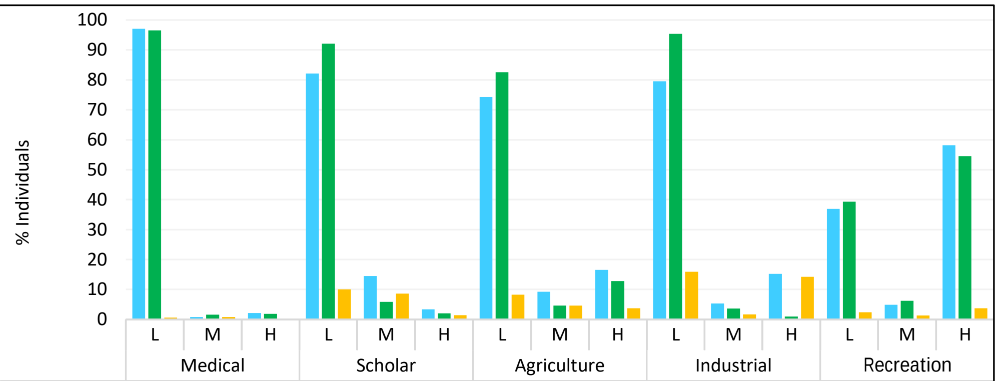
Figure 11 presents the outcomes of our experiments in area N5. Notably, the level pattern for all criteria is consistent for both methods in this area. Furthermore, when criteria are ordered based on the values of class H, the resulting pattern is identical for both methods, denoted as RAISM. Class H represents individuals who frequently visit related locations, while class B includes people with a lower frequency of visits to the desired locations. However, in this area, ordering criteria based on class B yields different patterns for the two methods. To assess the impact of this discrepancy with the adaptation of the class H pattern, we sorted the criteria in descending order according to the sum of classes H and B (similar to sorting criteria by class B in ascending order). The resulting pattern is consistent for both methods and aligns with the classification pattern based on the frequency of class H. Despite the notable differences between the two methods in criteria such as Agriculture, Industrial, and Recreation, the level of pattern matching is deemed acceptable for area N5.



	L	M	H	L	M	H	L	M	H	L	M	H	L	M	H
	Medical			Scholar			Agriculture			Industrial			Recreation		
Framework	97.8	1.0	1.2	90.2	4.7	5.1	46.8	15.8	37.4	56.5	17.0	26.5	32.2	18.5	49.3
Field Data	97.4	0.3	2.3	91.0	3.2	5.8	70.3	6.2	23.5	81.0	6.1	12.9	32.9	4.5	62.6
Difference	0.4	0.7	1.1	0.8	1.5	0.7	23.5	9.6	13.9	24.5	10.9	13.6	0.7	14.0	13.3

Figure 11. Comparison of the proposed framework and field data for area N5.

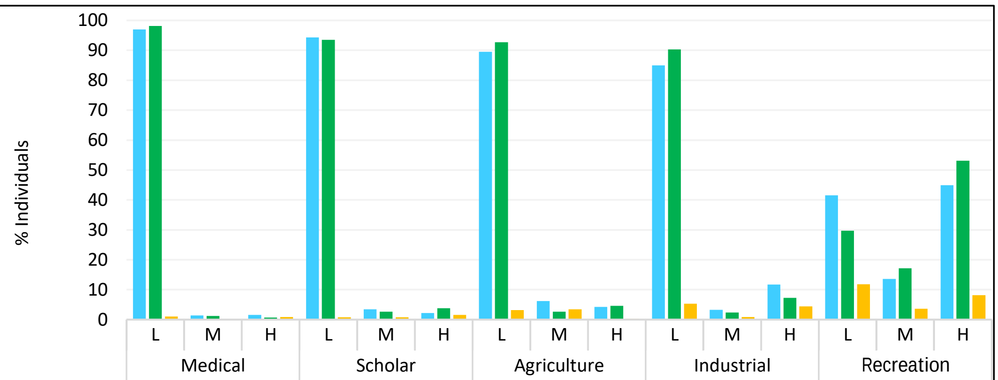
Figure 12 illustrates the outcomes of experiments in area N6. The level pattern of all criteria, except Industrial, is consistent for both proposed methods. When sorting the criteria based on the frequency of individuals in class H, the proposed framework yields the RAISM pattern, while the field data analysis leads to the RASMI pattern. However, removing the criterion Industrial results in identical patterns. Notably, the criterion Industrial exhibits a significant percentage difference of individuals in class H (14.2%). In terms of job-related criteria, the proposed framework suggests a demographic context in this area that includes a notable number of individuals who frequent industrial and agricultural places, with secondary connections to academic or medical locations. Similarly, according to field data analysis, the demographic context involves a significant number of people traveling to agricultural places, with secondary connections to academic or medical locations. Despite the major discrepancy in one of the five criteria, the level of pattern matching in this area is considered acceptable.



	Medical L	Medical M	Medical H	Scholar L	Scholar M	Scholar H	Agriculture L	Agriculture M	Agriculture H	Industrial L	Industrial M	Industrial H	Recreation L	Recreation M	Recreation H
Framework	97.1	0.8	2.1	82.1	14.5	3.4	74.3	9.2	16.5	79.5	5.3	15.2	36.9	4.9	58.2
Field Data	96.5	1.6	1.9	92.1	5.9	2.0	82.6	4.6	12.8	95.4	3.6	1.0	39.3	6.2	54.5
Difference	0.6	0.8	0.2	10.0	8.6	1.4	8.3	4.6	3.7	15.9	1.7	14.2	2.4	1.3	3.7

Figure 12. Comparison of the proposed framework and field data for area N6.

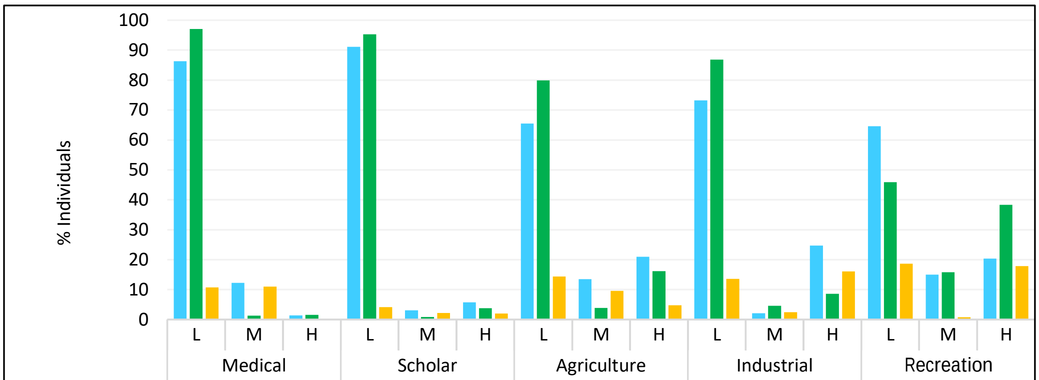
Figure 13 depicts the outcomes of our experiments in area N7. Notably, the level pattern for the three criteria Medical, Industrial, and Recreation is consistent, and for Scholar and Agricultural, where classes H and B are merged, their values are negligible and closely aligned. Hence, the compatibility of these two patterns is deemed acceptable. Furthermore, the pattern of ordering the criteria based on the values of class H is identical in both methods, denoted as RIASM. Regarding job-related criteria, based on the values of classes H and B, it can be inferred that the residents of this area predominantly visit industrial places, followed by agricultural locations.



	Medical L	Medical M	Medical H	Scholar L	Scholar M	Scholar H	Agriculture L	Agriculture M	Agriculture H	Industrial L	Industrial M	Industrial H	Recreation L	Recreation M	Recreation H
Framework	97.0	1.4	1.6	94.3	3.5	2.2	89.5	6.2	4.3	85.0	3.3	11.7	41.5	13.6	44.9
Field Data	98.1	1.2	0.7	93.5	2.7	3.8	92.7	2.7	4.6	90.3	2.4	7.3	29.7	17.2	53.1
Difference	1.1	0.2	0.9	0.8	0.8	1.6	3.2	3.5	0.3	5.3	0.9	4.4	11.8	3.6	8.2

Figure 13. Comparison of the proposed framework and field data for area N7.

Figure 14 illustrates the outcomes of our experiments in area N8. The level pattern is consistent for both methods in all criteria except Medical. Further investigations in this area revealed a significant increase in the number of people traveling to the hospital due to a prevalent disease during the test period, directly impacting the number of people in class B in the Medical criterion. It is crucial to note that in the proposed framework, we rigorously defined the membership function for classes B and H of the Medical criterion so that with the number of regular visits, the degree of membership in these classes increases. By ordering the criteria based on the percentage of people in class H, the IARSM pattern is obtained for the proposed framework, and the RAISM pattern is proposed for field data analysis. Since the values of class H in the three criteria Agriculture, Industry, and Recreation in the proposed framework are close to each other, with slight neglect, IAR and RAI can be considered similar here. Therefore, the patterns of class H in this area exhibit an acceptable similarity for both methods.



	Medical			Scholar			Agriculture			Industrial			Recreation		
	L	M	H	L	M	H	L	M	H	L	M	H	L	M	H
Framework	86.3	12.3	1.4	91.1	3.1	5.8	65.5	13.5	21.0	73.2	2.1	24.7	64.6	15.0	20.4
Field Data	97.1	1.3	1.6	95.3	0.9	3.8	79.9	3.9	16.2	86.8	4.6	8.6	45.9	15.8	38.3
Difference	10.8	11.0	0.2	4.2	2.2	2.0	14.4	9.6	4.8	13.6	2.5	16.1	18.7	0.8	17.9

Figure 14. Comparison of the proposed framework and field data for area N8.

We calculated the Root Mean Square Error (RMSE) for both anonymized mobile movement data and survey-based questionnaire (field) data to compare the measured lifestyle patterns across all studied areas. The analysis was also performed for the three classes (High, Medium, and Low) based on the number of visits to PoIs. The RMSE was calculated using Equation (1) to evaluate the consistency between the two methods. Here, d_i represents the i^{th} cell in the difference row within the desired tables (Figures 7–14).

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(d_i)^2}. \tag{1}$$

Table 1 presents the RMSE values indicating the discrepancy between mobile data and survey data across different regions and classes. Notably, the overall RMSE for all regions (5.167) suggests a moderate level of agreement between the two data collection methods. However, the specific RMSE values for each class (High, Balanced, Low) reveal nuanced differences in data consistency. Generally, RMSE values are higher in the Low visitation class across most regions, indicating that occasional frequency visits are more challenging to estimate accurately using mobile data or surveys. The Balanced and High categories show relatively lower RMSE values, suggesting that moderate or higher visit patterns are more consistently captured across data sources. In addition, the aggregate RMSE values across all regions (7.032 for Low, 3.869 for Balanced, and 3.959 for High) underscore the variability inherent in anonymous mobility data versus field data, especially in diverse urban settings. The discrepancies highlighted by the RMSE analysis suggest several factors influencing data accuracy, such as the anonymous nature of mobile movement data and potential variability in responses due to different individuals completing the surveys. These variations can arise from differences in individuals' perceptions of what constitutes a visit to a PoI or from limitations in the granularity and coverage of mobile data. Despite these differences, the findings indicate that the proposed framework is capable of providing a good estimation of lifestyle patterns across diverse urban areas. This capability underscores the framework's potential as a reliable tool for urban planners and policymakers in understanding and addressing the needs of urban populations.

Table 1. Comparison of RMSE between the proposed framework and field data.

$Area$	$RMSE_L$	$RMSE_M$	$RMSE_H$	$RMSE$
N1	4.913	5.282	6.070	5.443
N2	11.588	9.994	4.240	9.168
N3	6.716	3.132	4.237	4.928
N4	4.379	2.078	3.195	3.351
N5	15.191	9.052	10.552	11.889
N6	9.250	4.480	6.797	7.114
N7	5.990	2.311	4.244	4.444
N8	13.239	6.707	11.015	10.670
All	7.032	3.869	3.959	5.167

4.3. Study Limitations

Despite the strengths of the proposed framework, several limitations must be acknowledged. These limitations highlight areas for further research and refinement, particularly in terms of improving data resolution, enhancing validation efforts, and expanding the framework to different urban settings.

Data Granularity. The mobility data collected from BTS signals are less precise compared to GPS data. While BTS data allow for capturing general movement patterns, they may lack the granularity needed to accurately identify visits to small or closely located PoIs, such as shops or residential buildings.

Field Data Collection. While field data were collected for validation, the limited sample size and potential biases in self-reported data (e.g., inaccurate recall or subjective judgments) may affect the generalizability of the findings. Larger-scale field studies may be required for more robust validation.

Anonymization and Privacy Constraints. Due to privacy concerns, the dataset used in this study was anonymized, preventing us from linking specific mobility behaviors to socio-demographic factors.

4.4. Feature Applications

The proposed framework was built with a focus on adaptability to make it suitable for deployment in production environments across various domains and flexible enough to support the development of diverse APIs for seamless integration with other systems and services. The real-time processing of large datasets, particularly in densely populated areas, requires sufficient computational resources. Cloud-based solutions and distributed computing can mitigate this challenge, enabling real-time analysis and reporting. This section clarifies how our framework can be adapted for real-world applications.

Urban Planning and Infrastructure Development. The framework can be integrated into city planning systems to analyze the lifestyle patterns of residents and identify high-demand areas for transportation, commercial activities, or public services. By understanding the dynamic nature of urban mobility, planners can make more informed decisions regarding infrastructure investments and urban design.

Optimization of Public Services. Government agencies and municipalities can use the insights generated by this framework to optimize public service distribution. For example, transportation networks, waste collection, or emergency response services can be adjusted based on mobility patterns and lifestyle indicators in different urban subdivisions.

Healthcare and Epidemic Control. The framework can be applied in healthcare settings to monitor mobility patterns during disease outbreaks, enabling targeted interventions and efficient allocation of healthcare resources. Additionally, by tracking changes in population movement, healthcare providers can improve emergency response times and plan better for future healthcare needs.

Business and Commercial Strategy. Businesses can utilize the framework to identify areas of high commercial potential based on foot traffic and visitation patterns to specific points of interest. These data can guide location-based marketing strategies, optimize retail location planning, and improve customer experience by tailoring services to local lifestyle patterns.

Environmental Monitoring and Sustainability. The framework can help monitor the environmental impacts of human mobility and inform sustainability efforts. By understanding how residents move and congregate, urban planners can design more environmentally friendly transportation systems and mitigate the ecological effects of urban sprawl.

5. Conclusions

This article introduces a lifestyle approximation framework called LEAF that collects, indexes and analyzes human movement using cell phone tracking data. LEAF characterizes the human lifestyle based on the frequency of visiting specific PoI classes. The evaluation of the framework involves studying a mid-scale city, indexing data from eight neighborhoods based on 13 location types, and calculating approximations for five criteria at three levels: low, moderate, and high. Next, for each studied area, the criteria pattern at each level was calculated based on the frequency order of their visits. The framework's efficiency is assessed by comparing the obtained patterns with analyzed field data. The comparison involves extracting patterns based on criteria and subclasses to evaluate the alignment of corresponding patterns. Our systematic classification of visit frequencies provides a comprehensive framework for understanding and comparing lifestyle patterns across different regions. The RMSE of 5.167 between the proposed framework's results and survey-based data suggests that the framework provides a reliable estimation of lifestyle patterns across diverse urban areas. Additionally, the proposed ordered patterns distill complex data into accessible insights, revealing the primary activities and services that shape the daily lives of residents. The LEAF framework offers substantial potential beyond lifestyle pattern analysis by systematically categorizing human movement data based on visit frequencies to PoIs. It can provide valuable insights for urban planning, helping planners optimize infrastructure and services according to residents' mobility and usage patterns. High PoI frequencies could indicate the need for improved transportation or healthcare, while low frequencies might highlight underserved regions. LEAF also benefits urban design by aiding the creation of walkable, accessible spaces. Future research could expand LEAF by incorporating socioeconomic factors and enhancing its application in urban dynamics prediction and strategic planning. Its cross-regional validation creates opportunities for global smart city initiatives.

Author Contributions: Conceptualization, S.M.; methodology, M.K.F. and S.G.; software, M.K.F.; validation, S.M., S.G. and S.S.; formal analysis, M.K.F.; investigation, S.M., M.K.F. and S.G; resources, M.K.F. and S.G; data curation, M.K.F.; writing—original draft preparation, S.M., M.K.F., S.G. and S.S.; writing—review and editing, S.M., M.K.F., S.G. and S.S.; visualization, M.K.F. and S.G.; supervision, S.M.; project administration, S.G.; funding acquisition, S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korean government (MSIT) under Grant RS-2023-00278294.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to privacy reasons.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sobel, M.E. *Lifestyle and Social Structure: Concepts, Definitions, Analyses*; Elsevier: Amsterdam, The Netherlands, 2013.
2. Gehrke, S.R.; Singleton, P.A.; Clifton, K.J. Understanding stated neighborhood preferences: The roles of lifecycle stage, mobility style, and lifestyle aspirations. *Travel Behav. Soc.* **2019**, *17*, 62–71. [CrossRef]
3. Li, J.; Auchincloss, A.H.; Rodriguez, D.A.; Moore, K.A.; Diez Roux, A.V.; Sánchez, B.N. Determinants of residential preferences related to built and social environments and concordance between neighborhood characteristics and preferences. *J. Urban Health* **2020**, *97*, 62–77. [CrossRef] [PubMed]
4. Tiyachareonsri, S.; Chavarnakul, T.; Chandrachai, A.; Triukose, S. How consumer preference determines site selection in a metropolitan setting: Analysis of retailer perspective to stay ahead of the competition in the aftermath of a large-scale crisis. *J. Retail. Consum. Serv.* **2024**, *78*, 103704. [CrossRef]
5. Shieh, C.H.; Xu, Y.; Ling, I.L. How location-based advertising elicits in-store purchase. *J. Serv. Mark.* **2019**, *33*, 380–395. [CrossRef]
6. Rabindranath, M.; Singh, A.K. Types of Advertising. In *Advertising Management: Concepts, Theories, Research and Trends*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 39–78.
7. Wang, M.; Li, H.; Cui, J.; Bhowmick, S.S.; Liu, P. FROST: Movement History–Conscious Facility Relocation. *ACM Trans. Intell. Syst. Technol. (TIST)* **2020**, *11*, 1–26. [CrossRef]
8. Gao, Q.L.; Zhong, C.; Yue, Y.; Cao, R.; Zhang, B. Income estimation based on human mobility patterns and machine learning models. *Appl. Geogr.* **2024**, *163*, 103179. [CrossRef]
9. Moghari, S.; Ghorani, M. A symbiosis between cellular automata and dynamic weighted multigraph with application on virus spread modeling. *Chaos, Solitons Fractals* **2022**, *155*, 111660. [CrossRef]
10. Kato, H.; Takizawa, A. Human mobility and medical costs of lifestyle-related diseases during the COVID-19 pandemic: A cross-sectional study in Japan. *J. Transp. Health* **2024**, *34*, 101728. [CrossRef]
11. Kaluarachchi, Y. Implementing data-driven smart city applications for future cities. *Smart Cities* **2022**, *5*, 455–474. [CrossRef]
12. Ma, J.; Li, B.; Mostafavi, A. Characterizing urban lifestyle signatures using motif properties in network of places. *Environ. Plan. B Urban Anal. City Sci.* **2024**, *51*, 889–903. [CrossRef]
13. Xu, W.; Chow, C.Y.; Zhang, J.D. CALBA: Capacity-aware location-based advertising in temporary social networks. In Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Orlando, FL, USA, 5–8 November 2013; pp. 364–373.
14. Li, B.; Fan, C.; Chien, Y.H.; Hsu, C.W.; Mostafavi, A. Mobility behaviors shift disparity in flood exposure in US population groups. *Int. J. Disaster Risk Reduct.* **2024**, *108*, 104545. [CrossRef]
15. Li, Z.; Ding, B.; Han, J.; Kays, R.; Nye, P. Mining periodic behaviors for moving objects. In Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 25–28 July 2010; pp. 1099–1108.
16. Demissie, M.G.; Phithakkitnukoon, S.; Kattan, L.; Farhan, A. Understanding human mobility patterns in a developing country using mobile phone data. *Data Sci. J.* **2019**, *18*, 1. [CrossRef]
17. Yang, X.; Li, J.; Fang, Z.; Chen, H.; Li, J.; Zhao, Z. Influence of residential built environment on human mobility in Xining: A mobile phone data perspective. *Travel Behav. Soc.* **2024**, *34*, 100665. [CrossRef]
18. Yang, H.; Lv, S.; Guo, B.; Dai, J.; Wang, P. Uncovering spatiotemporal human mobility patterns in urban agglomerations: A mobility field based approach. *Phys. A Stat. Mech. Its Appl.* **2024**, *637*, 129571. [CrossRef]
19. Frantzich, H.; Fridolf, K.; Liljestrand, S.; Henningsson, A.; Lundin, J. Locating people in tunnels using Wi-Fi technology. *Fire Saf. J.* **2024**, *146*, 104178. [CrossRef]
20. Fu, R.; Xiao, D.; Fan, Y. A novel cell phone localization solution for trapped victims based on compressed RSSI fluctuation range and PSO-BP neural network. *Measurement* **2024**, *225*, 114014. [CrossRef]
21. Verma, R.; Mittal, S.; Lei, Z.; Chen, X.; Ukkusuri, S.V. Comparison of home detection algorithms using smartphone GPS data. *EPJ Data Sci.* **2024**, *13*, 6. [CrossRef]
22. Anderson, I.; Maitland, J.; Sherwood, S.; Barkhuus, L.; Chalmers, M.; Hall, M.; Brown, B.; Muller, H. Shakra: Tracking and sharing daily activity levels with unaugmented mobile phones. *Mob. Netw. Appl.* **2007**, *12*, 185–199. [CrossRef]
23. Chen, X.; Wan, X.; Ding, F.; Li, Q.; McCarthy, C.; Cheng, Y.; Ran, B. Data-driven prediction system of dynamic people-flow in large urban network using cellular probe data. *J. Adv. Transp.* **2019**, *2019*, 9401630.
24. Trogh, J.; Plets, D.; Surewaard, E.; Spiessens, M.; Versichele, M.; Martens, L.; Joseph, W. Outdoor location tracking of mobile devices in cellular networks. *EURASIP J. Wirel. Commun. Netw.* **2019**, *2019*, 115. [CrossRef]
25. Okmi, M.; Por, L.Y.; Ang, T.F.; Ku, C.S. Mobile phone data: A survey of techniques, features, and applications. *Sensors* **2023**, *23*, 908. [CrossRef] [PubMed]
26. Scholler, R.; Alaoui-Ismaïli, O.; Renaud, D.; Couchot, J.F.; Ballot, E. In-stream mobility and speed estimation of mobile devices from mobile network data. *Transportation* **2024**, 1–36. [CrossRef]
27. Kubik, A. The Use of Artificial Intelligence in the Assessment of User Routes in Shared Mobility Systems in Smart Cities. *Smart Cities* **2023**, *6*, 1858–1878. [CrossRef]
28. Šveda, M.; Hurbánek, P.; Madajová, M.S.; Rosina, K.; Förstl, F.; Záboj, P.; Výbošt'ok, J. When spatial interpolation matters: Seeking an appropriate data transformation from the mobile network for population estimates. *Comput. Environ. Urban Syst.* **2024**, *110*, 102106. [CrossRef]

29. Calabrese, F.; Ferrari, L.; Blondel, V.D. Urban sensing using mobile phone network data: A survey of research. *ACM Comput. Surv. (Csur)* **2014**, *47*, 1–20. [CrossRef]
30. Fontes, T.; Arantes, M.; Figueiredo, P.V.; Novais, P. A cluster-based approach using smartphone data for bike-sharing docking stations identification: Lisbon case study. *Smart Cities* **2022**, *5*, 251–275. [CrossRef]
31. Yang, Y.; Pentland, A.; Moro, E. Identifying latent activity behaviors and lifestyles using mobility data to describe urban dynamics. *EPJ Data Sci.* **2023**, *12*, 15. [CrossRef] [PubMed]
32. Jia, Y.; Ge, Y.; Ling, F.; Guo, X.; Wang, J.; Wang, L.; Chen, Y.; Li, X. Urban land use mapping by combining remote sensing imagery and mobile phone positioning data. *Remote Sens.* **2018**, *10*, 446. [CrossRef]
33. Yuan, G.; Chen, Y.; Sun, L.; Lai, J.; Li, T.; Liu, Z. Recognition of functional areas based on call detail records and point of interest data. *J. Adv. Transp.* **2020**, *2020*, 8956910. [CrossRef]
34. Nalin, A.; Vignali, V.; Lantieri, C.; Cappellari, D.; Zamengo, B.; Simone, A. Assessing veracity of big data: An in-depth evaluation process from the comparison of Mobile phone traces and groundtruth data in traffic monitoring. *J. Transp. Geogr.* **2024**, *118*, 103930. [CrossRef]
35. Peng, Y.; Song, G.; Guo, M.; Wu, L.; Yu, L. Investigating the impact of environmental and temporal features on mobile phone distracted driving behavior using phone use data. *Accid. Anal. Prev.* **2023**, *180*, 106925. [CrossRef] [PubMed]
36. Liu, Y.; Lu, A.; Yang, W.; Tian, Z. Investigating factors influencing park visit flows and duration using mobile phone signaling data. *Urban For. Urban Green.* **2023**, *85*, 127952. [CrossRef]
37. Chen, Z.; Li, P.; Jin, Y.; Bharule, S.; Jia, N.; Li, W.; Song, X.; Shibasaki, R.; Zhang, H. Using mobile phone big data to identify inequity of aging groups in transit-oriented development station usage: A case of Tokyo. *Transp. Policy* **2023**, *132*, 65–75. [CrossRef]
38. Ijeh, S.; Okolo, C.A.; Arowoogun, J.O.; Adeniyi, A.O.; Omotayo, O. Predictive modeling for disease outbreaks: A review of data sources and accuracy. *Int. Med. Sci. Res. J.* **2024**, *4*, 406–419. [CrossRef]
39. Gupta, P.; Pandey, M.K. Role of AI for Smart Health Diagnosis and Treatment. In *Smart Medical Imaging for Diagnosis and Treatment Planning*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2024; pp. 23–45.
40. Eriskin, L.; Karatas, M.; Zheng, Y.J. A robust multi-objective model for healthcare resource management and location planning during pandemics. *Ann. Oper. Res.* **2024**, *335*, 1471–1518. [CrossRef]
41. Washington, V.; Guikema, S.; Mondisa, J.L.; Misra, A. A data-driven method for identifying the locations of hurricane evacuations from mobile phone location data. *Risk Anal.* **2024**, *44*, 390–407. [CrossRef]
42. Osorio Arjona, J. Analyzing post-COVID-19 demographic and mobility changes in Andalusia using mobile phone data. *Sci. Rep.* **2024**, *14*, 14828. [CrossRef]
43. Östh, J.; Toger, M.; Türk, U.; Kourtit, K.; Nijkamp, P. Leisure mobility changes during the COVID-19 pandemic–An analysis of survey and mobile phone data in Sweden. *Res. Transp. Bus. Manag.* **2023**, *48*, 100952. [CrossRef]
44. Coleman, N.; Liu, C.; Zhao, Y.; Mostafavi, A. Lifestyle pattern analysis unveils recovery trajectories of communities impacted by disasters. *Humanit. Soc. Sci. Commun.* **2023**, *10*, 1–13. [CrossRef]
45. Qu, X.; Xu, G.; Qi, J.; Bao, H. Identifying the spatial patterns and influencing factors of leisure and tourism in Xi'an based on point of interest (POI) data. *Land* **2023**, *12*, 1805. [CrossRef]
46. Yao, Y.; Zhang, H.; Shi, X.; Chen, J.; Li, W.; Song, X.; Shibasaki, R. LTP-Net: Life-Travel Pattern Based Human Mobility Signature Identification. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 14306–14319. [CrossRef]
47. Martin, H.; Wiedemann, N.; Reck, D.J.; Raubal, M. Graph-based mobility profiling. *Comput. Environ. Urban Syst.* **2023**, *100*, 101910. [CrossRef]
48. Fallah, M.K.; Moghari, S.; Nazemi, E.; Zahedi, M.M. Fuzzy ontology based document feature vector modification using fuzzy tree transducer. In Proceedings of the 2008 IEEE International Conference on Signal Image Technology and Internet Based Systems, Bali, Indonesia, 30 November–3 December 2008; pp. 38–44.