Gray Zone Detection: Using AI to Spot Urban Development Gaps

CAPSTONE PROJECT PHASE-1

Phase – I Report

Submitted by

Group No. 189

1. Shivansh Kaintura 21BCE10089
2. Mayank Singh 21BCE10097
3. Nikhil Santosh Thale 21BCE10377
4. Kopal Soni 21BCE10455
5. Dhruv Singh 21BCE10502

in partial fulfilment of the requirements for the degree of

Bachelor of Engineering and Technology

VIT Bhopal University
Bhopal
Madhya Pradesh

September 2024

Bonafide Certificate

Certified that this project report titled **"Gray Zone Detection: Using AI to Spot Urban Development Gaps"** is the Bonafede work of "21BCE10089 Shivansh Kaintura, 21BCE10097 Mayank Singh, 21BCE10377 Nikhil Santosh Thale, 21BCE10455 Kopal Soni, 21BCE10502 Dhruv Singh "who carried out the project work under my supervision.

This project report (DSN4095-Capstone Project Phase-I) is submitted for the Project Viva-Voce examination held on …………...

(Dr. Vijendra Singh Bramhe) Supervisor

(Dr. Vipin Jain) Reviewer -1

(Dr. Vivek Parasher) Reviewer-2

INDEX

CHAPTER 1: INTRODUCTION

As India experiences rapid urbanisation, smart city infrastructure plays an increasingly important role in meeting the needs of its expanding population. One of the most essential aspects of smart cities is the integration of healthcare systems, which is critical for providing equitable access to medical facilities. Healthcare services not only improve quality of life, but they also have a significant impact on economic productivity and social well-being. However, gaps in healthcare accessibility might result in the formation of "grey zones"—areas where healthcare services are limited or unavailable, leaving citizens vulnerable during medical emergencies.

Recent advances in deep learning and computer vision have transformed the processing of large-scale satellite and overhead data, enabling new methods of assessing infrastructure deficiencies. When high-resolution spatial data is analysed using convolutional neural networks (CNNs), it can provide unprecedented insights into urban planning difficulties. These systems can now detect patterns in overhead photos, identify infrastructure bottlenecks, and predict places with insufficient healthcare services. These advancements allow politicians to optimise resource allocation, ensuring that smart cities are both technologically advanced and socially inclusive. Despite their potential, problems such as inconsistent data resolution, complicated urban landscapes, and nonuniform object classes continue to prevent widespread use of these solutions.

Despite substantial advances in urban planning and healthcare integration, many Indian cities still struggle to provide uniform healthcare access. Identifying and eliminating grey zones is critical to building more resilient, inclusive, and efficient smart cities that provide optimal healthcare services to all people. We hope to create a system that maps hospital locations and identifies grey zones, utilising AI and geospatial data to promote urban planning and increase healthcare access.

1.1 Motivation

The pressing need to ensure equitable access to healthcare in India's rapidly urbanising regions is what motivates this effort. Disparities in the accessibility of essential services, particularly healthcare, worsen as cities grow and the population rises. Urban growth frequently results in the creation of "grey zones," or places with insufficient access to healthcare services. This poses a major risk to public health, especially in emergency situations when prompt access to care is crucial. In these grey areas, vulnerable groups frequently suffer terrible outcomes such as postponed medical care, higher rates of illness, and avoidable deaths, which feeds the cycle of poverty and inequality.

Technological developments in the fields of artificial intelligence, computer vision, and geospatial mapping have transformed our understanding of how to analyse and visualise urban settings, giving us the means to effectively tackle these obstacles. We can uncover important

information about the spatial distribution of healthcare services by utilising deep learning algorithms and satellite imagery. This will allow us to pinpoint underserved areas with previously unheard-of accuracy. The goal of this initiative is to use these technologies to provide policymakers and urban planners with useful information.

Our ultimate goal is to assist in the creation of more intelligent and inclusive communities where everyone has equal access to healthcare as a basic right. By filling in these gaps in healthcare, we hope to make a positive impact on a more equitable society where everyone has access to timely and quality medical care, improving public health and well-being in urban settings. We hope that this project will not only increase access to healthcare but also promote sustainability and resilience in the face of an urban setting that is changing quickly.

1.2 Objective

This project's main goal is to create a thorough mapping system that makes use of cutting-edge technologies, such as computer vision, artificial intelligence, and geographic information systems (GIS), in order to locate and assess "grey zones" in urban and rural areas where access to healthcare is insufficient. The project's specific goal is to map healthcare facilities in order to produce a precise geographic database of all the healthcare services that are currently provided, such as clinics, hospitals, and emergency rooms. It also aims to use deep learning algorithms to identify regions with inadequate healthcare facilities by analysing relevant data and satellite photos. The research will assess how easily accessible healthcare services are in designated grey areas, taking into account variables including the density of hospitals in an area and the distance between them.

By guaranteeing fair access to healthcare services, the ultimate objective is to enhance public health outcomes by promoting healthier urban populations and lowering inequities in health equity. By achieving these goals, we hope to establish a strong foundation for resolving healthcare inequalities in areas that are quickly urbanising and open up new opportunities for inclusive and wiser urban development.

CHAPTER 2: LITERATURE REVIEW

Numerous research has looked at the best conditions for identifying the differences between urban developed regions and the absence of daily supplies.

[1]. Manonmani et al. (2012) conducted a study titled "GIS design and application for urban planning using Cartosat-1 and IRS LISS-IV imagery." The researchers utilized sharpened georeferenced Cartosat-1 (PAN), and IRS LISS-IV (MX) fused satellite imagery to co-register the georeferenced area of interest (AOI) derived from administrative boundary shapefiles, following the spatial reference framework. The AOI delineates the city core boundary, including municipal or corporation limits, village boundaries, and settlement names. They classified the area into residential, institutional, commercial and industrial, road areas, and religious and cultural places, interpreted using on-screen digitization with key elements. This mapping technique supports urban planning by identifying vacant land, updating base maps, and aligning major roads and railways. The study also examined how urban quality depends on resources like water supply, infrastructure, and services such as health and education. Technologies like remote sensing, GPS, and GIS analysed the accessibility of key services using road network, buffer, and connectivity analysis. The mapping process was supported by IRS P6 LISS IV data. In conclusion, the study demonstrated the use of GIS for site management decisions, urban planning, and network analysis, especially for cultural heritage sites, providing key insights for future town planning.

[2]. Al Fanatseh and Saqallah (2021) conducted a study entitled "Evaluation of Accessibility to Public Services in the City of Aqaba Using Geographic Information Systems." The research aimed to evaluate public service accessibility in Aqaba, Jordan, using Geographic Information Systems (GIS). The researchers analysed various service sectors, including administrative, commercial, educational, health, and entertainment services. To measure accessibility, three primary methods were employed: (a) distance to the nearest service, (b) number of services within a certain distance, and (c) the average distance to all services. Both actual distances and population-weighted distances were used in the analysis. The study's results indicated that educational services had the highest accessibility, followed by entertainment and commercial services, while health services were generally less accessible. The research highlights the effectiveness of GIS in analysing and planning the distribution of public services. Recommendations from the study included accounting for population growth in underserved areas, prioritizing accessibility in service distribution planning, and creating planning standards tailored specifically to Jordanian cities.

[3]. Zhang et al. (2023) published a study entitled "The Convergence of Deep Learning and Computer Vision: Smart City Applications and Research Challenges." This research explores how deep learning, particularly in computer vision, is revolutionizing smart city development by providing innovative solutions across various domains. It highlights the critical role of computer vision in effectively analysing visual data, surpassing traditional methods. The study identifies key applications, such as enhancing urban transportation through advanced surveillance and traffic management, improving healthcare via real-time patient monitoring to detect anomalies, and bolstering security through intelligent video surveillance to identify suspicious activities in public areas. Additionally, it illustrates the impact of deep learning on agriculture, emphasizing technologies for yield prediction and disease detection to meet global food demands. The authors also address ongoing challenges, including improving accuracy across diverse environments, ensuring data privacy, building resilient infrastructures, and fostering community engagement in smart city initiatives.

[4]. Lee et al. (2023) published a study entitled "Smart City Architecture: Vision and Challenges." This study provides an in-depth analysis of the challenges and potential solutions related to smart city development, emphasizing the importance of data management in optimizing urban operations. Key challenges identified include the complexity and cost of establishing scalable and reliable IT infrastructure, the necessity of safeguarding sensitive data in a highly connected environment, and the effective management of the vast amounts of data generated. Other challenges encompass the high costs of investment, difficulties in achieving interoperability among diverse devices and networks, the need for scalability in large systems, and the social adaptation required for citizens to embrace new technologies. The authors propose a comprehensive design approach that addresses these issues by managing the city as a network of interconnected subsystems through a centralized data management system (CDMS). This includes organizing local data into zones for progressive development, ensuring interoperability with service-oriented architecture, and promoting innovation through an open data model while maintaining security. Overall, this study presents a valuable framework for understanding the complexities of smart city development and offers insights into potential solutions for the key challenges identified.

[5]. Head, Tran, and Blumenstock (2017) published a study entitled "Can Human Development be Measured with Satellite Imagery?" This research presents a preliminary investigation into the generalizability of satellite-based methods for estimating human development. The study replicated earlier work that demonstrated the potential of satellite imagery for predicting asset-based wealth in Rwanda. However, it found that this method could not be easily adapted to predict other "softer" development outcomes, such as health and access to clean drinking water, with the same accuracy in other countries, specifically Haiti and Nepal. The researchers identified several limitations in their approach. One issue was the insufficient "signal" in the satellite imagery, which is essential for providing relevant information about the development indicator in question. Additionally, the study faced high levels of measurement error in the "ground truth" data, making it difficult to accurately assess these softer outcomes. The focus on Haiti and Nepal further highlighted the challenge of generalizing the original model to different countries and contexts.

[6]. Abburi and Golla (2017) published a study entitled "Satellite Image Classification Methods and Techniques: A Review." The paper provides a comprehensive review of various methods and techniques used in satellite image classification. It emphasizes the critical role of satellite imagery in offering geographical information and the necessity for effective techniques to extract meaningful data from these images. The study categorizes satellite image classification methods into three main types: automated, manual, and hybrid. It primarily focuses on automated methods, discussing both supervised techniques—such as Artificial Neural Networks, Binary Decision Trees, and Image Segmentation—and unsupervised techniques, including ISODATA, K-Means, and Support Vector Machines.

[7]. Law and Deng (2018) introduced *Corner Net Detecting Objects as Paired KeyPoint's*, a groundbreaking method in the field of object detection. Instead of relying on anchor boxes—a standard approach in previous detection models—they proposed detecting objects by identifying paired key points, specifically the top-left and bottom-right corners of an object's bounding box. This approach simplifies object detection by reducing the complexity introduced by anchor boxes. To further enhance the precision of corner detection, the study also introduced the concept of corner pooling, a specialized pooling layer designed to improve the localization of object corners. This method significantly improved average precision (AP) scores on the MS-COCO dataset. The introduction of paired keypoints and corner pooling provided a more efficient and accurate way of detecting objects, streamlining the detection process and reducing computational overhead.

[8]. Duan et al. (2019) presented a paper titled CenterNet: Key point Triplets for Object Detection, introducing a novel approach to object detection that builds upon the previous Corner Net method.

While Corner Net detects objects using two key points (top-left and bottom-right corners), CenterNet enhances this by detecting objects as a triplet of key points: the top-left corner, bottom-right corner, and the centre of the object. This triplet approach improves detection accuracy, particularly for smaller objects. The study also introduced advanced techniques such as cascade corner and centre pooling to improve precision and recall. These strategies resulted in a marked reduction in false discovery rates and significant gains in average precision and recall, especially when evaluated on the MS-COCO dataset. The CenterNet model thus provides a more robust and efficient framework for object detection, improving upon its predecessor in several key aspects.

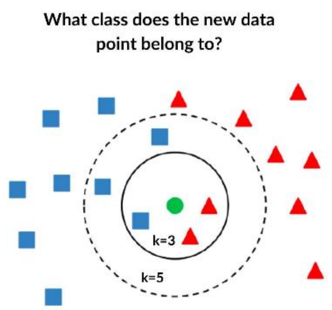
CHAPTER 3: METHODOLOGY

3.1 What is the KNN Algorithm?

K-Nearest Neighbours (KNN) is a simple, supervised machine learning technique that may be used for both classification and regression; however, it is most typically employed for classification. The core principle of KNN is that similar data points tend to be close together in the feature space. KNN assigns a label or value to a new data point using the 'K' nearest points (neighbours) from a known dataset. It is non-parametric, which means that it makes no assumptions about the underlying data distribution, making it adaptable and simple to implement.

3.2 How does the algorithm work?

K-Nearest Neighbours (KNN) begins with selecting a number, 'K', to signify how many neighbouring data points will impact the categorisation or prediction of a new data point. It then calculates the distance between the new data point and all other points in the dataset, usually using the Euclidean distance. Once the distances have been calculated, the algorithm selects the 'K' nearest data points and examines their labels. For classification tasks, the new point is assigned the most common label among its neighbours, but for regression, the forecast is based on the average of the neighbours' values. The images provided below show a new data point is clustered according to its nearest neighbours.



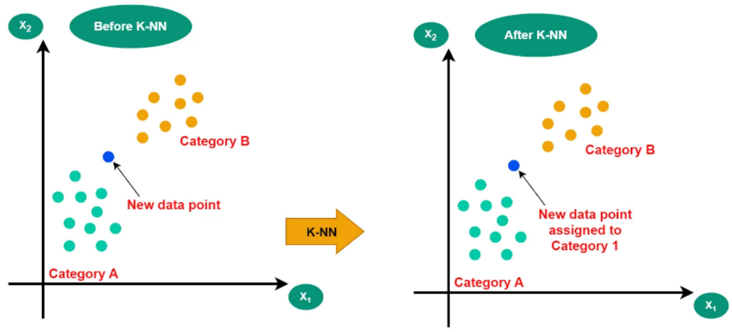


Fig 1. Choosing a value for number of neighbours

Fig 2. Process of clustering a new data point using KNN

3.3 Supporting Algorithms:

And with the help of and with the help of **Voronoi Diagrams**, space can be divided into regions based on the distance to points, where each point (or "seed") has a corresponding region of points closer to it than any other. In your project, hospitals act as seeds, and the study area (Area A) becomes the plane, with Voronoi cells representing each hospital's theoretical service area. This visualization aids in detecting grey zones by showing which regions are covered by hospitals and highlighting underserved areas. Large cells may indicate hospitals covering extensive areas, potentially pointing to grey zones at the edges. Voronoi edges show equidistant points between hospitals, useful for identifying the most distant areas from any hospital. Overlaying population data can reveal high-density areas with limited access to hospitals. Additionally, analysing boundaries of Voronoi cells helps understand transition zones between hospital service areas. Comparing the size and shape of these cells can identify disparities in hospital distribution, while large or irregularly shaped cells suggest optimal locations for new hospitals.

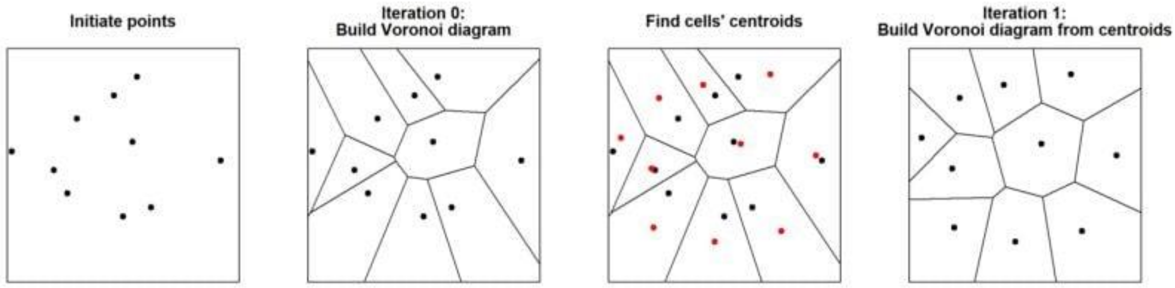


Fig 3. Example of Voronoi Cells

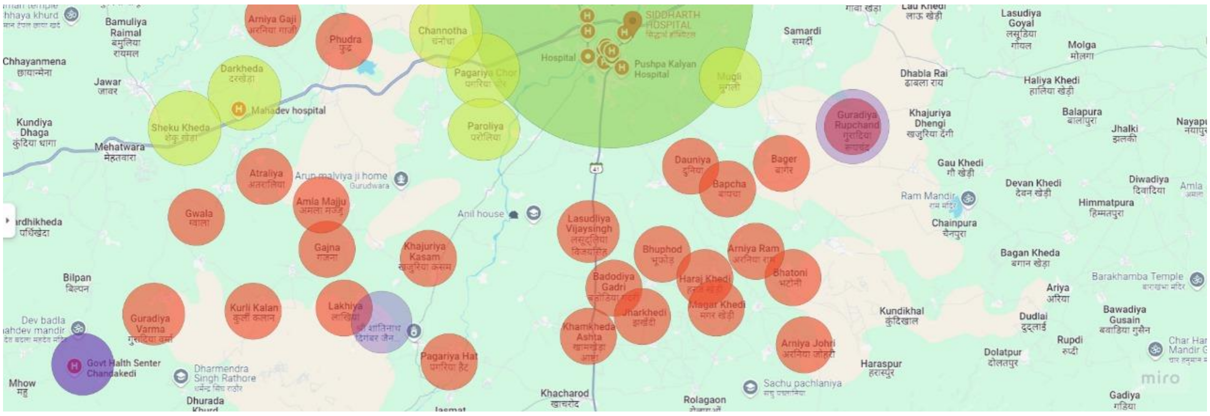
Voronoi diagrams are geometric structures that partition space into regions based on proximity to a set of seed points. Each region contains all points closest to a particular seed point. In the context of our research, hospitals serve as the seed points, and the map is divided into regions where each area is closest to a specific hospital. These regions, known as Voronoi cells, naturally represent hospital service areas, making them useful for visualizing how hospitals cover population centres.

To implement Voronoi diagrams, we first input the geographical coordinates of hospitals as seed points. The algorithm then divides the map into regions where each point is closest to its corresponding hospital. Population data can be overlaid onto these regions to identify grey zones—areas either outside the Voronoi cells or located far from hospital boundaries within the cells. These zones can be flagged as underserved. Tools like SciPy, QGIS, or Shapely can help visualize and compute Voronoi diagrams for more accurate analysis.

DBSCN: Density-Based Spatial Clustering of Applications with Noise, it is a powerful clustering algorithm that groups data points based on their density. Unlike other clustering algorithms, DBSCAN can identify clusters of arbitrary shapes and sizes, making it highly versatile for spatial data analysis. It can also detect noise or outliers, which are points that do not belong to any cluster. In the context of our research, DBSCAN can help identify clusters of population centres and hospitals, representing well-covered areas. Points that are not clustered, or are considered noise, can highlight potential grey zones—areas lacking adequate hospital coverage.

To implement DBSCAN, we begin by inputting geographical coordinates of population centres and hospitals. The algorithm groups points where population density or hospital density is high, marking these areas as clusters with sufficient hospital coverage. Points outside of dense clusters are flagged as noise, potentially identifying grey zones. DBSCAN requires careful tuning of two key parameters: epsilon (ε), which defines the maximum distance between two points to be considered part of the same cluster, and minimum samples, the minimum number of points required to form a cluster. This flexibility makes DBSCAN a valuable tool for detecting hospital coverage gaps.

3.4 Sample Implementation



The map uses color-coded regions to depict healthcare accessibility in the district. Dark green areas represent regions covered by 11 registered hospitals located in Astha, indicating reliable healthcare access. Surrounding these are the light green areas, which mark the villages that fall within the hospital coverage zones, also indicating adequate access to medical facilities. In contrast, red areas highlight villages with no nearby hospitals, signalling a lack of healthcare infrastructure. Additionally, purple areas represent locations incorrectly marked as hospitals in the dataset, where no actual hospital facilities exist, pointing out data irregularities

Step 1: Data Collection:

After we get the collected information about hospitals in the district, including the name, longitude, and latitude.

```
HospitalName, Longitude, Latitude
Hospital A, 77.58, 12.97
Hospital B, 77.60, 12.92
```

Similarly, for the villages, including their names, longitude, and latitude.

```
VillageName, Longitude, Latitude
Village X, 77.55, 12.90
Village Y, 77.53, 12.85
```

Step 2: Distance Calculation:

Use the **Haversine formula**, which computes the great-circle distance between two points on the Earth's surface, based on their longitude and latitude.

$$d = 2r \cdot \arcsin\left(\sqrt{\sin^2\left(\frac{\Delta\phi}{2}\right) + \cos(\phi_1)\cdot\cos(\phi_2)\cdot\sin^2\left(\frac{\Delta\lambda}{2}\right)}\right)$$

d = distance between the two points (in kilometers)

r = Earth's radius (mean radius = 6,371 km)

$\Delta\phi$ = difference in latitude (village and hospital)

$\Delta\lambda$ = difference in longitude (village and hospital)

$\phi 1$ and $\phi 2$= latitude of the village and hospital, respectively

Step 3: Step-by-Step KNN Implementation:

1. For each village in the dataset, iterate over all hospitals and calculate the distance using the Haversine formula.
2. Sort the hospitals by the calculated distance in ascending order.
3. Select the top 3 nearest hospitals for each village.

Step 5: Identify Grey Zones

- After finding the 3 nearest hospitals, if the closest hospital is more than 10 km away, flag the village as a "grey zone."

3.5 Individual Contribution by members:

Mayank:

I worked to identify areas where there were few hospitals, which we called "Grey zones. To this end, I established two priorities for our study. Area A is a large rectangle with 50 km on each side. Within this large area we have Zone B and the size of the zone changes depending on where the hospitals are located. To figure out how big Area B should be, I used a special statistical tool called the KNN algorithm. This helps compare more hospitals to places that may not have enough. To make our study more efficient, I collected the data myself. This was important because I found some mistakes in the information we had before. Some places that were called hospitals were not really hospitals. By eliminating these errors, I ensured that our study used correct information.

Shivansh:

I heavily engaged in identification of grey zones -places where there are no hospitals across urban regions based on satellite images. While reading through most research papers entirely, I learned about methodologies that other studies used along with challenges they came across while selecting mine for better insight and avoiding common pitfalls; it helped us form a better strategy.

I suggested that such mapping techniques can be carried out with the marking of crucial points on population centre and special radius marking the coverage area of every hospital. By such distance calculations between the population centres and hospitals, one would differentiate between so-called safe zones and grey zones.

My findings in my analysis were able to bring out huge gaps which exist in the covering of hospitals across urban regions. With such accuracy of detection, we included the K-nearest neighbours' algorithm, which allows areas to be classified not only based on proximity to healthcare facilities but also assists in explaining medically underserved areas more easily and has practical implications for improving access in cities.

Nikhil:

In this research I focused on analysing previous studies to strengthen our approach to identifying grey areas. This is an area without hospital coverage in an urban environment. After carefully reviewing research like our topic, I have identified keyways and areas that can be improved. This analysis not only shapes our approach; But it also lays the foundation for the use of accurate mathematical models.
Especially I support the use of a K-nearest neighbours (KNN) algorithm that incorporates Euclidean mathematics to improve the accuracy of our distance calculations. This way, I can calculate various mathematical formulas that will be more accurate. Important to refine our model and provide more confidence. Accurate identification of grey areas Helps ensure that results are measured. This makes the gaps in access to health services more clearly visible.

Dhruv:
In this project, I looked at the concept of smart cities, focussing on how computer vision (CV) and image detection technologies may improve healthcare accessibility. My contribution focused on how these technologies might identify urban development shortages, notably in the healthcare sector.
I studied various research articles to understand the characteristics of a "smart" city, such as the incorporation of AI, IoT, and data-driven decision-making processes. My research led me to investigate how CV is currently used in urban planning and healthcare contexts, specifically to map healthcare institutions and identify underserved areas. This insight led to the development of a complex mapping system that uses AI and GIS technology to successfully identify healthcare grey zones in metropolitan environments, with the goal of improving public health outcomes.

Kopal:

In this project, I began by researching advanced object detection models like Cascade R-CNN and CenterNet, which are commonly used for detecting objects in images. I thought these models might help us map healthcare facilities by analysing satellite images. However, while they were effective at identifying individual objects, they didn't meet our key requirement of measuring distances between places like hospitals and villages, which is crucial for identifying healthcare 'Gray zones. 'Through a process of evaluating these models and discussing their limitations with the team, I contributed to narrowing down the choices for a more suitable approach. This led the group to eventually adopt the K-Nearest Neighbors (KNN) algorithm, which is better suited for calculating distances between locations. My role was pivotal in guiding the team away from complex object detection models and toward a more practical solution that allowed us to effectively analyse the proximity of villages to healthcare facilities.

CHAPTER 4: CONCLUSION

This project focusses on identifying "grey zones"—areas without enough hospital coverage—with the purpose of improving healthcare accessibility in accordance with the United Nations' sustainable development objectives. We examined a 50km-by-50km region, identifying hospital clusters and using the K-Nearest Neighbours (KNN) algorithm, a non-parametric supervised learning method, to calculate ideal cluster sizes. Furthermore, tools such as DBSCAN and Voronoi diagrams assisted us in grouping hospitals and visualising their service regions.

To assure accuracy, we collected and corrected our own data, using tools like Folium, Geo Pandas, and Matplotlib to generate maps and visualise healthcare gaps. We also intend to use computer vision techniques such as image registration and fusion to analyse satellite images and automatically detect grey zones. These approaches will aid in the integration of custom maps with satellite imagery, resulting in a more thorough picture of the area.

Our work has the potential to help governments and organisations plan strategic hospital placement and healthcare improvements, particularly in underserved rural areas. Using a combination of innovative computer tools, this project provides a comprehensive solution to improve healthcare accessibility, ensuring that no location is left without necessary health services.

Reference and Publication

[1] (Manonmani, R., Prabaharan, S., Vidhya, R., & Ramalingam, M. (2012). GIS design and application for urban planning using Cartosat-1 and IRS LISS-IV imagery. Published in Geo-spatial Information Science.)

[2] (Al Fanatseh, A., & Saqallah, S. (2021). "Evaluation of Accessibility to Public Services in the City of Aqaba Using Geographic Information Systems." Published on ResearchGate.)

[3] (Zhang, X., et al. (2023). The Convergence of Deep Learning and Computer Vision: Smart City Applications and Research Challenges. Published in IEEE Access.)

[4] (Lee, J., et al. (2023). Smart City Architecture: Vision and Challenges. Published in IEEE Transactions on Industrial Informatics.)

[5] (Head, A., Tran, N., & Blumenstock, J. E. (2017). Can Human Development be Measured with Satellite Imagery? Published on arxiv.)

[6] (Abburi, S., & Golla, S. B. (2017). Satellite Image Classification Methods and Techniques: A Review. Published on IEEE Xplore.)

[7] (Law, H., & Deng, J. (2018). Corner Net: Detecting Objects as Paired KeyPoint's. Published in arxiv.)

[8] (Duan, K., et al. (2019). CenterNet: KeyPoint Triplets for Object Detection. Published in arXiv.)