Optimizing Healthcare Accessibility: An Integration of GIS and Machine Learning for Strategic Hospital Facility Location Planning

Kahlil Sebastian G. Ramos, Elmer C. Peramo
DOST–Advanced Science and Technology Institute, Quezon City 1101 Philippines
elmer@asti.dost.gov.ph

***Abstract*—Political and practical challenges impede access to primary healthcare in the Philippines, with approximately 50utilizes Geographic Information System (GIS) tools and artificial intelligence (AI) to thoroughly analyze hospital accessibility, focusing on Olongapo City, Zambales, Philipines. Using QGIS, current hospital distribution was mapped, and the spatial availability of healthcare services was evaluated through location-allocation analysis. Land Use Semantic Segmentation Models were developed to consider land suitability for establishing healthcare infrastructure. The methodology involves data preprocessing, shortest route analysis, and AI models for automatic land-use categorization. This research provides insights for optimizing resource allocation under the Universal Healthcare Act of 2019, aiming to enhance accessibility and potentially impact job availability and the national gross domestic product positively.**

***Index Terms*—artificial intelligence, geographic information system, hospital facility accessibility, optimization, semantic segmentation**

I. Introduction

A. The Philippine Healthcare System

In the Philippines, political and practical factors influence the selection of sites for primary care facilities, often leading to suboptimal placements of healthcare institutions [1]. Consequently, around 50% of the population needs primary care facilities within a 30-minute travel radius [2]. The Universal Healthcare (UHC) Act of 2019 allocates significant resources to ensure social equity in accessing essential services and address challenges in providing healthcare services. This republic act acknowledges the pivotal role of well-planned healthcare networks in improving public health and boosting employment opportunities and GDP [3].

This study analyzes hospital facility accessibility in Olongapo City, Zambales, Philippines. Leveraging Geographic Information System (GIS) technology and Artificial Intelligence (AI), the research aims to map the current distribution of hospital facilities to assess spatial healthcare availability. The study's primary objective is to propose strategic locations for new hospitals or expand existing facilities to bridge accessibility gaps.

This research seeks to provide data-driven insights and recommendations to enhance healthcare service accessibility in Olongapo City, aligning with the broader objectives of equitable access and resource optimization.

B. Facility Location Optimization

Facility location problems (FLPs) assume a fundamental role in optimizing the allocation of resources and services in various domains, including healthcare infrastructure planning. These problems involve determining the most strategic locations for facilities while considering specific objectives and constraints. Common optimization parameters in FLPs include minimizing average travel time or maximizing coverage within a predefined radius. Three widely used models in FLPs are the P-median problem, the location set covering problem (LSCP), and the maximal covering location problem (MCLP) [4].

The P-median problem focuses on identifying p facility locations to efficiently allocate demand, often associated with objectives such as minimizing travel distances [4]. This facility location model is used in the Nearest Facility Analysis, which aims to reduce the distance between users and the closest facility.

In contrast, the location set covering problem (LSCP) seeks to find the minimum number of facilities and their specific locations to ensure the coverage of all demand points [4]. LSCP is employed in city planning scenarios where the primary objective is to encompass all demand points within a service area.

The maximal covering location problem (MCLP) entails locating a predetermined number of facilities to maximize the coverage within a specified service distance or time, often applicable in situations where budget constraints limit the number of facilities that can be established [4].

In the context of hospital facility location optimization, Flores et al. have explored utilizing a Cooperative Maximal Covering Model that prioritizes maximizing the coverage of healthcare facility locations from a set of predefined candidate sites [1]. In their approach, the authors introduced the concept of distance decay to model the attractiveness of candidate facilities. This method encompasses two key metrics: the first metric considers the population residing within a 30-minute travel time, aligning with the objectives outlined in the Philippine Health Facility Development Plan (PHFDP), while the second metric incorporates an equation that factors in population, bed capacity, and travel time to estimate the expected number of visitors to a healthcare facility.

Furthermore, Flores et al. introduced two distinct demand assumptions: the zeroed demand assumption allocates zero demand to areas within a 30-minute drive of an existing rural health unit (RHU) since these regions are already adequately covered by an existing healthcare facility. In contrast, in the reduced demand assumption, demand is adjusted for areas already served by an existing RHU, accounting for the unit's capacity [1].

It was observed that both the zeroed demand and reduced demand methods identified candidate sites near existing RHUs, implying that these rural health units are likely insufficient in meeting the anticipated healthcare service demands.

Flores et al. recommended incorporating land use analysis through satellite imagery to assess a candidate location's suitability for healthcare facility establishment [1]. Hence, this research developed a land use semantic segmentation model to identify locations appropriate for healthcare infrastructure construction automatically.

II. Related Work

A. Integration of GIS in Healthcare Facility Optimization

The application of GIS in healthcare has been pivotal in spatial analysis and decision-making processes. In recent years, studies have emphasized the role of GIS in optimizing healthcare facility locations. Amoah-Nuamah et al. (2023) demonstrated using GIS to analyze the spatial distribution of healthcare services, identifying areas with inadequate healthcare coverage in rural communities in Ghana [5]. Their work highlighted the potential of GIS in revealing geographical disparities in healthcare access, thereby guiding policymakers in resource allocation decisions. A similar study by Das et al. (2021) extended this approach by incorporating population density data, elevation, and road network information into GIS models to identify optimal locations for new healthcare facilities [6]. Their research underscored the importance of considering demographic and geographical data in healthcare planning.

B. Machine Learning in Healthcare Facility Location Optimization

Machine learning (ML) has increasingly been adopted to augment decision-making in healthcare facility location optimization. A notable study by Vargas-Santiago et al in 2023 introduced two novel methods to enhance the effectiveness of conventional problem-solving techniques like heuristics, metaheuristics, and genetic algorithms [7]. Their first method involves leveraging collective intelligence through crowdsourcing, offering a unique solution to bridge the research void in utilizing crowdsourcing for Facility Location Problems (FLP). The second method integrates machine learning, specifically predictive modeling, to effectively navigate the solution space. The study reveals that machine learning not only augments existing problem-solving methods but also provides a more holistic approach to addressing FLP, filling previously identified gaps. The predictive capabilities of machine learning models proved to be instrumental in decision-making processes, offering rapid insights into system dynamics. Their research demonstrated how machine learning could be used to forecast future healthcare needs, informing the strategic placement of facilities. Furthermore, More et al delved into the potential of machine learning as a solution for facility location problems. They thoroughly analyzed different machine learning algorithms and techniques, assessing their performance based on accuracy, computational speed, and scalability. The findings from their study suggest that machine learning holds significant promise as an effective approach to these problems, especially when integrated with other optimization methods. Nevertheless, the study also highlighted the need for further investigation to comprehensively understand the constraints and capabilities of machine learning in addressing facility location issues [8]. The 2023 study of Salami et al. explored a comprehensive model for optimal placement and allocation of healthcare facilities over multiple periods [9]. This model uniquely considers both referral between services and equitable access. A practical example illustrates the model's effectiveness. Furthermore, they developed a hybrid Genetic Algorithm-Sequential Quadratic Programming (GA-SQP) technique to solve large-scale problems efficiently. The GA-SQP's performance was rigorously tested against various randomly generated scenarios and benchmarked with exact methods. The findings underscore the model's robustness in addressing healthcare facility location and service allocation while prioritizing service shortage and equity across different time frames.

C. Combining GIS and Machine Learning for Enhanced Healthcare Facility Optimization

Recent research has focused on the synergy of GIS and machine learning in healthcare optimization. In 2019, Kamel Boulos et al. presented a comprehensive survey of GeoAI technologies, encompassing a variety of methods, tools, and software. They explored the present and prospective uses of these technologies in numerous fields, including public health, precision medicine, and smart healthy cities enhanced by the Internet of Things [10]. Additionally, the paper briefly addressed the existing challenges that GeoAI faces in its research and application within the health and healthcare sectors. Moreover, Noon and Hankins' (2024) study presented the use of spatial data visualization to aid in determining the location and size of a new Neonatal Intensive Care Unit (NICU) within a network of rural hospitals. Utilizing a Geographic Information System (GIS), they analyzed both public and specific system data. This analysis revealed patterns in the healthcare system's market share and patient travel behavior, which were crucial in guiding decisions. Their application serves as an effective demonstration of employing spatial data mining as a vital component in the Knowledge Discovery in Databases (KDD) process [11].

D. Implications for Policy and Planning

The research in this domain has significant implications for healthcare policy and planning. These studies demonstrate that integrating GIS and machine learning provides a robust framework for optimizing healthcare facility locations. This approach not only enhances accessibility but also ensures efficient resource allocation, ultimately leading to improved healthcare outcomes.

III. Research Methodology

The research methodology flowchart follows a systematic approach to investigate hospital facility accessibility within the study area.

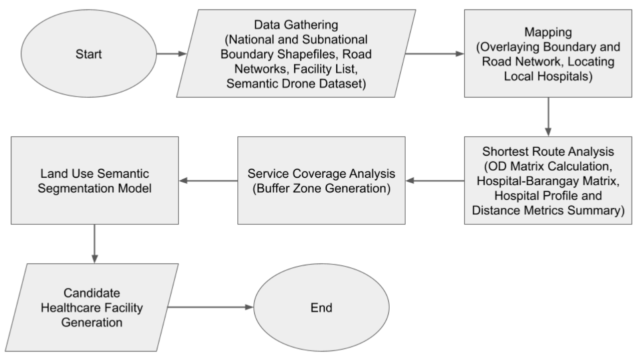


Fig. 1: Flowchart of the research methodology.

The flowchart begins with data gathering, which includes the acquisition of open-source national and subnational boundary shapefiles, road networks, Philippine population density, the national hospital facility list from the Department of Health - National Health Facility Registry (DOH - NHFR), and the semantic drone dataset provided by the Graz University of Technology.

The subsequent mapping phase involves overlaying national and subnational boundary shapefiles in QGIS and deriving a local road network using an intersection algorithm. Simultaneously, geolocation of local hospitals was achieved through a Python script utilizing the Google Geotagging API for obtaining latitude and longitude values.

Two assumptions were made to represent local population centers: the geometric centroid assumption and the population density centroid assumption. For the geometric centroid assumption, the local population center of each barangay was assumed to be the geometric centroid of a barangay. On the other hand, the population density centroid assumption determines local population centers by getting the mean coordinate of an aggregate of points representing a barangay's demand points (i.e., its residents).

After mapping, the shortest route analysis was performed to assess hospital utilization and determine the extent of coverage for Barangays. Through shortest route analysis, Hospital-Barangay Matrices were developed; these matrices report the distances between barangays and their servicing hospital and the utilization status of the local hospitals. To summarize the Hospital-Barangay Matrices, a Hospital Profile and Distance Metrics Summary were accomplished in order to classify which hospitals are underutilized, normally utilized, and heavily utilized.

The methodology proceeds to service coverage analysis, which was done by generating 5-kilometer and 10-kilometer radius buffer zones; these buffer zones highlight the spatial reach of healthcare facilities.

Finally, U-Net-based land use semantic segmentation models were developed to obtain detailed land use information for potential facility locations.

IV. Geospatial Mapping

The geospatial mapping in QGIS was initiated by defining the boundaries of Olongapo City using the Philippines - Subnational Administrative Boundary Shapefile, obtained from the Humanitarian Data Exchange. Subsequently, the national road network data from OpenStreetMap (OSM) was imported and processed through the QGIS Intersection Algorithm to extract the local road network of Olongapo City.

An updated list of healthcare facilities in the Philippines was obtained from the DOH - NHFR. After obtaining the updated facility list, the comma-separated value (.csv) file was filtered to retain only Zambales healthcare facilities. Automation of the geotagging process for Zambales healthcare facilities was accomplished by developing a Python script that accepts the facility list in .csv format and generates a corresponding .csv file containing facility names, longitude, and latitude values.

The resulting .csv file from the automated geotagging script was subsequently imported into QGIS as a delimited text layer. All layers were imported using EPSG:32651 - WGS 84 / UTM zone 51N as the coordinate reference system (CRS) since this CRS is suitable for the Philippines.

For the initial local population center assumption, the geometric centroid of each barangay within Olongapo was determined using the Centroid function in QGIS. For the population density centroid assumption, the Philippine population density raster data was obtained from the Humanitarian Data Exchange; this raster data was then converted into vector points to approximate demand point locations. These vector points were subsequently clustered per barangay, with the QGIS mean coordinate function employed to derive the population density centroid for each barangay.

Using the QGIS Network Analysis Toolbox (QNEAT3), Origin-Destination (OD) matrices were computed for both local population center assumptions, with centroids serving as Origin Points (O) and healthcare facilities as Destination Points (D). Additionally, QNEAT3 was used to generate shortest routes using the OD matrices and the localized road network as inputs.

V. Shortest Route Analysis

Two sets of shortest routes for the geometric centroid assumption and population density centroid assumption were generated using the QNEAT3 plugin.

Hospital-Barangay Matrices were developed for both centroid assumptions after completing the shortest route generation. These matrices report distances for every combination of barangays and servicing hospitals and the number of barangays each hospital serves.

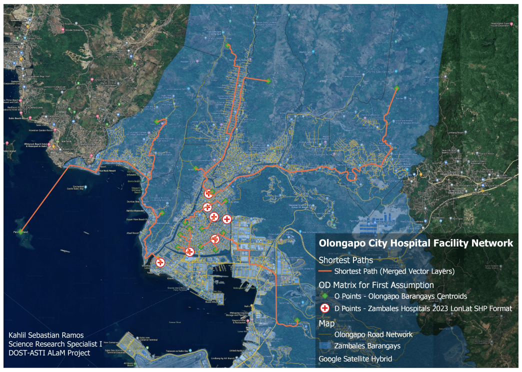


Fig. 2: Shortest route generation for the geometric centroid assumption.

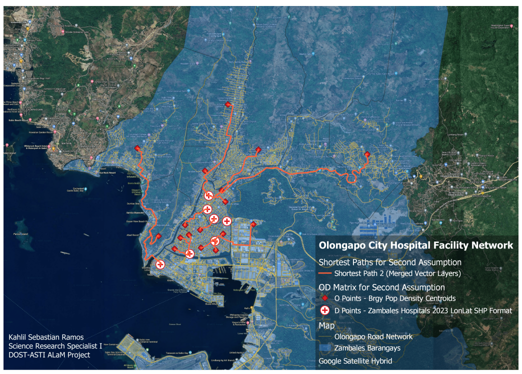


Fig. 3: Shortest route generation for the population density centroid assumption.

For the geometric centroid assumption, Alferos Hospital services no barangay, while the Zambales Medical Mission Group Coop Hospital services five barangays (see Appendix A, Subsection A). In this assumption, the maximum distance between a barangay and a hospital was 9.27 kilometers.

Similarly, Alferos Hospital serves no barangay, while Zambales Medical Mission Group Coop Hospital serves six barangays for the population density centroid assumption (see Appendix A, Subsection B). Under this assumption, the maximum distance between a barangay and a hospital was determined to be 6.39 kilometers.

Subsequently, Hospital Profile and Distance Metrics Summary Table was constructed to assess the utilization status of each hospital based on the number of barangays served, employing an arbitrary classification method. Hospitals serving none to one barangay were categorized as underutilized, hospitals serving two to three barangays were deemed sufficiently utilized, and hospitals serving more than four barangays were classified as heavily utilized. For both centroid assumptions, the mean and maximum distances in kilometers were computed as these distance metrics are relevant parameters in assessing hospital facility accessibility (See Appendix B, Subsection A).

The analysis revealed that the Geometric Centroid Assumption results in a more balanced distribution of Hospital-Barangay combinations than the Population Centroid Assumption (See Appendix B, Subsection A). Under the Geometric Centroid Assumption, two hospitals were classified as underutilized, while three were considered sufficiently utilized. In contrast, three hospitals were classified as underutilized, and two were classified as sufficiently utilized for the Population Density Centroid Assumption.

It is important to note that for barangays characterized by uneven population distributions, as is common in many Olongapo City barangays due to the geographic profile of some barangays, the Population Density Centroid Assumption aligns more closely with the actual mean coordinates of a local population center. Hence, it represents a more appropriate framework for evaluating utilization status and assessing hospital accessibility. Additionally, the distance metrics present relatively better accessibility from local population centers when population density, rather than geometric shape, was considered. The Population Density Centroid Assumption has mean and maximum distances of 1.71 and 6.39 kilometers, respectively (See Appendix B, Subsection B). In comparison, the Geometric Centroid Assumption resulted in mean and maximum distances of 2.37 and 9.27 kilometers, respectively. Table I presents a comparison of distance metrics, contrasting the mean and maximum distances for both the geometric and population density centroids.

TABLE I: Summary of Distance Metrics

Distance Metrics	Geometric Centroid	Pop. Density Centroid
Mean Distance	2.3655 km	1.7051 km
Max Distance	9.2686 km	6.3942 km

VI. Service Coverage Analysis

Local hospital coverage within Olongapo City was assessed by generating two distinct sets of buffer zones: 5-kilometer radius buffers and 10-kilometer radius buffers. These buffers were produced using the Buffer tool of QGIS. The total area enclosed by the buffer zones for each set was obtained through the Dissolve tool in QGIS.

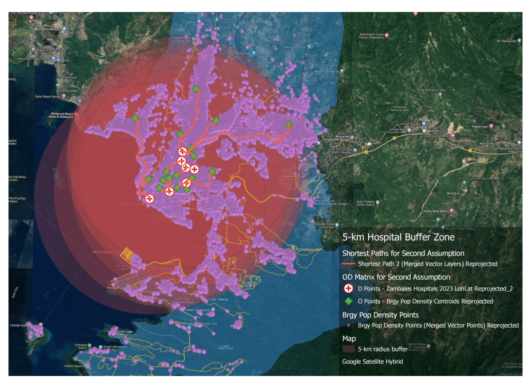


Fig. 4: 5-km hospital buffer zones.

Subsequently, the intersection between population density vector points and buffer zones was determined using the intersection algorithm. In this algorithm, the population density

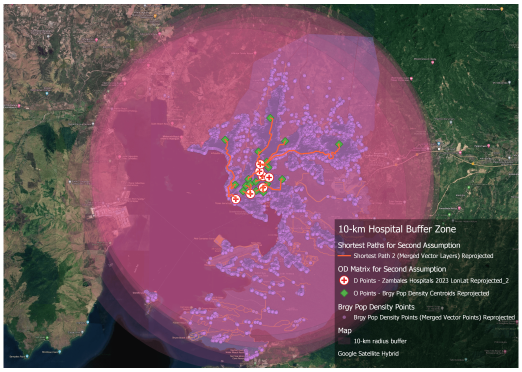


Fig. 5: 10-km hospital buffer zones.

vector points were specified as the input layer, while the generated hospital buffer zones served as the overlay layer.

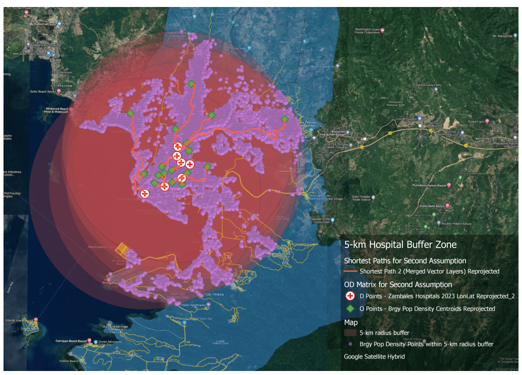


Fig. 6: Intersection between vector points and the 5-km buffer zones.

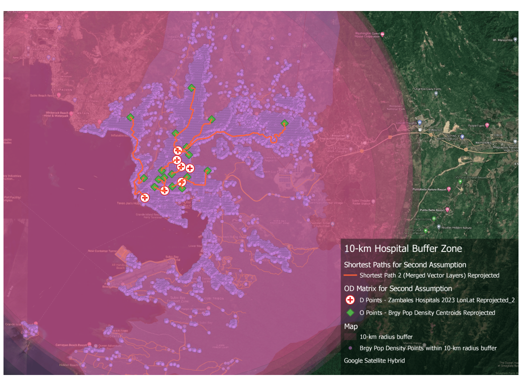


Fig. 7: Intersection between vector points and the 10-km buffer zones.

After employing the intersection algorithm, the number of population density vector points enclosed within both the 5-kilometer and 10-kilometer radius buffers was obtained from the attribute table. It was identified that 21,261 population density vector points were enclosed within the 5-kilometer radius buffers and 23,290 vector points within the 10-kilometer radius buffers. Since the total number of barangay population density vector points amounts to 23,290, it can be deduced that for the 10-kilometer radius buffers, all population density vector points were covered by the local hospitals.

The computation of hospital percent coverage was performed using (1). With the 5-kilometer radius buffer zones, the hospital percent coverage was calculated as 91.29%. On the other hand, for the 10-kilometer radius buffer zones, the hospital percent coverage was determined to be 100%. These findings signify that 91.29% of the Olongapo population resides within a 5-kilometer radius from a local hospital, while the entire population of Olongapo is considered covered within a 10-kilometer radius.

$$\%_{\text{covered}} = \frac{\text{vector points within } n\text{-km buffer}}{\text{total vector points}} \times 100\% \qquad (1)$$

VII. Land Use Semantic Segmentation

A total of six semantic segmentation models were developed using the Semantic Drone Dataset provided by Graz University of Technology [12]. This dataset comprises 23 classes, such as trees, grass, vegetation, gravel, dirt, water bodies, and paved areas, among others. The dataset has 400 images for both training and testing, each accompanied by corresponding segmentation masks.

TensorFlow and ONNX were employed as deep learning frameworks for the models. TensorFlow has a dedicated segmentation models library, which was utilized for this project. However, since most QGIS deep learning plugins exclusively accept ONNX models, tf2onnx was used to convert the Keras models into ONNX format.

The architecture selected for the models was U-Net, while the encoders used were ResNet-34, ResNeXt-50, and VGG-16. For each encoder, two models were developed: one with augmented data and a learning rate scheduler and one with only the original data and no learning rate scheduler.

The ResNet-34 U-Net model without data augmentation demonstrated the most acceptable performance among the developed models. It can be observed from the inference that this model captured general details and patterns from the test images. The model has a training loss of 18.37%, training accuracy of 94.16%, validation loss of 91.14%, and validation accuracy of 79.17%.

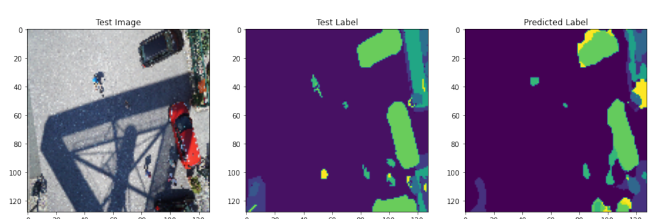
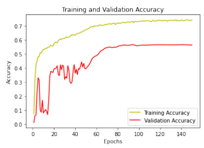
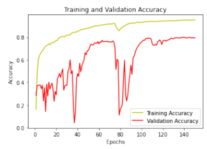


Fig. 8: Model inference for ResNet-34 U-Net model without data augmentation.

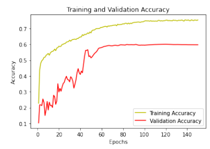
The remaining five models exhibited suboptimal performance, characterized by poor performance metrics.



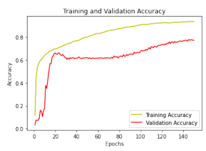
(a) ResNet-34 U-Net model with data augmentation.



(b) ResNeXt-50 U-Net model without data augmentation.



(c) ResNeXt-50 U-Net model with data augmentation.



(d) VGG-16 U-Net model without data augmentation.

(e) VGG-16 U-Net model with data augmentation.

Fig. 9: Performance metrics of five semantic segmentation models.

Model inference was conducted using the ResNet-34 U-Net Keras model on a satellite image of Olongapo City. While the results were not a precise representation of the test image, the model was able to extract generic image features such as the shape of the river.

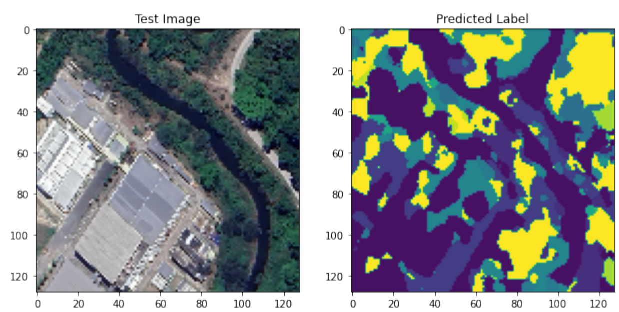


Fig. 10: Model inference for ResNet-34 U-Net model without data augmentation using a satellite image of Olongapo City, Zambales.

VIII. Conclusions

This study employed geospatial analysis to assess the accessibility of existing hospital facilities in Olongapo City, Zambales, Philippines. AI models were developed to obtain essential land use information for potential facility locations. The Population Density Centroid Assumption (PDCA) was adopted as the framework for evaluating hospital utilization and accessibility since it better represents local population centers. Utilizing the PDCA, hospital utilization status was evaluated: it was determined that three hospitals are underutilized, two are sufficiently utilized, and two are heavily utilized. The mean and maximum distances for PDCA were 1.71 and 6.39 kilometers, respectively. Service coverage analysis using 5-km and 10-km radius hospital buffer zones indicated high hospital percent coverage (91.29% and 100%, respectively), suggesting comprehensive city coverage by existing hospitals. Six U-Net land use semantic segmentation models were developed; among these models, the ResNet-34 U-Net model demonstrated the best performance. Training and validation metrics for this model include a training loss of 18.37%, training accuracy of 94.16%, validation loss of 91.14%, and validation accuracy of 79.17%. For future studies, the ONNX-converted Keras model may be utilized using QGIS in order to automatically segment different land uses and select locations that are suitable for healthcare infrastructure establishment.

References

[1] L. J. Y. Flores, R. R. Tonato, G. A. dela Paz, and V. G. Ulep, “Optimizing health facility location for universal health care: A case study from the Philippines,“ PLOS ONE, vol. 16, no. 9, p. e0256821, Sep. 2021, doi: https://doi.org/10.1371/journal.pone.0256821.

[2] Department of Health, Philippine Health Facility Development Plan 2020-2040. Manila, PH: Department of Health, 2020.

[3] Q. Peng, and H. Afshari, “Challenges and Solutions for Location of Healthcare Facilities,“ Industrial Engineering & Management, vol. 03, no. 02, 2014, doi: https://doi.org/10.4172/2169-0316.1000127.

[4] J. Yao, GIS-based Facility Location Analysis for the Public and Private Sectors [PDF document]. Available: https://www.ubdc.ac.uk/media/2133/facility-location-analysis-webinar-presentation-231020.pdf

[5] John Amoah-Nuamah, Williams Agyemang-Duah, Gmakikube Prosper Ninorb Borbor Gladstone Ekeme (2023) Analysis of Spatial Distribution of Health Care Facilities and its Effects on Access to Primary Healthcare in Rural Communities in Kpandai District, Ghana, Cogent Public Health, 10:1, DOI: 10.1080/27707571.2023.2183566

[6] Das M, Dutta B, Roy U, Das S, Rath S. Spatial accessibility modeling to healthcare facilities in the case of health shocks of Midnapore municipality, India. GeoJournal. 2023 Feb 11:1–24. doi: 10.1007/s10708-023-10838-1. Epub ahead of print. PMCID: PMC9919752.

[7] Vargas-Santiago, Mariano, Diana A. León-Velasco, Ricardo Marcelín Jiménez, and Luis Alberto Morales-Rosales. 2023. ”Complementing Solutions for Facility Location Optimization via Video Game Crowdsourcing and Machine Learning Approach” Applied Sciences 13, no. 8: 4884. https://doi.org/10.3390/app13084884

[8] S. K. More, L. R. Gupta, A. Gehlot, S. K, A. A. Al-Hilali and M. B. Alazzam, ”Exploring the Effectiveness of Machine Learning in Facility Location Problems,” 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2023, pp. 154-158, doi: 10.1109/ICACITE57410.2023.10182954.

[9] Salami A, Afshar-Nadjafi B, Amiri M. A Two-Stage Optimization Approach for Healthcare Facility Location- Allocation Problems With Service Delivering Based on Genetic Algorithm. Int J Public Health. 2023 Feb 28;68:1605015. doi: 10.3389/ijph.2023.1605015. PMID: 36926284; PMCID: PMC10011119.

[10] Kamel Boulos, M.N., Peng, G. VoPham, T. An overview of GeoAI applications in health and healthcare. Int J Health Geogr 18, 7 (2019). https://doi.org/10.1186/s12942-019-0171-2

[11] C. E. Noon and C. T. Hankins, ”Spatial data visualization in healthcare: supporting a facility location decision via GIS-based market analysis,” Proceedings of the 34th Annual Hawaii International Conference on System Sciences, Maui, HI, USA, 2001, pp. 10 pp.-, doi: 10.1109/HICSS.2001.926573.

[12] C. Mostegel, M. Maurer, N. Heran, J. P. Puerta, and F. Fraundorfer, “Semantic Drone Dataset,“ Graz University of Technology Institute of Computer Graphics and Vision, Jan. 25, 2019. https://www.tugraz.at/index.php?id=22387