¡Research Title¿

Theodore Utomo, Kometh, Aritro, Dev

March 31, 2025

1 Introduction

Understanding the location of healthcare centers with respect to the population that require its services is important for a multitude of reasons. Healthcare centers are a necessary piece of infrastructure that a healthy population needs to thrive. The need for healthcare centers often arise out of urgency and it is crucial to contextualize the time that a person in need of urgent healthcare requires to commute to these healthcare centers. We assembled a spatio-temporal database consisting of population density, settlement locations, healthcare centers and factors affecting travel in Kenya.

2 Relevant Past Papers

For each paper, consider including:

- One sentence summary of the paper.
- Gap/limitation of the paper.

The main focus of our experiment is to see how accessible it is for Kenyan people to access healthcare. Our idea is to map geospatial population data along with data for Kenyan healthcare centers to see the average or maximum time it would take for someone to commute to or from a healthcare center. Relevant research has been conducted in a similar manner to map the commute time to education centers such as primary schools. The techniques used take into account the nuances of travel time such as walking, traffic congestion, difficult roads, etc. (see `https://www.tandfonline.com/doi/full/10.1080/14733285.2022.2137388#d1e808`).

3 Motivation

- **Limitation or problem:** Patient experience in accessing healthcare is critical, yet many developing countries, including Kenya, face infrastructural limitations that make accessing healthcare challenging.
- **Importance:** Adequate access to healthcare is crucial for improving patient outcomes, particularly in low-income and rural areas.
- **Proposed solution:** Our idea uses existing population, settlement, and healthcare center data, combined with travel time nuances, to identify accessibility issues and suggest optimal areas for developing new healthcare infrastructure.

- **Why it should work:** By integrating multiple geospatial datasets and modeling realistic travel conditions, our approach should offer a more accurate picture of healthcare accessibility.

Healthcare is an essential piece of infrastructure in a developing country. Patient experience is a crucial aspect of healthcare that faces many limitations within a developing country. Poverty is a reality that many people face in developing countries and low-income communities often find it difficult to find adequate access to healthcare resources. Kenya is a rapidly growing country with its population and economy seeing steady growth in recent years (find source), but due to its developing infrastructure, it is difficult to assess access to healthcare. Kenya's population also includes nomadic pastoralists who travel from settlement to settlement, which makes it difficult to target optimal areas for infrastructure development. However, by identifying potential settlement areas, we can create estimates for future infrastructure improvements.

4 Key Contributions/Ideas

- Integration of spatial population data, settlement data, and healthcare center data.
- Incorporation of factors affecting travel time (e.g., road quality, congestion, age of commuters).
- Development of an accessibility model that suggests optimal locations for new healthcare infrastructure.

5 Methods

5.1 Datasets

We plan to gather data for three primary components:

- **Spatial Population/Population Density:** We obtained our estimated population density map through GRID3. This dataset is a heat map of Kenya with data collected at a spatial resolution of 3 arc-seconds, providing population estimates for each 100m x 100m grid cell. Metadata, including age, is provided to account for varying travel speeds.
- **Healthcare Infrastructure:** Our database of healthcare facilities comes from OpenStreetMap's Kenya Health Facilities dataset, which includes details on location, amenity type, service type, capacity, etc.
- **Factors Affecting Travel:** Additional datasets will be collected to account for factors such as road networks (available via OpenStreetMap shapefiles), traffic congestion, transportation methods, and the age of commuters.

5.2 Travel Time Modeling

Once all input layers are gathered, we will use AccessMod's merging tools to combine the data and conduct an analysis of health coverage. Our model will compute the average travel time to the nearest healthcare center while considering various transportation challenges.

5.3 AccessMod Analysis Features

After computing the travel times, we can use AccessMod's several analytical features including a scaling up feature that identifies optimum locations for constructing new healthcare centers.

5.4 CNN Model For Identifying Optimal Types Of Health Care Facilities

Once we have AccessMod's suggestions for optimum locations to create new health care facilities, we can leverage deep learning models to identify the optimal types of health care facilities to construct. We can execute this by fine tuning existing CNN models to identify already existing health care infrastructures to allow it to identify optimal existing health care structures in relation to the overall population. We can then use unsupervised learning methods to identify the types of health care infrastructure that will be optimal to build in the locations suggested by AccessMod.

6 Evaluation and Metrics

To assess the effectiveness of our proposed method, we will evaluate the improvements in healthcare accessibility using several quantitative metrics. Our evaluation metrics include:

- **Commute Time Reduction:** Measure the reduction in both average and maximum travel times to the nearest healthcare facility after the implementation of our suggested infrastructure improvements.
- **Accessibility Improvement:** Compare the healthcare accessibility scores (e.g., coverage indices) before and after the proposed changes against established benchmarks.
- **Model Performance Metrics:** For the CNN component that identifies optimal types of healthcare facilities, we will use standard classification metrics such as accuracy, precision, recall, and F1-score on a validation set, in addition to confusion matrices to evaluate class-wise performance.
- **Ablation Studies:** Systematically disable or modify parts of our model (e.g., using or excluding specific travel factors) to isolate and quantify the impact of each component on the overall system performance.

In addition to these metrics, we will perform sensitivity analyses to assess how changes in input data quality and travel time modeling assumptions affect the overall results.

7 Experimental Setup

Our experimental setup is divided into several stages to systematically evaluate and validate the proposed methodology:

1. **Data Collection and Preprocessing:**
    - **Population Data:** Acquire high-resolution population density maps from GRID3, ensuring that spatial resolution and demographic metadata (e.g., age distribution) are retained.
    - **Healthcare Facilities:** Gather healthcare facility data from OpenStreetMap's Kenya Health Facilities dataset, including geolocations and facility attributes.
    - **Travel Factors:** Collect additional geospatial layers (e.g., road networks from OpenStreetMap shapefiles, traffic congestion data, and public transit routes) to accurately model travel times.

- **Preprocessing:** Normalize and georeference all datasets so they can be integrated into a unified spatial framework.

2. **Geospatial Analysis using AccessMod:**
   - Utilize AccessMod to compute baseline travel times to existing healthcare facilities, accounting for terrain, road quality, and other travel-affecting factors.
   - Leverage AccessMod's scaling and optimization features to generate initial suggestions for new healthcare facility locations based on minimizing travel times.

3. **CNN Model Training for Facility Type Identification:**
   - **Dataset Preparation:** Assemble a labeled dataset of existing healthcare facilities (with various types such as clinics, hospitals, etc.) using the collected data.
   - **Model Architecture:** Fine-tune a pre-trained CNN to classify healthcare facility types, adapting the network to our specific dataset and task.
   - **Training:** Train the CNN using standard backpropagation and optimization techniques (e.g., Adam optimizer), with a portion of the data reserved for validation.
   - **Unsupervised Analysis:** Apply unsupervised learning methods (e.g., clustering) on the features extracted by the CNN to determine the optimal facility types that should be constructed in the proposed new locations.

4. **Integration and End-to-End Evaluation:**
   - Combine the results from the AccessMod analysis and the CNN model to propose an integrated plan for healthcare infrastructure improvement.
   - Evaluate the proposed plan using the previously defined metrics, and perform ablation studies to determine the contribution of each component.
   - Visualize the proposed improvements using geospatial maps and statistical plots to clearly demonstrate accessibility gains.

5. **Benchmarking and Validation:**
   - Compare our method against existing benchmarks and alternative approaches in the literature to validate the effectiveness of our approach.
   - Conduct cross-validation and sensitivity analysis to ensure that the improvements are robust to variations in the input data and model parameters.

7.1 Visualizations and Statistics

Visual representations of population density, healthcare facility locations, and computed travel times will be used to illustrate findings. Statistical analysis will compare our model's suggested improvements with current accessibility metrics.

8 Potential Limitations

While our approach integrates multiple datasets and realistic travel factors, potential limitations include:

- Incomplete or outdated datasets.
- Challenges in accurately modeling the nuances of real-world travel conditions.
- Variability in data quality across different regions.

9 Conclusion

References