Poverty, Environment, and Neglected Tropical Diseases in Senegal: A Computational Sustainability Study

Kamwoo Lee, PhD candidate in Systems and Information Engineering
(Advisor: Dr. William Scherer in Engineering Systems and Environment)

Project Overview

Neglected tropical diseases (NTDs) are a collection of infectious diseases caused by a variety of pathogens that are usually found in tropical and subtropical regions (WHO, 2017). Although these diseases have historically received far less attention than other major killers such as malaria, HIV/AIDS, and tuberculosis, they pose significant social and economic challenges such as blindness, disfigurement, and developmental impairment, and they disproportionately affect poor and marginalized populations who live in unsanitary conditions (Hopkins, 2016).

While some mapping of the geographic presence of NTDs in sub-Saharan Africa is currently underway, increased coordination across multiple fields is necessary in order to completely map NTDs. Predicting the distribution of NTDs requires understanding how key elements of poverty, pathogens, human behavior, and the environment interact with each other to facilitate NTD presence in communities. Data on these factors is crucial, and information such as where the disease is prevalent geographically and which vectors carry the disease is incredibly informative to any mapping project. However, such high-level information, which can be found in national statistics and clinical records are exorbitantly expensive, if not impossible, to acquire in lower-income countries.

This study aims to map NTD distribution in Senegal and West African countries through satellite remote sensing so that we can identify vulnerable populations both regionally and socially. I propose a method that combines two engineering techniques — machine learning and simulation. I believe this method is an effective and efficient way to tackle the biggest obstacles to reaching the goal – lack of data and complexity of the diseases.

This project requires qualitative field research in Senegal for constructing a poverty index, choosing environment variables, and designing the causal links for the simulation, which is why I apply for iGrant of the Global Infectious Disease Institute (GIDI). If selected by GIDI, I attest that I will not receive funding from other sources to support this research trip and I will be happily present this search project at GIDI events as requested. Further, I will acknowledge GIDI for support in any of my scholarly products arising from this project as well as allow GIDI to publicize my project.

Project Description

With the recent advancement of machine learning techniques, we can now extract high-level information from unstructured data such as images and texts, which are abundant, even in lower-income countries. For instance, many researchers have used machine-learning methods to convert satellite images into accurate estimates of household consumption and assets (Jean et al., 2016; Pandey et al., 2017). I seek to take these machine-learning methods an additional step further. Since poverty and poor environment and sanitation factors influence the prevalence of NTDs, I aim to estimate these factors at the village level using satellite images.

Moreover, I plan to incorporate the machine learning-supported information into an agent-based model (ABM), which can embrace the complexity of the issue and can model key characteristics of NTDs. Predicting NTD prevalence with ABM at a local scale has shown promising results for future NTD mitigation interventions (Alderton et al., 2016). I hope that our unstructured data-driven simulation method can scale up NTD prediction for vulnerable populations and can provide a more efficient approach.

1) Methodology

I will utilize the Demographic and Health Surveys (DHS) to study poverty and health issues with a primary focus on Senegal, as Senegal's data is particularly useful to this study. Senegal's DHS has GPS coordinate for each respondent, enabling us to link information on the ground with satellite images. Since the DHS has rich information about household assets and health status, I can utilize it for supervised learning with satellite images. Although deep learning models such as convolutional neural networks (CNN) could in principle be trained to directly estimate NTD prevalence from satellite imagery, the structure of deep learning prohibits discovering relationships between variables, which is necessary to implement policy interventions.

Thus, I plan to train CNN for poverty and environmental conditions separately to extract a poverty index and environment variables from the high-dimensional satellite images. I will then construct a multi-host agent-based model to set up causal links from poverty and environmental variables to NTDs through modeling disease transmission in natural and changing ecologies. ABM can be an especially effective tool for modeling the transmission of pathogens for vector-borne, infectious diseases because it can extract a complex aggregate phenomenon from interactions between multiple entities in question.

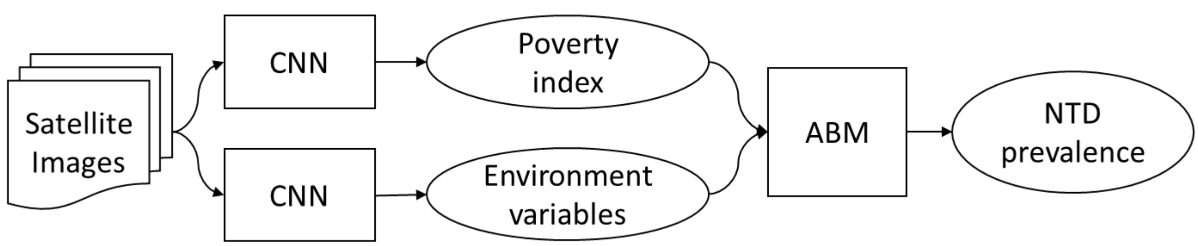

Figure: Illustration of the pipeline for the proposed method

I will first pre-train the two CNN modules with general satellite images for object detection and classification with ImageNet before training for the poverty index and environment variables with the DHS Senegal data. For the ABM, I define three types of agents: human, domestic animals, and pathogens for each disease, and then model their interactions with the environment and between agents. I consider utilizing existing ABM software dedicated to disease modeling such as Epidemiological MODeling software (EMOD, Institute for Disease Modeling).

Since this approach incorporates two layers of modeling abstraction, CNN and ABM, it requires validation for each model. For validating the generated poverty index, I will utilize the Poverty Monitoring Survey conducted by Agence Nationale de la Statistique et de la Démographie (ANSD), the national statistics agency. With the help of the World Bank Senegal office, I have already acquired the survey data, which I will use to aggregate consumption at the household level.

For the environmental model, I will further investigate relevant factors that impact NTD prevalence. If the iGrant funding permitted, I will gather validation data through field research, which I plan to conduct through interviews and focus groups in Senegal during summer 2019. I will conduct the field research in partnership with our in-country mentor Professor Aly Mbaye who is the Dean of the Centre de Recherches Économiques Appliquées (Center of Applied Economic Research) at the Université Cheikh Anta Diop of Dakar (UCAD) in Dakar, Senegal.

Currently, no comprehensive map of NTD prevalence in sub-Saharan Africa exists. As I generate our own NTD distribution map, I will validate our model with smaller datasets such as the Global Neglected Tropical Disease Database, the Global Atlas of Helminth Infections and the Global Atlas of Trachoma. These open datasets are geographically limited, but will still be useful for validation because they are manually collected and contain relatively accurate information on disease prevalence in Senegal and other sub-Saharan countries.

2) Expected Outcome

Most directly, this study will map NTD prevalence throughout Senegal and will help to identify populations which are particularly vulnerable to NTDs. These insights will aid policymakers who seek to reduce the high rate of NTDs in the country, as effective mapping will enable these policymakers to more efficiently direct limited resources into interventions such as vector controls, improved access to clean water and improved sanitation facilities. However, simply identifying vulnerable populations might not be enough to achieve significant reductions in NTDs. Various policy alternatives should be explored and evaluated for their effectiveness prior to implementation. Our agent-based model could serve as a testbed for potential policy options to support a particular program before a costly policy intervention is pursued.

In addition, there are broader implications of this project for the intersection of the fields of international development and data science. By providing insights into identifying NTD-vulnerable population groups, I will further explore how mathematical modeling can inform policymakers in an international development context. Mathematical modeling of NTD transmission systems can have a significant impact on intervention strategies and policy formulation. I expect that this trained model can be used in other West African countries which share many of the same poverty and environmental issues but do not have enough data to build a separate model. Additionally, the combination of CNN and ABM has never been applied in this domain, and I hope this study will extend the boundary of the computational sustainability research field.

References

Alderton, S., Macleod, E. T., Anderson, N. E., Schaten, K., Kuleszo, J., Simuunza, M., ... & Atkinson, P. M. (2016). A multi-host agent-based model for a zoonotic, vector-borne disease. A case study on trypanosomiasis in Eastern Province, Zambia. PLoS neglected tropical diseases, 10(12).

Global Atlas of Helminth Infections. (n.d.). http://www.thiswormyworld.org/maps. Accessed February 24, 2019.

Global Atlas of Trachoma. (n.d.). http://www.trachomaatlas.org/. Accessed February 24, 2019.

Global Neglected Tropical Diseases Database. (n.d.). https://www.gntd.org/. Accessed February 24, 2019.

Hopkins, A. D. (2016). Neglected tropical diseases in Africa: a new paradigm. International health, 8 (suppl_1), i28-i33.

Hotez, P. J., & Kamath, A. (2009). Neglected tropical diseases in sub-Saharan Africa: review of their prevalence, distribution, and disease burden. PLoS neglected tropical diseases, 3(8).

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. Science, 353(6301), 790-794.

Pandey, S. M., Agarwal, T., & Krishnan, N. C. (2018). Multi-Task Deep Learning for Predicting Poverty from Satellite Images, 6.

Uniting to Combat NTDs. (2018). Africa and neglected tropical diseases. Retrieved February 24, 2019 from https://unitingtocombatntds.org/africa/.

World Health Organization (2017). Integrating tropical disease in global health and development. Fourth WHO Report on Neglected Tropical Diseases.