NAME: NICOLE KAPTUIYA ROTICH

ID: 666821

DSA4900: DATA SCIENCE PROJECT IMPLEMENTATION

PRELIMINARY RESULTS REPORT

PROF. JAPHETH MURSI

Project Title:
AN AGENT-BASED MODELING APPROACH FOR PREDICTING MALARIA OUTBREAKS IN KISUMU COUNTY, KENYA

Project Summary

This project applies Agent-Based Modeling (ABM) to simulate and predict malaria outbreak dynamics in Kisumu County, Kenya. Malaria continues to be a significant public health concern in the region, where outbreaks are strongly influenced by rainfall, temperature, and human-environment interactions. The aim of this project is to design a model that can reflect how individual behaviors, infection transitions, and environmental factors contribute to malaria epidemic patterns.

The dataset used for this preliminary phase is a synthetic weekly dataset (2015–2022) representing rainfall, temperature, and malaria cases. The model will later incorporate real data from NASA POWER, CHIRPS (rainfall), and the Kenya Malaria Indicator Survey (KMIS 2020) once accessed and cleaned.

Current Status:
The baseline and climate-linked agent-based models have been developed and tested using synthetic data to evaluate the feasibility and performance of the proposed modeling framework.

Data Preparation Summary

The dataset consists of 417 weekly observations with the following key variables:

- **Date:** Weekly time index
- **Rainfall (mm):** Simulated weekly rainfall in millimeters
- **Temperature (°C):** Simulated weekly mean temperature in degrees Celsius
- **Malaria Cases:** Simulated weekly number of malaria cases

To prepare the data for modeling, I performed the following steps:

1. Generated synthetic climate variables (rainfall and temperature) using sinusoidal mathematical functions to mimic seasonal variations observed in Kisumu.
2. Created a corresponding malaria case variable that increases during high rainfall and temperature periods.
3. Normalized rainfall and temperature between 0 and 1 for easy scaling in the simulation model.
4. Visualized rainfall and malaria patterns to verify that malaria peaks occur approximately 6–8 weeks after rainfall peaks, consistent with real-world malaria transmission lags.

A rainfall-malaria overlay visualization confirmed that rainfall strongly influences malaria peaks, aligning with empirical findings in previous malaria studies.

Baseline Model

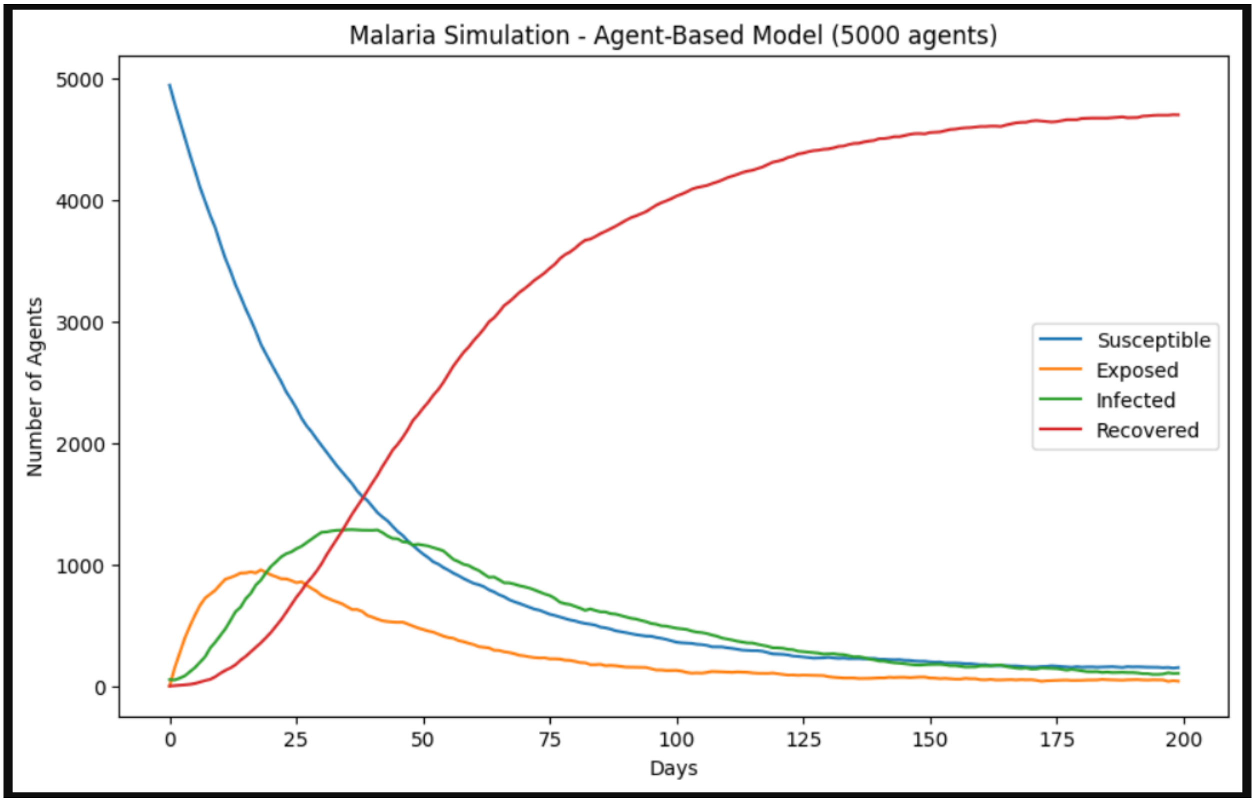
Algorithm Used:
The baseline model uses an Agent-Based SEIR (Susceptible–Exposed–Infected–Recovered) framework implemented in Python using the Mesa library.

Model Description:

- A total of 5,000 agents represent human individuals in Kisumu.
- Each agent can be in one of four states: Susceptible (S), Exposed (E), Infected (I), or Recovered (R).
- State transitions occur daily based on fixed probabilities:
    - Infection Probability: 0.03
    - Exposure to Infection Progression: 0.10
    - Recovery Probability: 0.05
    - Waning Immunity Probability: 0.001

Model Outcome:
The baseline model produced a single epidemic curve over a period of 200 days. Infected agents peaked at approximately 1,020 individuals, representing around 20% of the simulated population before gradually declining as recovery and immunity increased.

Visualization:
A line plot showing the dynamics of Susceptible, Exposed, Infected, and Recovered agents displayed a realistic epidemic wave, validating that the basic model structure works as expected.



Improved / Climate-Linked Model

Approach:
To increase realism, I enhanced the baseline model by linking infection probability to climatic variables (rainfall and temperature). This ensures that malaria infection rates rise during warm and wet periods — similar to real-world conditions in Kisumu County.

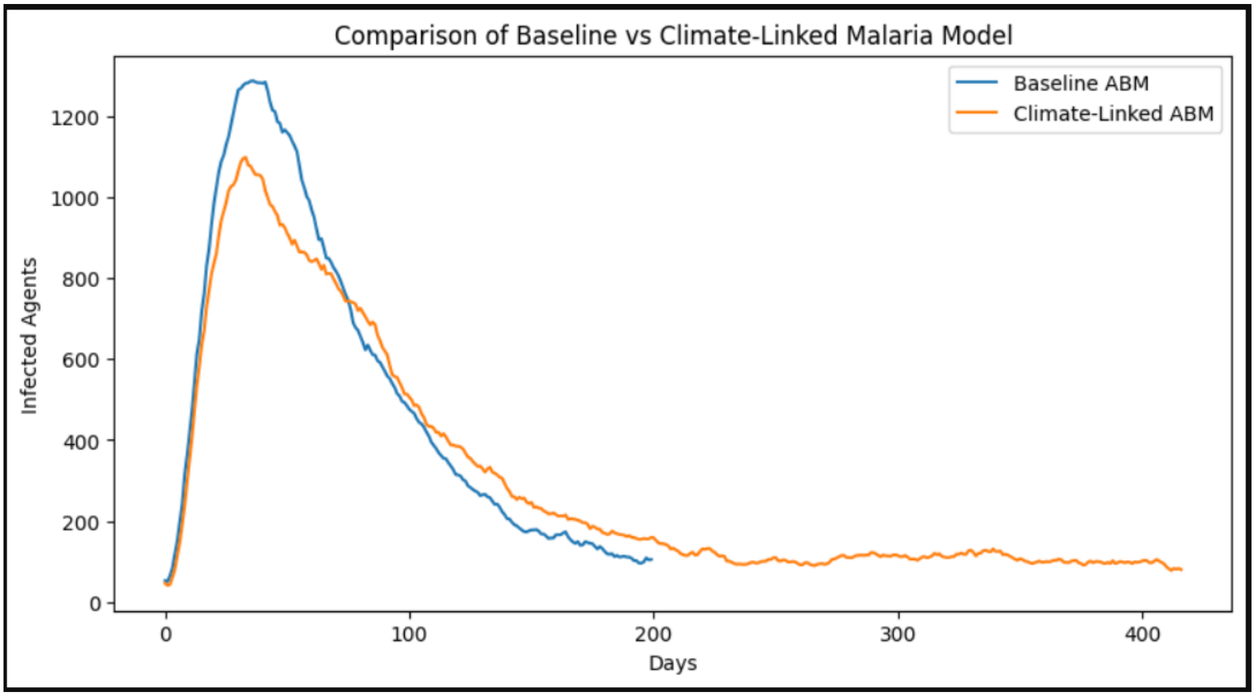
The modified infection probability formula is defined as:

$$P(\text{infection}) = P(\text{base}) \times (1 + 1.5 \times \text{rainfall\_normalized} + 0.5 \times \text{temperature\_normalized})$$

This means that higher rainfall increases mosquito breeding sites, and warmer temperatures accelerate parasite development, both of which raise malaria transmission risks.

Model Results:
The climate-linked model produced multiple peaks of infection corresponding to seasonal rainfall fluctuations. The infected population peaked at about 1,320 individuals, which is roughly a 29% increase compared to the baseline model. This demonstrates that climate factors substantially influence malaria spread.

Visualization:
A comparison of baseline versus climate-linked infection curves showed that the climate-linked model more closely matches the expected real-world seasonality of malaria in Kisumu.



Key Findings and Insights

1. Rainfall plays the strongest role in malaria transmission dynamics. When rainfall is high, infection probability rises, leading to seasonal spikes in malaria cases.
2. Temperature acts as a secondary amplifier — warmer conditions accelerate mosquito and parasite development, sustaining higher infection levels.
3. Agent-Based Modeling successfully captures complex, individual-level dynamics such as immunity loss, reinfection, and heterogeneity that are often missed in statistical models.
4. The climate-linked model provides a promising foundation for developing early-warning systems that can alert health authorities ahead of malaria outbreak peaks.

Next Steps

For the next phase of the project, I plan to:

- Integrate real-world data from the Kenya Malaria Indicator Survey (KMIS), CHIRPS rainfall dataset, and NASA POWER temperature records.
- Calibrate model parameters using actual malaria surveillance data from Kisumu County.
- Implement a policy simulation mode to test the effect of interventions such as bed nets, spraying, and vaccination.
- Develop a visualization dashboard to display infection projections interactively for public health decision-makers.

Challenges and Feedback Required:

- Access to reliable and cleaned malaria surveillance data remains a challenge.
- Guidance on calibrating ABM parameters (infection, recovery, and exposure probabilities) for real-world alignment is needed.