Sumaiya Sultana

Mustafa Ahmed AFEEF

Ayan Hassan Olad

Stacy-Kelvin Adaisky Irutingabo

**EpiGuard: Malaria Early Warning**

**1.Project Idea:**

Our project is aligned with the Sustainable Development Goals (SDGs), specifically focusing on SDG 3: Good Health and Well-being. The central aim of our initiative is to combat malaria, a significant global health challenge that disproportionately affects the most vulnerable populations and contributes to poverty and inequality.

**Specific Goal of the Project:**

The specific goal of our project is to develop a predictive model for malaria outbreaks, leveraging advanced data analysis and machine learning/deep learning techniques. By accurately forecasting potential outbreaks, our model aims to enable governments and public health organizations to implement early intervention measures. These measures include mobilizing resources, optimizing the distribution of medical supplies, and targeting vector control efforts more effectively. Ultimately, our project seeks to reduce the incidence and impact of malaria, contributing to the improvement of health outcomes in affected regions and supporting the broader objective of enhancing global health and well-being.

**2. Relevance to Sustainable Development Goals (SDGs):**

Our project directly aligns with the United Nations Sustainable Development Goal (SDG) 3: Good Health and Well-being. By focusing on predicting and mitigating malaria outbreaks, our initiative contributes significantly to reducing the global health burden posed by malaria, a major disease that impacts millions annually, particularly in low-income regions. This effort is crucial for improving health outcomes and enhancing the quality of life, especially among vulnerable populations.

Additionally, our project indirectly supports several other SDGs:

- SDG 1: No Poverty

- SDG 4: Quality Education

- SDG 6: Clean Water and Sanitation

- SDG 13: Climate Action.

**3.Literature that are similar to our project**

* **Prediction of malaria incidence using climate variability and machine learning**

A machine learning model predicts malaria outbreaks in six Sub-Saharan African countries using climate data over 28 years. It identifies key climate factors affecting malaria, outperforms other models, and aids early detection for public health decisions. This research helps prepare for future outbreaks and informs government responses to climate-driven disease transmission.

* **Predicting malaria outbreak using Machine Learning and Deep Learning approach: A review and analysis**

Authors used machine learning and deep learning to predict malaria outbreaks in Visakhapatnam, India, using six years of meteorological and malaria case data. Various algorithms were tested, with XGBoost performing best at 96.26% accuracy and 93.89% recall. Weather forecasts could aid outbreak prediction, helping prevent malaria-related deaths. The study suggests atmospheric factors and historical cases effectively predict outbreaks, highlighting XGBoost's efficiency. Future work may focus on refining precision and enhancing predictive capability by combining algorithms and incorporating additional atmospheric factors.

**4.Description of our data**

For our project aimed at predicting malaria outbreaks, we plan to use a comprehensive and diverse set of data sources. Here's a description of the data we will be utilizing:

**Reported Malaria Cases and Deaths:**

Source: World Health Organization (WHO)

Format: CSV or Excel spreadsheets

Size: Varies based on temporal and geographical coverage, potentially ranging from thousands to millions of records.

Preprocessing: Includes data cleaning, handling missing values, aggregating data to the desired temporal resolution, and normalizing case counts.

1. **Population Data:**

* Source: The World Bank
* Format: CSV
* Size: Varies based on the granularity of population data.
* Preprocessing: Involves standardizing population data across different regions, handling missing values, and aligning the data with other datasets.

1. **Economic Indicators:**

* Source: OurWorldInData.org
* Format: CSV or Excel files
* Size: Varies based on the number of indicators and geographical coverage.
* Preprocessing: Requires standardizing indicators, handling missing values, and aggregating data to match other datasets.

1. **Environmental Data (Temperature, Rainfall, Humidity):**

* Source: APIs like OpenWeatherMap
* Format: JSON or CSV
* Size: Varies based on the temporal and spatial resolution of the data requested.
* Preprocessing: Entails data normalization, handling missing values, and aligning the data with other datasets used in the project.

1. **Total Density of Hospitals per 100,000 Population:**

* Source: WHO
* Format: Likely CSV or Excel files, based on the standard data dissemination formats of WHO.
* Size: The dataset size is not explicitly mentioned but can be inferred to cover a comprehensive range of countries, given WHO's global scope.
* Preprocessing: Involves standardizing the hospital density data across different regions, handling any missing values, and aligning the data with other datasets to ensure consistency and accuracy in the analysis.

These datasets will provide a robust foundation for our analysis. The preprocessing steps outlined are essential for ensuring the data is clean, consistent, and ready for analysis. With these datasets, we aim to develop a predictive model that can accurately forecast malaria outbreaks by understanding the interplay between healthcare infrastructure, population dynamics, environmental conditions, and socio-economic factors.

**5.Approach (Machine Learning and Deep Learning)**

For our project on predicting malaria outbreaks, we will primarily use a machine learning approach due to its effectiveness with structured data, interpretability, and flexibility in handling diverse data sources. Machine learning models require less computational power and are more practical given the potential limitations in data volume and quality.

However, we will also incorporate deep learning techniques for specific tasks where they offer advantages, such as analyzing unstructured environmental data, identifying complex patterns, and modeling the temporal dynamics of malaria outbreaks. This combined approach allows us to leverage the strengths of both machine learning and deep learning to develop a robust and efficient predictive model.