Lab 5: Mapping Malaria Risk Using AI-Based Clustering in Google Earth Engine with ChatGPT

🧠 Goal of the Lab

In this lab, you will use environmental data (NDVI and rainfall) to identify areas with similar conditions that could support malaria transmission. You'll perform **unsupervised classification** (specifically, **K-means clustering**) using **Google Earth Engine (GEE)** and use **ChatGPT** to assist with writing, understanding, and modifying code. You'll also learn how to export the resulting malaria risk map for visualization in **QGIS**.

This lab builds directly on **Lab 3 (NDVI and Rainfall visualization)** and **Lab 4 (ChatGPT-assisted coding)**.

🔧 Tools and Requirements

- **Google Earth Engine (GEE)**: for accessing and analyzing NDVI and rainfall data
- **ChatGPT**: for generating and modifying GEE code
- **QGIS**: for visualizing the exported cluster map (optional but recommended)
- **Datasets**:
  - **NDVI**: `MODIS/006/MOD13Q1`
  - **Rainfall**: `UCSB-CHG/CHIRPS/DAILY`
  - **Boundary**: `USDOS/LSIB_SIMPLE/2017` (Uganda)

🧭 Part 1: Use ChatGPT to Generate a Clustering Script

Step 1: Open ChatGPT

1. Go to https://chat.openai.com/
2. Start a new chat session.

Step 2: Type the Following Prompt

"Write a GEE script to perform k-means clustering on stacked NDVI and rainfall data for Uganda to map malaria risk zones."

Step 3: Review ChatGPT's Output

It should generate something like:

```
// Load NDVI and rainfall data
var ndvi = ee.ImageCollection('MODIS/006/MOD13Q1')
  .filterDate('2022-01-01', '2022-12-31')
  .select('NDVI')
  .mean();

var rainfall = ee.ImageCollection('UCSB-CHG/CHIRPS/DAILY')
  .filterDate('2022-01-01', '2022-12-31')
  .sum();

// Load Uganda boundary
var uganda = ee.FeatureCollection('USDOS/LSIB_SIMPLE/2017')
  .filter(ee.Filter.eq('country_na', 'Uganda'));

// Stack the layers
var input = ndvi.addBands(rainfall).clip(uganda);

// Sample the data
var training = input.sample({
  region: uganda,
  scale: 5000,
  numPixels: 5000
});

// Perform k-means clustering
var clusterer = ee.Clusterer.wekaKMeans(3).train(training);
var result = input.cluster(clusterer);

// Visualize the clusters
Map.centerObject(uganda, 6);
Map.addLayer(result.randomVisualizer(), {}, 'Malaria Risk Clusters');
```

Step 4: Copy and Paste into GEE

1. Open https://code.earthengine.google.com/
2. Create a new script: `Lab5_Clustering_Malaria.js`
3. Paste the code into the editor.
4. Click **Run** to visualize clusters.

🔁 Part 2: Modify and Explore Using ChatGPT

🔄 Example Prompts

Copy-paste these into ChatGPT:

- “Change the number of clusters to 5 and re-run the script.”
- “How can I export the malaria risk map to Google Drive?”
- “Can I mask a specific cluster to highlight high-risk areas only?”

✅ What You Learn:

- How to adjust the number of clusters (change `.wekaKMeans(3)` to `.wekaKMeans(5)`)
- How to use `Export.image.toDrive()` to export data
- How to filter or mask raster values for specific cluster classes

---

💾 Part 3: Export Risk Zones to Google Drive

Sample Prompt:

"Help me export the malaria cluster map from GEE as a GeoTIFF to Google Drive."

Sample Code from ChatGPT:

```
Export.image.toDrive({
  image: result,
  description: 'Malaria_Risk_Clusters',
  folder: 'GEE_exports',
  fileNamePrefix: 'malaria_risk_uganda',
  region: uganda.geometry(),
  scale: 5000,
  crs: 'EPSG:4326',
  maxPixels: 1e13
});
```

Instructions:

1. Paste the `Export.image.toDrive` block at the bottom of your script.
2. Click **Run**, then **Tasks**, and click **Run** in the task panel.
3. The map will be exported to your Google Drive.

🗺️ Part 4: Visualize the Cluster Map in QGIS

Step-by-Step:

1. Open **QGIS**.
2. Use **Layer → Add Layer → Add Raster Layer**.
3. Browse to the downloaded GeoTIFF file (from Google Drive).
4. Load the raster.

Customize:

- Go to **Layer Properties → Symbology**
- Set render type to **Categorized**
- Choose a color palette (Color Brewer recommended)
- Add **Legend**, **Title**, and **Labels**

Save:

Export the final map as a **PDF or PNG** for presentations.

🧩 Understanding the Concepts

📦 What is Clustering?

Clustering is an unsupervised machine learning technique. In this case, **K-means** divides Uganda into areas with similar environmental features (NDVI + rainfall) that might indicate similar **malaria suitability**.

🩹 What is NDVI?

Normalized Difference Vegetation Index indicates plant health/greenness, relevant because mosquitoes prefer vegetated environments.

🌧️ What is CHIRPS Rainfall Data?

CHIRPS provides **daily precipitation estimates**—important because water bodies enable mosquito breeding.

✅ Expected Outcome

You will generate a **thematic malaria risk map** showing zones with similar vegetation and rainfall patterns—potential malaria hotspots. You will:

- Understand and apply clustering in GEE
- Use ChatGPT for AI-assisted geospatial coding
- Export and visualize results in QGIS

⚠️ Important Disclaimer

The clusters are **not confirmed malaria zones**, but rather areas with **similar environmental risk conditions**. To confirm actual transmission risk, field validation or overlaying epidemiological data is needed.