# Project Title: AI-Driven Malaria Outbreak Prediction System

**Group 7 Members:**

**NAME Email**

1. Thomas S. McCay [tboyemccay2000@gmail.com](mailto:tboyemccay2000@gmail.com)
2. David T Paye [paye6060@gmail.com](mailto:paye6060@gmail.com)

# Literature Review

## 1. Introduction

Malaria remains one of the most deadly infectious diseases, particularly in Sub-Saharan Africa. According to the WHO (2023), over 247 million malaria cases and 619,000 deaths were reported globally, with most deaths occurring among children under five. Predicting outbreaks is critical because timely interventions—such as distributing mosquito nets, improving access to treatment, and indoor residual spraying—can save thousands of lives. A literature review is necessary to understand past efforts, existing gaps, and successful methods in malaria prediction. This ensures that our project builds upon proven research while addressing limitations.

## 2. Organization

We review malaria prediction research thematically:  
1. Climate and Environmental Factors in Prediction  
2. Machine Learning and AI in Malaria Prediction  
3. GIS and Data Visualization Tools for Malaria  
4. Challenges and Gaps in Existing Studies

## 3. Summary and Synthesis

a. Climate and Environmental Factors

Pascual et al. (2006) highlighted the strong correlation between rainfall, humidity, and temperature with malaria incidence. Caminade et al. (2014) used climate models to simulate malaria transmission under climate change, finding that warming trends shift transmission zones.  
Contribution: Established weather as a core predictor of malaria.  
Limitation: Climate-only models lack patient-level or health facility data.

b. Machine Learning and AI Approaches

Loha et al. (2019) applied Random Forests and Support Vector Machines (SVM) to predict malaria incidence in Ethiopia, showing higher accuracy than traditional statistical models. M’boga et al. (2021) developed a deep learning model combining climate and population data, achieving over 80% accuracy.  
Contribution: AI significantly improves predictive accuracy.  
Limitation: Requires large datasets and computing power.

c. GIS and Data Visualization

Machault et al. (2011) demonstrated that GIS mapping of malaria cases helps identify hotspots and plan interventions. Gething et al. (2016) produced global malaria maps using spatial-temporal modeling, improving decision-making in public health.  
Contribution: Visualization makes data accessible for policymakers.  
Limitation: High dependency on geocoded health facility data, often missing in rural Africa.

d. Challenges and Gaps

Many studies rely on historical health data, which may not be timely. Few systems integrate real-time climate data with AI models for outbreak forecasting. Additionally, limited deployment exists at the community level; most are academic models, not practical tools.

## 4. Conclusion

Key takeaways from existing literature:  
- Climate and environment are strong predictors.  
- AI and ML outperform traditional models but require robust datasets.  
- GIS mapping enhances usability for decision-makers.  
Gap: Lack of real-time, community-focused, integrated malaria prediction systems in rural Sub-Saharan Africa.  
Our Contribution: This project will integrate real-time weather data, AI prediction models, and GIS dashboards to provide practical, localized, and early warnings, directly supporting SDG 3: Good Health and Well-Being.

# Data Research

## 1. Introduction

Our research question is: “Can machine learning models predict malaria outbreaks using climate, health, and population data?” Exploring reliable data is necessary because prediction accuracy depends on data quality. Without strong datasets, models will fail to produce useful results for health interventions.

## 2. Organization

We categorize data into:  
1. Climate & Environmental Data  
2. Health Facility & Case Data  
3. Demographic & Population Data

## 3. Data Description

- Climate Data: Sources include NASA POWER, WorldClim, and NOAA. Format: CSV/GeoTIFF. Variables: temperature, rainfall, humidity.  
- Health Data: WHO Malaria Surveillance Data, DHS (Demographic and Health Surveys). Format: CSV/Excel. Contains malaria case counts, mortality, treatment.  
- Population Data: WorldPop or census datasets. Format: GeoJSON/CSV. Provides population density, age groups, rural/urban breakdowns.  
- Data Size: Climate datasets are large (~GBs), health datasets smaller (~MBs).

## 4. Data Analysis and Insights

Preliminary analysis shows high correlation between rainfall peaks and malaria outbreaks. Health data shows seasonal spikes in malaria cases (after rainy seasons). Population density affects outbreak severity (urban vs rural).  
Visualizations: Time-series charts of rainfall vs malaria cases; heatmaps of malaria incidence.

## 5. Conclusion

Key findings:  
- Weather strongly influences malaria outbreaks.  
- Combining climate + health + population data yields better predictive power.  
Importance: Our model can support proactive malaria prevention by alerting communities before outbreaks occur.

# Technology Review

## 1. Introduction

Technology selection is critical to ensure accuracy, scalability, and usability of our malaria prediction system. Reviewing available tools helps us identify the most effective ML algorithms, cloud platforms, and visualization frameworks for our project.

## 2. Technology Overview

- Machine Learning Frameworks: Scikit-learn, TensorFlow, PyTorch.  
- Data Platforms: Google Earth Engine (for climate data), Kaggle datasets.  
- Visualization Tools: Power BI, or web dashboards with Plotly/D3.js.  
- Cloud Deployment: AWS, Google Cloud

## 3. Relevance to Project

ML frameworks enable predictive modeling. GIS platforms provide geospatial insights. Cloud ensures real-time updates and scalability.

## 4. Comparison and Evaluation

- Scikit-learn: Easy to implement, good for classical ML (Random Forests, SVM).  
- TensorFlow/PyTorch: Best for deep learning, higher accuracy but needs more compute.  
- Tableau/Plotly: Visualization dashboards, user-friendly.  
- Google Earth Engine: Free, powerful for handling large climate datasets.  
- AWS/GCP: Scalable but cost considerations.

## 5. Use Cases and Examples

Loha et al. (2019) used Random Forests in malaria prediction. Gething et al. (2016) mapped malaria with GIS models. IBM AI for Social Good applied ML for predicting disease outbreaks in Africa.

## 6. Identify Gaps and Research Opportunities

Existing studies rarely combine real-time weather + health data. Few are mobile-friendly for use in rural health centers. Our project can bridge this by offering a lightweight dashboard + SMS alerts.

## 7. Conclusion

Key takeaway: ML (Random Forest/Deep Learning) + GIS + Cloud forms the strongest tech stack for malaria prediction.  
Importance: Ensures scalable, accurate, and user-friendly solutions for public health workers.  
Contribution: Our system will integrate these technologies into a real-world, deployable early warning system.