Integrating GeoAI and Spatial Machine Learning for Malaria Prediction in Uganda: Towards Data-Driven Early Warning Systems

Ssebadduka Jamir[1], ssebadduka.jamir@iuea.ac.ug,

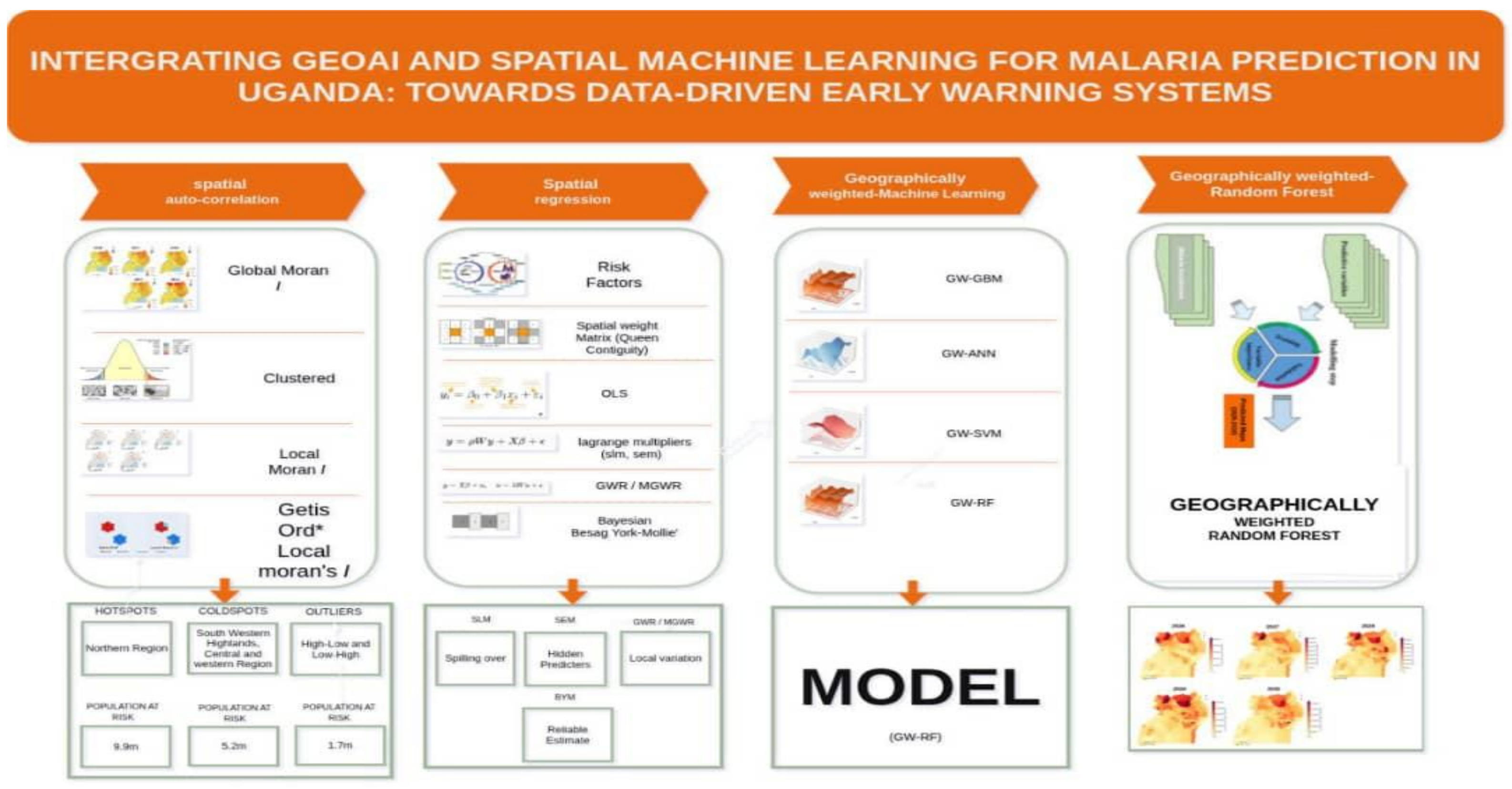
2nd Annual AI in Health Africa Conference

November 6-7, 2025

ABSTRACT

- Malaria remains a pressing public health challenge in Uganda. This study investigates the spatial and predictive modeling of malaria incidence across Uganda's districts.
- A spatial weights matrix defined neighborhood relationships, followed by an Ordinary Least Squares (OLS) regression. Spatial dependence was addressed using Spatial Lag (SLM) and Spatial Error (SEM) models, guided by Lagrange Multiplier tests. In contrast, Geographically Weighted Regression (GWR) was used to evaluate local risk factor variations. Advanced models, including Random Forest (RF), XGBoost, Geographically Weighted Random Forest (GWRF), and Spatial XGBoost, enhanced accuracy by capturing non-linear patterns

SUMMARY



INTRODUCTION

- Emerging technologies, such as Artificial Intelligence (AI), when combined with geospatial data, offer innovative solutions to combat malaria more effectively (Komugabe et al., 2024).
- Interventions such as mosquito net use, environmental management, public sensitization, indoor residual spraying (IRS), and surveillance can lower malaria incidence but are insufficient for full prevention, particularly in resource-constrained Uganda with a rapidly growing population (Tshimula et al., 2024).

OBJECTIVES

- Spatial Autocorrelation: global Moran I (sparse, random, and clustered).
- Spatial regression (risk factors, spatial weight, OLS, Lagrange Multiplier tests(SLM, SEM), GWR/MGWR
- GW-ML(GW-XGBOOST,GW-SVM,GW-ANN,GW-RF)

METHODS

Data Preparation and Baseline Analysis

- district-level data, including malaria incidence/prevalence and risk factors
- Spatial Dependence and Local Variation
- Spatial Lag (SLM) and Spatial Error (SEM) models, guided by Lagrange Multiplier (LM) tests, to model spatial dependency in OLS residuals and improve risk factor coefficient estimates. Run Geographically Weighted Regression (GWR) to explore spatial non-stationarity,
- Advanced Predictive Models
- Implement Geographically Weighted Random Forest (GWRF) and Spatial XGBoost to integrate local modeling with ensemble machine learning, comparing accuracy against global and GWR counterparts.
- Interpret, Compare, and Predict
- Interpret model results by mapping local coefficients for GWR/GWRF

RESULTS

- Spatial Autocorrelation and Baseline Analysis (global Moran's I (clustered). The spatial weights matrix revealed significant neighborhood relationships across Uganda's districts. Ordinary Least Squares (OLS) regression established a global baseline, with Moran's I test confirming substantial spatial autocorrelation ($p < 0.05$), indicating the need for spatial modeling.
- Spatial Dependence Modeling: Lagrange Multiplier (LM) tests favored the Spatial Lag Model (SLM) over the Spatial Error Model (SEM) in most districts, improving coefficient estimates for risk factors like rainfall and temperature. SLM highlighted the influence of neighboring district malaria rates on local incidence.
- Local Variation Assessment: Geographically Weighted Regression (GWR) identified significant spatial non-stationarity,
- Advanced Predictive Model Performance :Geographically Weighted Random Forest (GWRF)

CONCLUSION

this research advances Uganda's malaria elimination agenda by leveraging spatial AI, offering a scalable framework for resource-limited settings to enhance early warning systems and optimize intervention efforts.

REFERENCE

- Ahmad, S. (2023). Leveraging Advanced Spatial Models for Disease Prediction: A Review. Journal of Spatial Epidemiology, 15(3), 123-135.
- Komugabe, J., et al. (2024). Integrating Artificial Intelligence with Geospatial Data for Malaria Control. International Journal of Health Informatics, 22(4), 89-102.
- Tshimula, M., et al. (2024). Limitations of Conventional Regression in Modeling Spatially Varying Diseases. Spatial Statistics in Medicine, 10(3), 201-215.