Slippery Snails: Determining Indicators for Schistosomiasis in Côte d'Ivoire

Ashal Ali; Andrew Chamberlin, BS; Giulio De Leo, PhD

Department of Statistics, Stanford University

More than **200,000 deaths** are caused by schistosomiasis ("snail fever") every year in Africa. While attempts to develop seasonal indicators of schistosomiasis exist in the literature focused on this disease, not all frameworks are transferrable due to the unique geographic and environmental considerations in schistosomiasis afflicted regions. Here, indicators that can be used to model the spread of schistosomiasis in Côte d'Ivoire are proposed. The indicators have been derived by regression and tree-based modelling across terrain, climate, vegetation, geographic, and human variable groupings.

Introduction

Schistosomiasis is mainly present in tropical and subtropical regions. The vectors of the disease are freshwater snails. Intestinal damage, bladder cancer, and liver cirrhosis are some complications of schistosomiasis. This project uses machine learning to develop indicators for predicting the spread of the disease in Côte d'Ivoire where each observation is extracted from schools in different regions of the country. Indicators for two main parasites are developed: *S. haematobium* and *S.mansoni*.
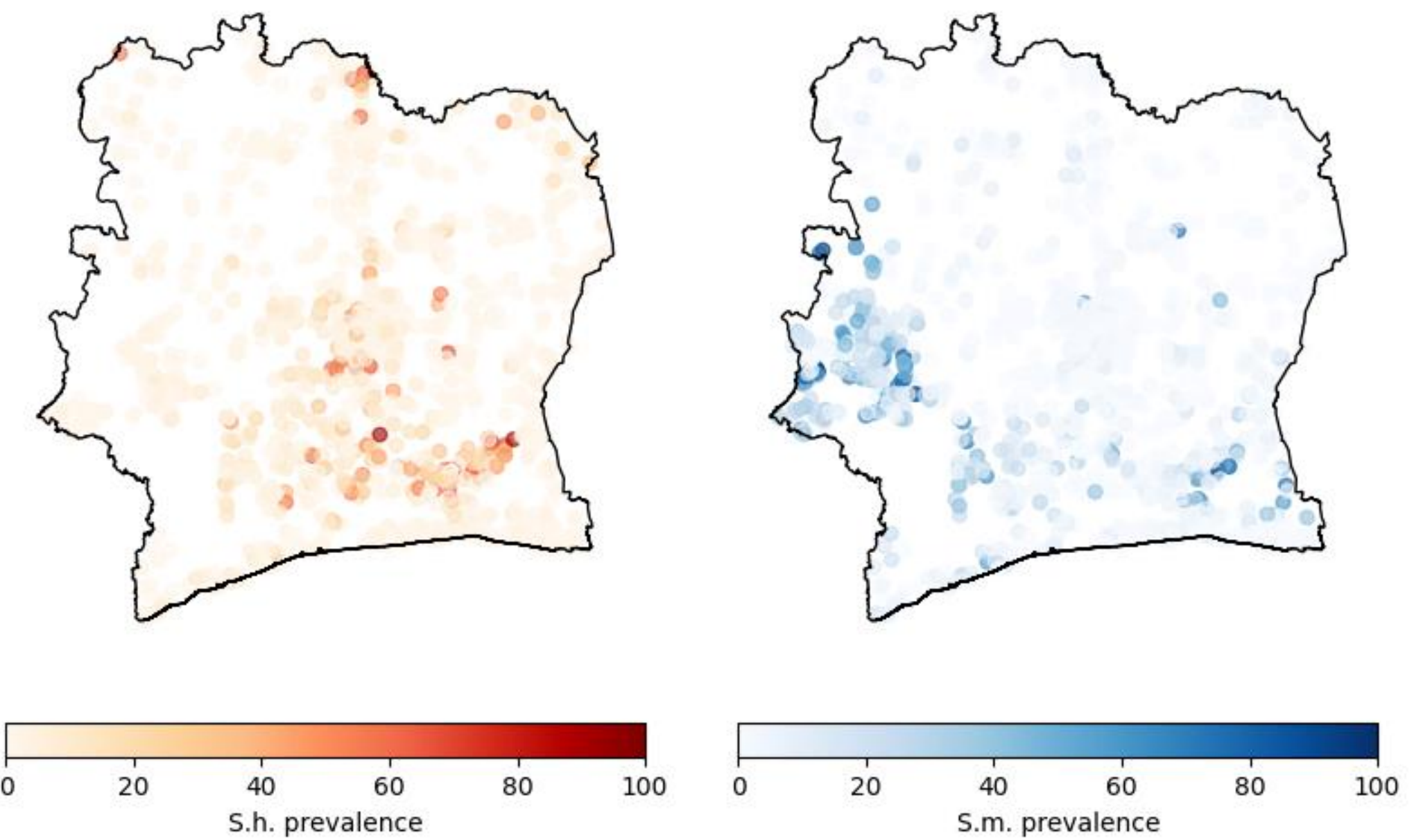

Figure 1. *Prevalence of S.haematobium and S.mansoni in Côte d'Ivoire.* Each observation is a school where schistosomiasis was detected. A key insight to note is that each type of the disease is focused in different regions. This informs a hypothesis regarding each type of the disease being potentially indicated by different indices.

Methods

Five groups were used to code the variables by theme.
- **Terrain:** Soil acidity, elevation, slope, soil water etc.
- **Climate:** temperature, precipitation, vapor pressure etc.
- **Vegetation:** soil index, water index, infrared index etc.
- **Geographic:** irrigated area, built-up area, forest area etc.
- **Human:** population, gross domestic product etc.
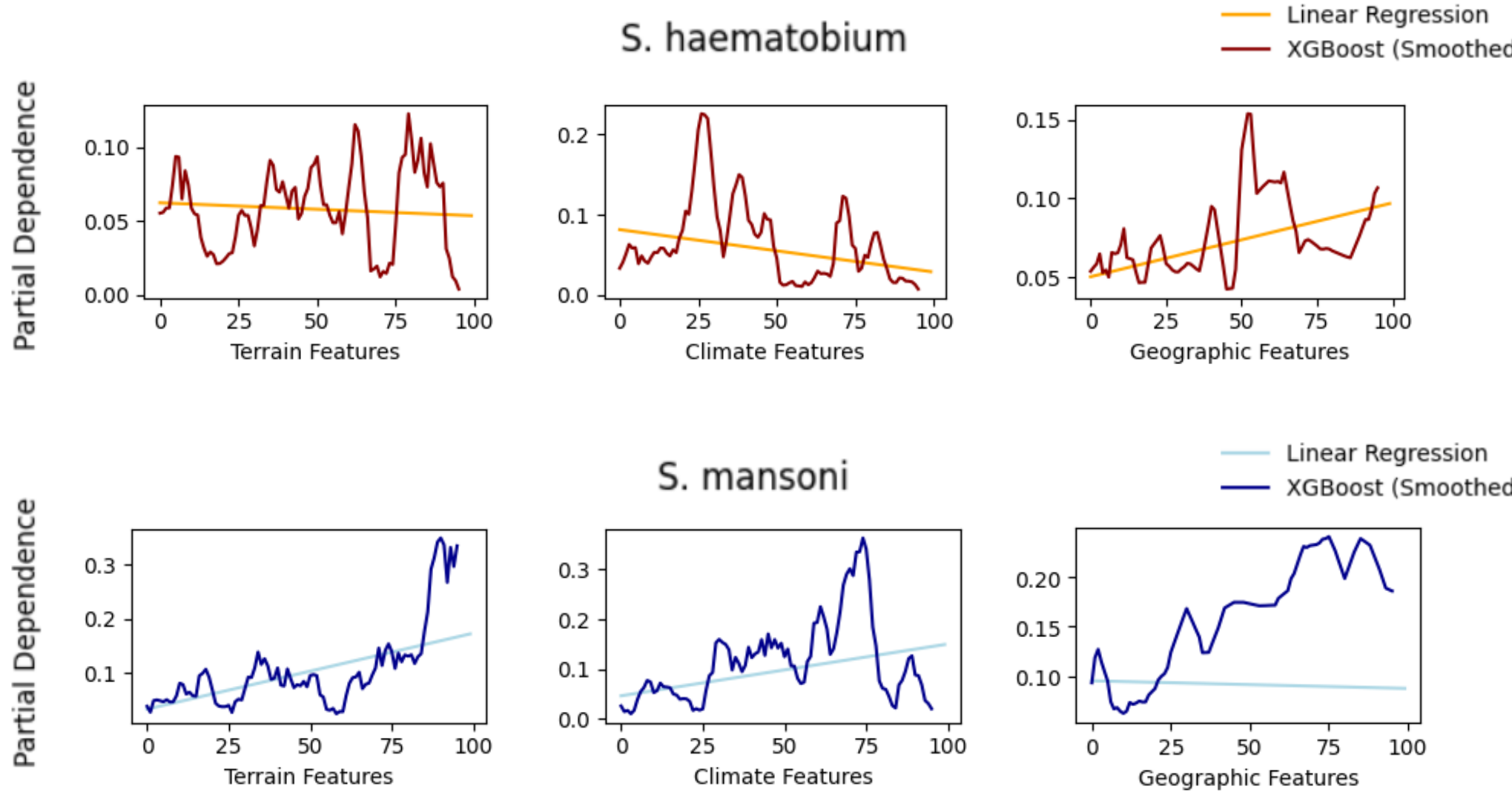
Figure 2. *Partial Dependance plots of Linear Regression and XGBoost applied to three most significant feature groups for each schistosomiasis type.* The visualization compares the most important feature groups determined through a cumulative model for *S. haematobium* and *S. mansoni*. The key insight derived is that the most important feature groups are the same, but the individual variable dependance is different.

Results

Group	Variables
Terrain	aspect, bedrock depth, bulk density, clay, dam distance, nitrogen, organic carbon, soil water, stone, total carbon, upstream drainage area, and water distance
Climate	evapotranspiration, climate water deficit, temperature, precipitation accumulation, run-off, soil moisture, vapor pressure, wind speed, and respective lags
Vegetation	enhanced vegetation index, soil adjusted vegetation index, normalized difference index, normalized difference vegetation index, and respective lags
Geographic	agricultural area, developed area, and change in forest area
Human	gross domestic product mean lag (3yr and 5yr), total population, total population lag (3yr, encoded)

Figure 3. *Indicators for Schistosomiasis Spread in Côte d'Ivoire.* This table highlights the indicators that are common among the variables that are used to predict *S. haematobium* and *S. mansoni* best. It is important to note that there were some differences in the variables used in the groupings for both types due to contrasting individual feature importances within the groups. However, there were very few such variables (n = 6).

Conclusion

This poster uses machine learning to propose a base framework for indices that can be used and finetuned for predicting schistosomiasis using data on *S. haematobium* and *S. mansoni*. To build upon this, future research factoring in other types of schistosomiasis and other regions within Africa can contribute towards the consolidation of a more universal index to predict the spread of the disease.