GINKGO
BIOSECURITY

Measles forecasting model

Ginkgo Biosecurity

August 15, 2024

AI enabled Measles Forecasting (AIM-F)

The Problem	The Project
Measles continues to cause significant mortality: ~136K deaths in 2022, mostly in small children **Forecasting is critical for vaccine deployment**, but currently relies on granular vaccination and transmission data that is seldom available, particularly in LMICs - More innovative approaches are needed to forecast future vaccine needs	- Build a **global, data-driven predictive model** leveraging: **Traditional and non-traditional data**, incl. public health reports, travel patterns, economic activity, and other factors **AI approaches,** incl. machine learning and deep learning, to structure and analyze a multitude of data sources to produce actionable insights,

Model overview

Objective: Develop a machine-learning modeling framework which allows for the inclusion of various types of input data to determine which countries are at highest risk of a measles outbreak in the coming months

Modeling Approach

- Employing ML to overcome the challenge of incomplete and disparate training and case data.
- Leveraging various datasets, many not traditionally used in measles forecasting, to estimate the risk of measles introduction and transmission in a given geography.
- Initial model uses NeuralProphet and similar ML driven regression models to project seasonal predictors and epicurves, fusing traditional time series algorithms and deep learning methods.
- Sensitivity analyses and clustering are being applied to optimize model performance as a function of countries' social, political, ecological, environmental, and/or epidemiological descriptors.

Predictor variables

Climate measures are associated with the disease seasonality

Climate	Rationale	Source
Precipitation	Precipitation is associated with measles incidence in tropical regions	NOAA
Temperature	Temperature is associated with measles incidence in temperate regions	NOAA

Demographic measures may be associated with disease importation and the size of the susceptible population

Demographic	Rationale	Source
Birth rate	Influx of susceptible population	UN
Migration rate	Influx of susceptible population; disease importation	UN

Predictor variables

Socioeconomic measures may be associated with access to, or trust in, preventative and response care

Socioeconomic	Rationale	Source
Nighttime lights	Proxy for urban/rural population as well as socioeconomic status at a granular resolution	NASA
Human Development Index	Summarizes average achievement in key dimensions of human development: a long and healthy life, being knowledgeable, and having a decent standard of living	UN
Income group	Associated with state's ability to fund public health and social programs; associated with the average citizen's standard of living	World Bank
Minorities at risk	Minorities at risk may be less trusting of public health authorities or have reduced access to care	University of Maryland

Predictor variables

Public health preparedness and response measures can influence the likelihood and severity of outbreaks

Public health preparedness and response	Rationale	Source
MCV1 and MCV2 vaccination coverage	Susceptible population	WHO, UNICEF, IHME, Georgetown University
State vaccination requirements and enforcement	Susceptible population; public health infrastructure	Georgetown University
Global Health Security Index	Provides assessment of country's of health security and related capabilities	Brown University
State Capacity Scores	Proxy for ability of government to enact and enforce public health policy	Lee and Zhang 2017
Road access datasets	Road access is important for vaccine delivery and public health response efforts	SEDAC, European Commission

Neural Prophet

Initial development model

- ML-tuned seasonal autoregression model with integrations for lagged and leading predictors
- Focused upon extracting and projecting the periodic trends in data at user defined granularities (eg: day of week, month of year, year to year)
- Tolerant to missing or irregular data
- Capable of global-local modelling

Continuous improvement framework

- An automated sensitivity analysis of individual predictors by country and variable time spans is underway
- Clustering analysis of Neural Prophet performance as a function of regional social, economic, and environmental indicators is underway
- Next steps will include building wrappers and generalizing the sensitivity analysis pipeline to integrate and evaluate reference models