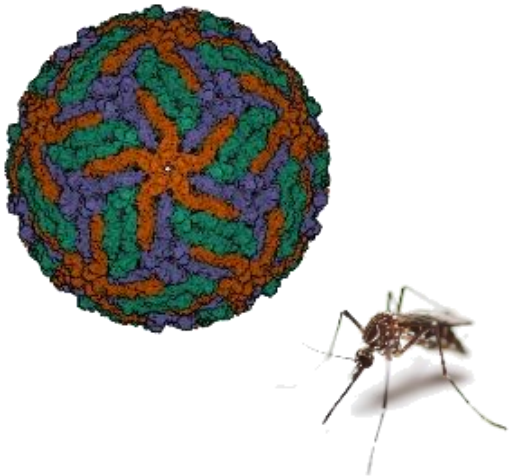
Machine Learning Algorithms for Predicting the Force of Infection of Japanese Encephalitis

Sun Jie, Nguyen Manh Duy, Quan Tran Minh, Shreya Khandelwal, Chris Erwin G. Mercado, Hannah Eleanor Clapham

29 November 2021

Background- Japanese Encephalitis

- JE is a vaccine preventable mosquito-borne flavivirus.
- JE primarily affects children. Most adults in endemic countries have natural immunity after childhood infection, but individuals of any age may be affected.
- Most JEV infections are mild (fever and headache), but approximately 1 in 250 infections results in severe clinical illness.
- Approximately 67,900 JE cases typically occur annually, of which only about 10% are reported to the World Health Organization.



Background- Force of Infection and disease burden

- Force of Infection (FOI) is the rate at which susceptible individuals in a population acquire infection.
- The FOI is a key parameter used in estimation of disease burden in susceptible countries.
- Current FOI estimates assume homogeneity across the entire country- limits assessment of disease burden on a finer scale.
- Due to lack of data- burden estimates are based on broad assumptions.
- Hence- there is a need to estimate the FOI on smaller geographical scales

Methodology

We use tree based machine learning models to estimate the FOI for pixel-level (10s x 10s) data points.

A. Data:
   - Demographic variables (population density, age groups)
   - Geographical data (elevation, Pigs and rice distribution)
   - Bioclimatic data

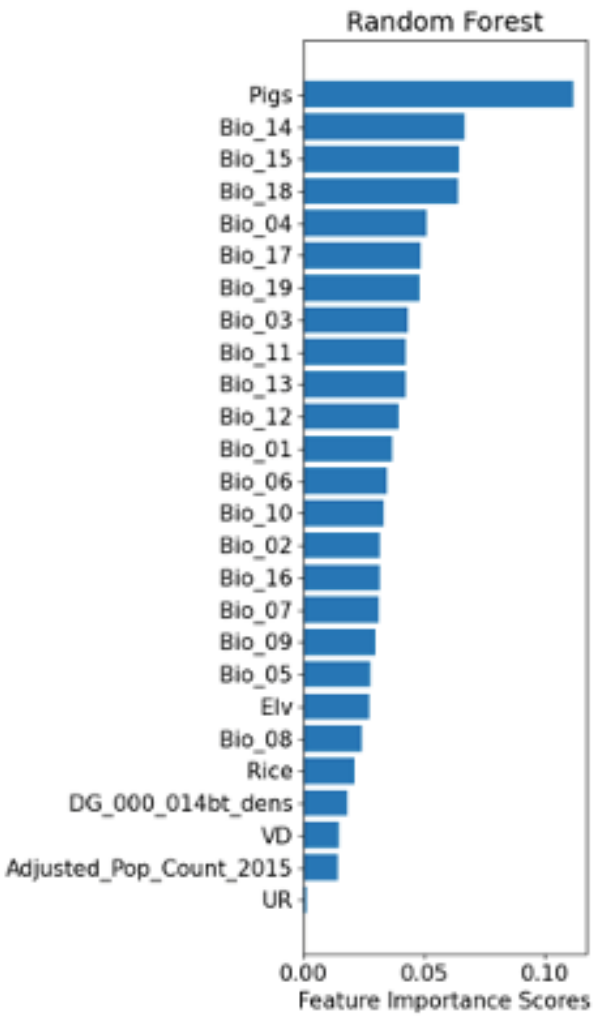
B. Models used:
   - Principal Component Regression (PCR)
   - Random Forest (RF)
   - Gradient Boosting (GB)
   - XGBoost (XGB)

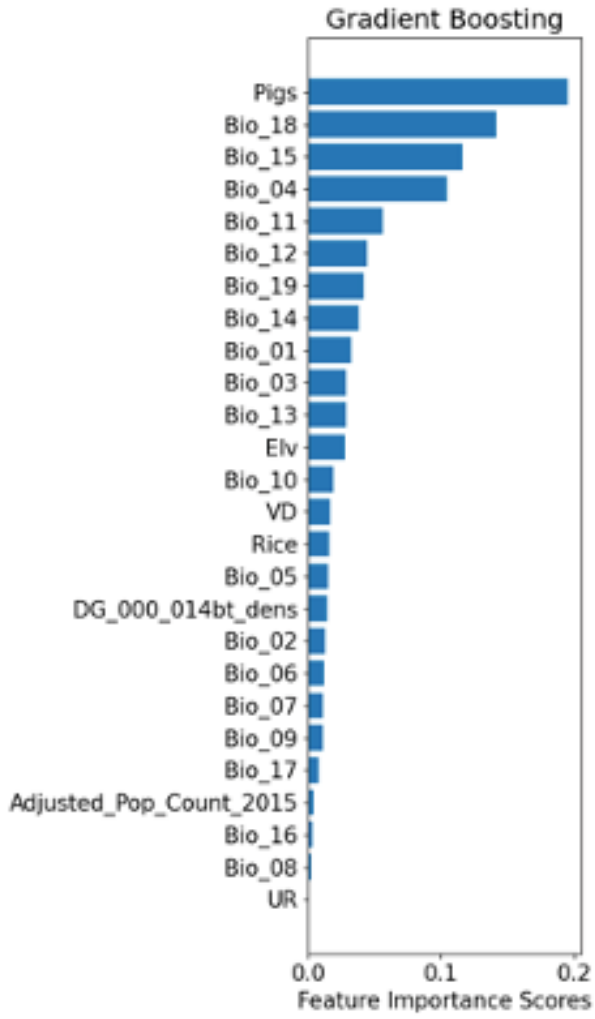
Results- MSE

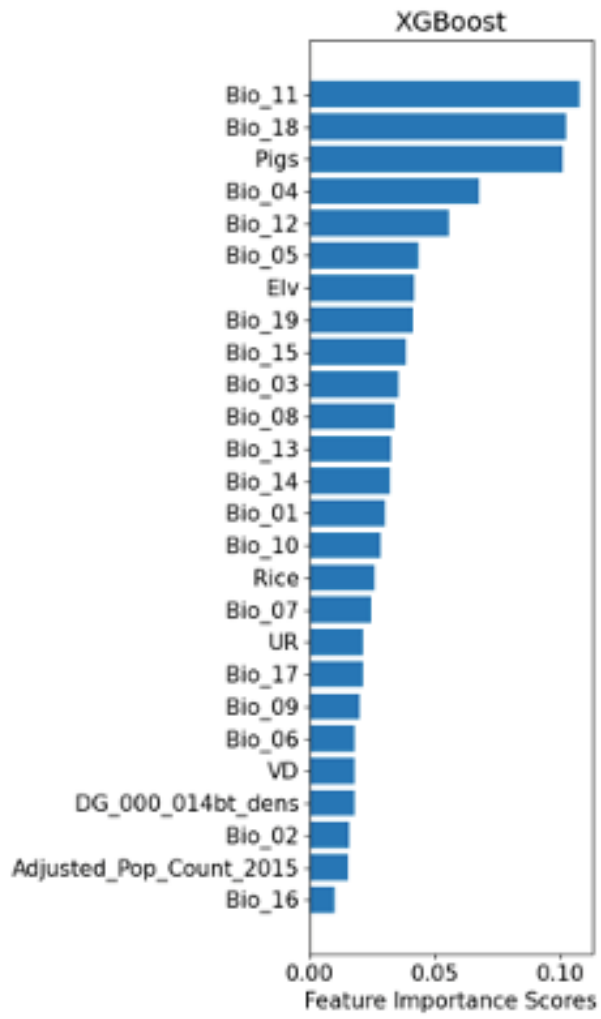
We compared the test Mean Squared Error (MSE) for the four modelling methods:

PCR	0.00496
Random Forest	0.00197
Gradient Boosting	0.00183
XGBoost	0.00172

Results- Feature Importance







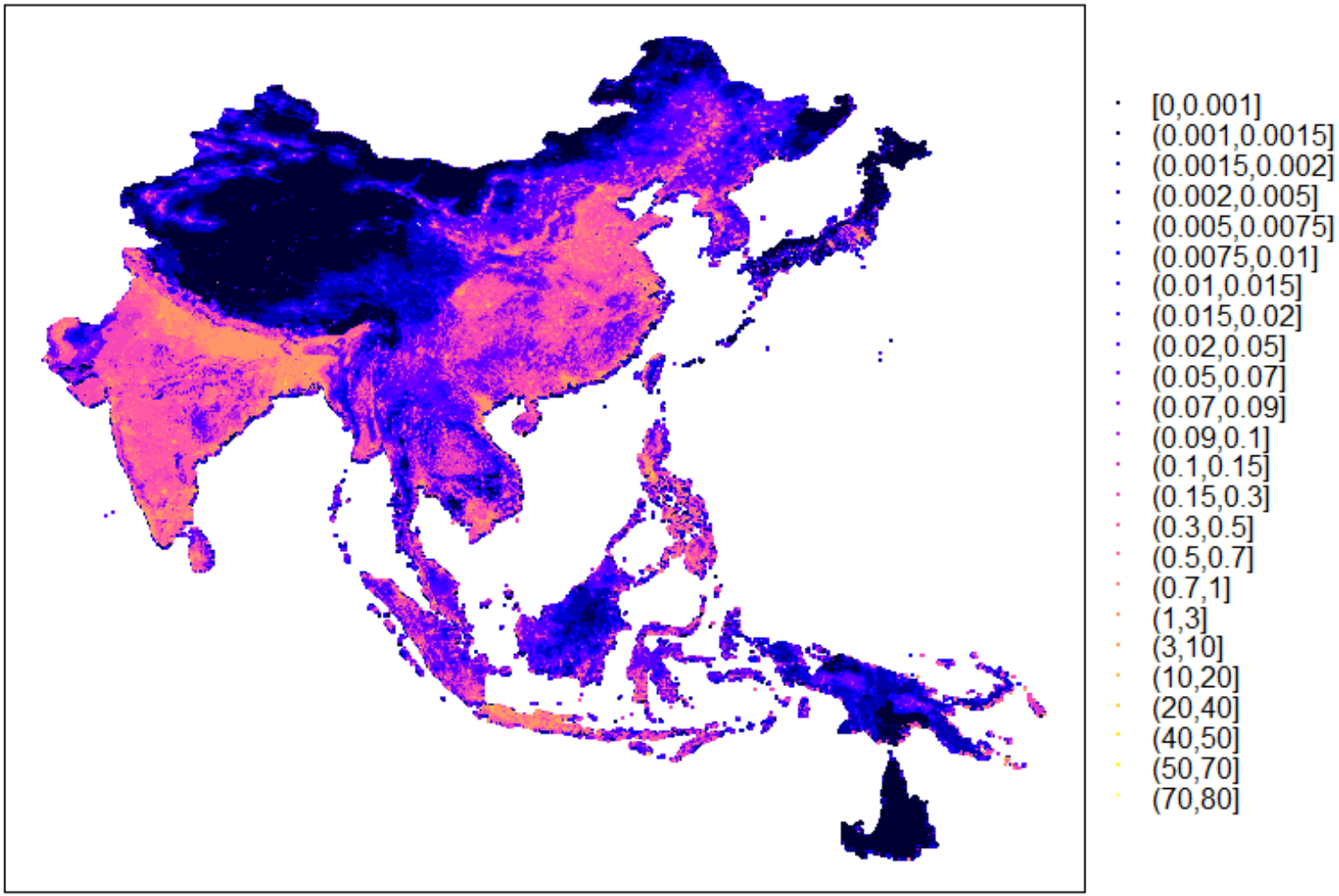
Results

- We also used Individual Conditional Expectation (ICE) plots.

- Our results suggested that the top 10 features are relatively consistent across methods. These features included various bioclimatic variables and the pig distribution.

- Features with low importance scores across all three methods, such as adjusted population count and the urban-rural divide (UR), might indicate that they do not contribute much to the prediction of FOI.

Output- Generating number of JE cases



Conclusion

- The finer scale estimates of the FOI allowed for the precise calculation of disease burden, even for countries without surveillance systems in place for JE.
- More accurate burden estimates lead to more effective understanding of vaccination impact in countries with significant JE burden.
- The study offers insight on using tree based algorithms for understanding public health problems involving heavily correlated predictors.