Interpretability in Machine Learning for Epidemiological Forecasting

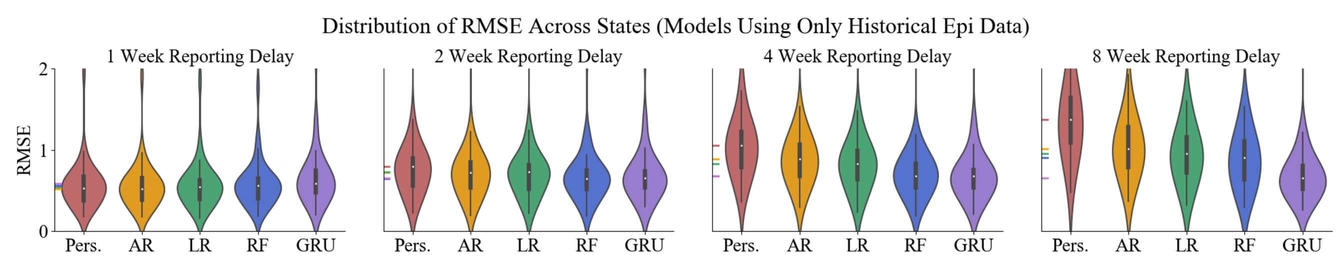
HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH

Emily Aiken & Jonathan Waring, APCOMP221

HARVARD John A. Paulson School of Engineering and Applied Sciences

Problem Statement

- Time-series machine learning methods provide accurate predictions of state-level influenza activity in the United States up to eight weeks in advance, but lack interpretability.



- We evaluate feature extract feature importance for several machine learning methods.
- We use averaging and clustering methods to evaluate whether machine learning models pick up on known spatiotemporal patterns of influenza spread.

Data and Methods

- **Data:** Weekly influenza case-counts for 37 states for the years 2010-2017 from the CDC.
- **Modeling Methods:** Each model uses 52 historic lags from 37 states as input, we extract feature importance for the 37-by-52 input matrix in each week
- **Clustering Methods:** We examine average influence of each state on each other state, and cluster states by their major predictors.

Method	Architecture	Feature Importance
Linear Regression	LASSO (L1 regularization)	Regression coefficients represent the influence of each input feature on the prediction.
Random Forest	Ensemble of 50 decision trees	Feature importance represents how much each feature contributes to decreasing variance in the target after splitting.
Gated Recurrent Unit Neural Network (GRU)	One-layer GRU with 5 nodes in the hidden layer	We extend CNN saliency maps [5] to our regression problem by visualizing attention over inputs that increase or decrease regression output.

Results: Feature Importance

- Southern states are important features to all other states.
- For short time horizons of prediction (1-2 weeks), early lags are important, for longer time horizons (4-8 weeks), seasonal lags are more important and GRU attention extends further back.
- Whereas the important features in the linear regression and random forest are vary across states, the same states come up as the most important for GRU predictions.

Figure 1: Feature Importance for One Week in MA (8-week Forecasts)

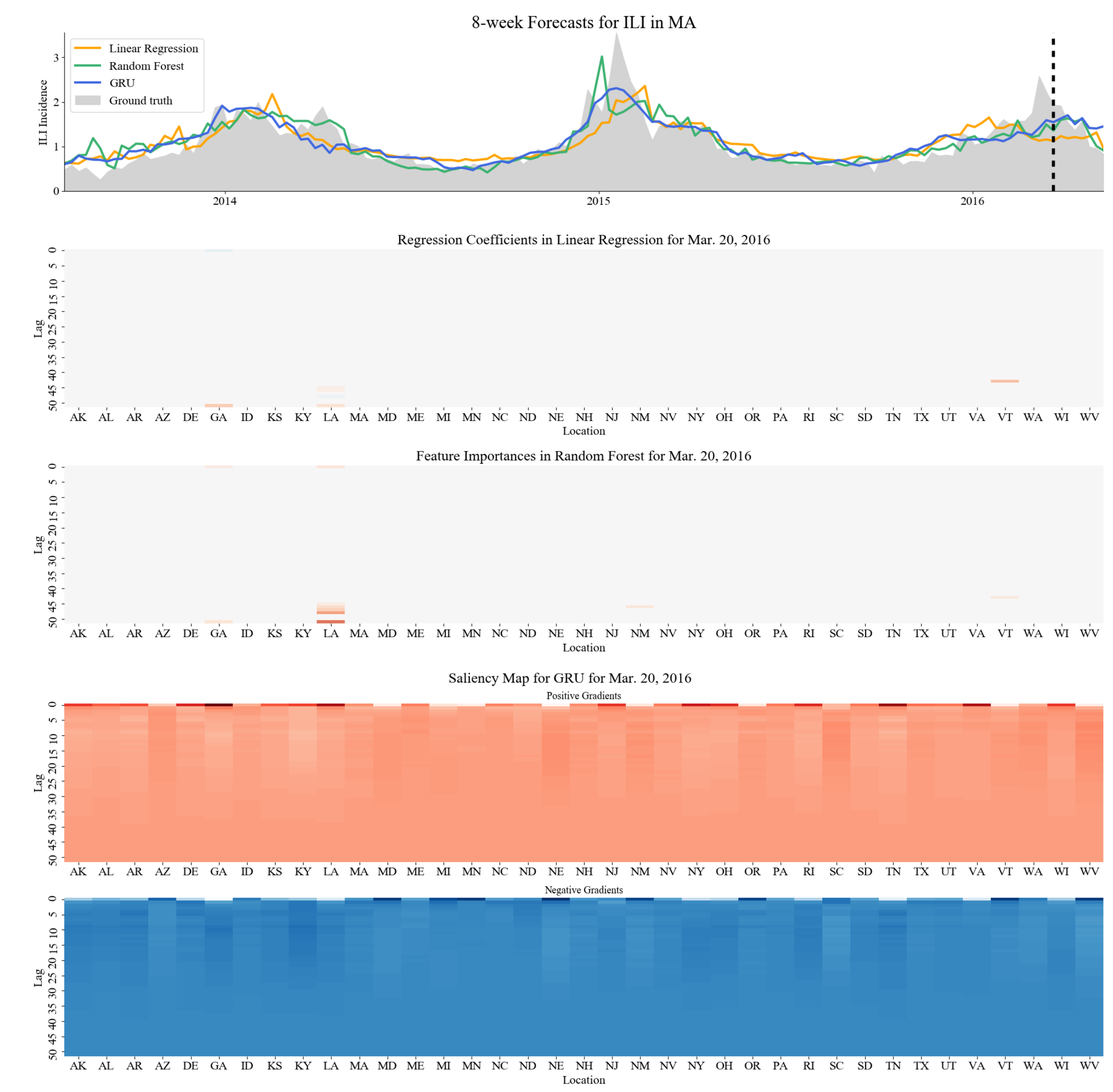
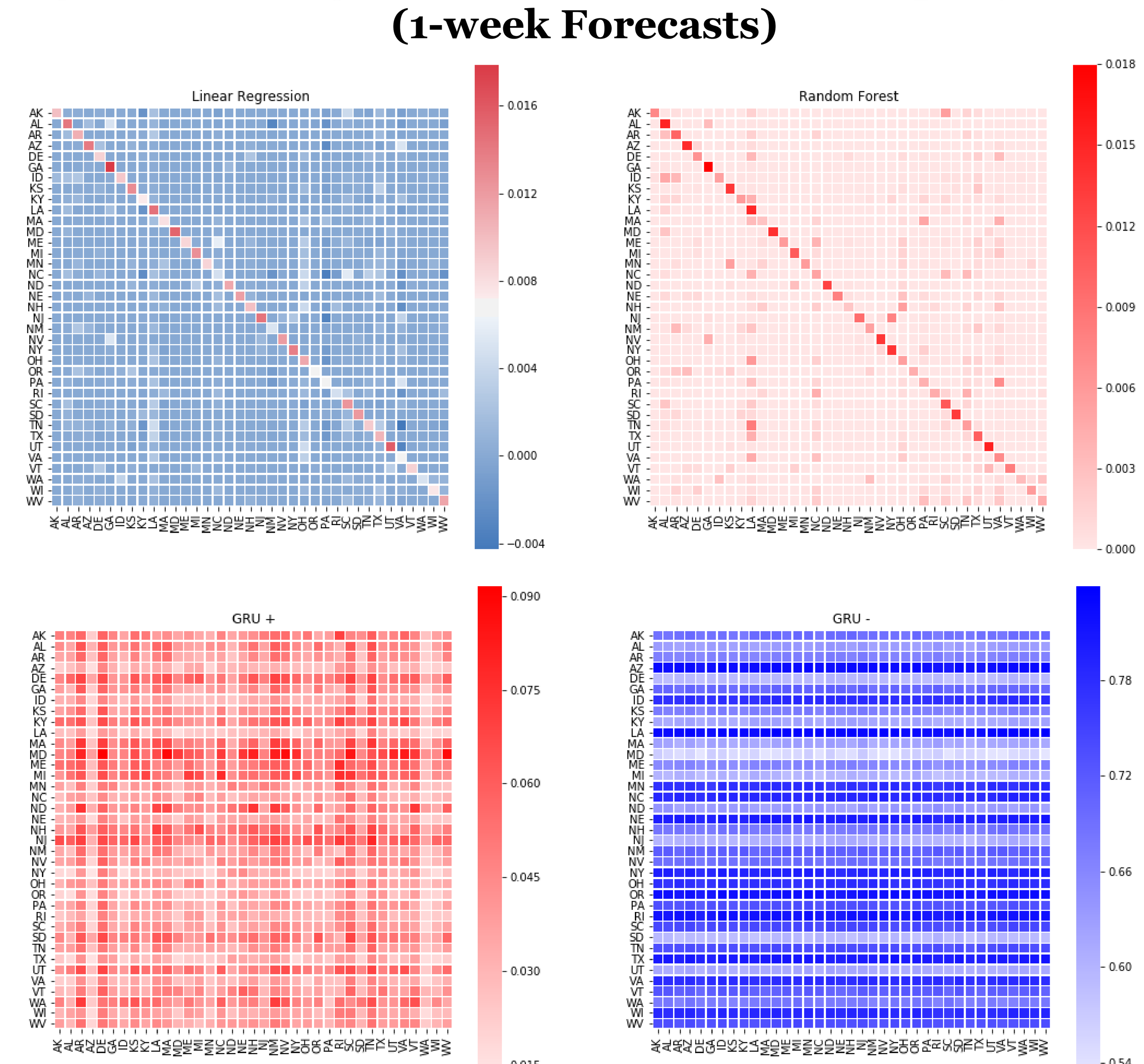


Figure 2: Averaged Cross-Influence Heatmaps by Model (1-week Forecasts)



Results: Clustering

Figure 3: K-means Clustering on Average Feature Importance for Random Forest (4-week Forecasts)

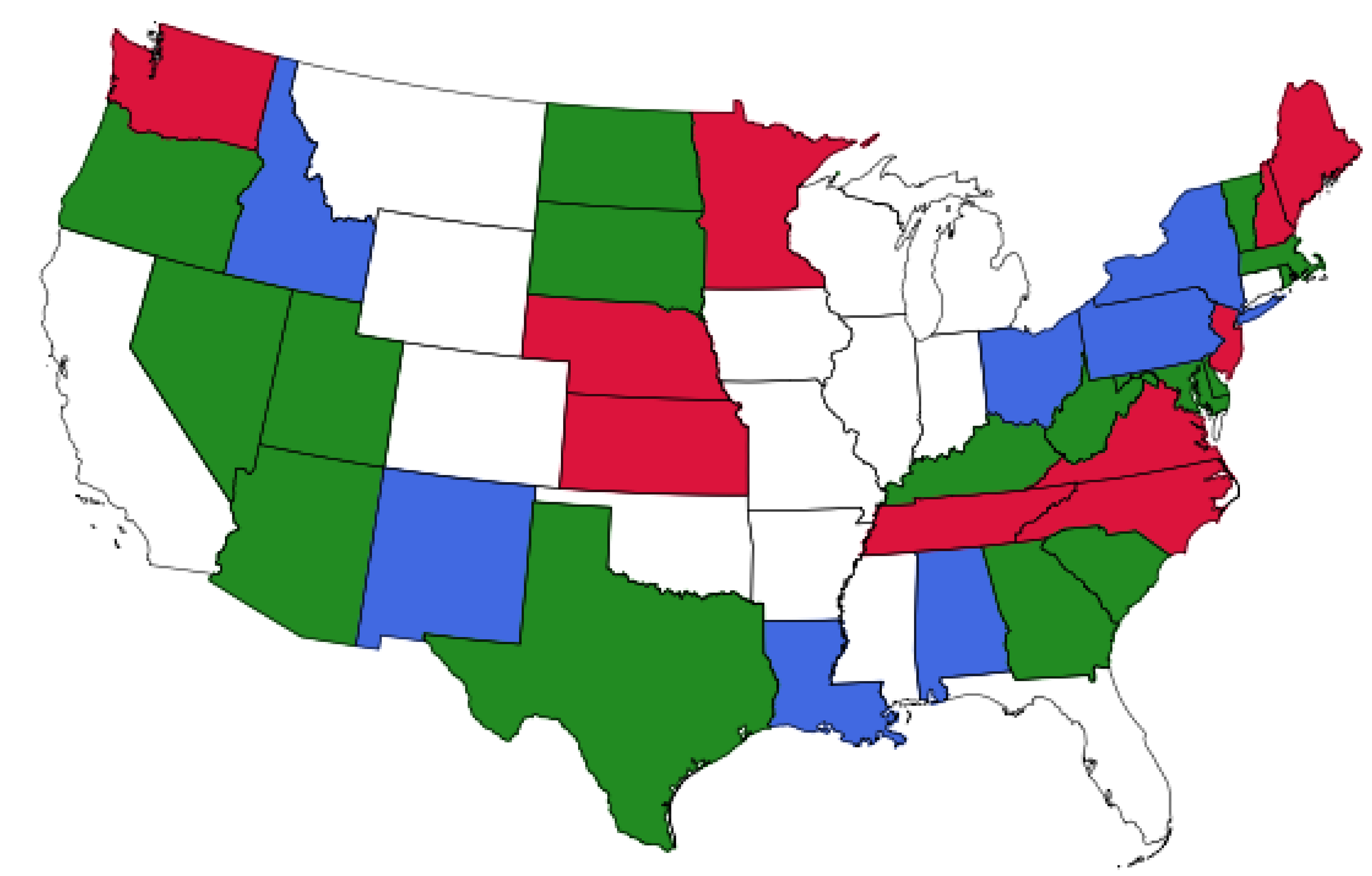
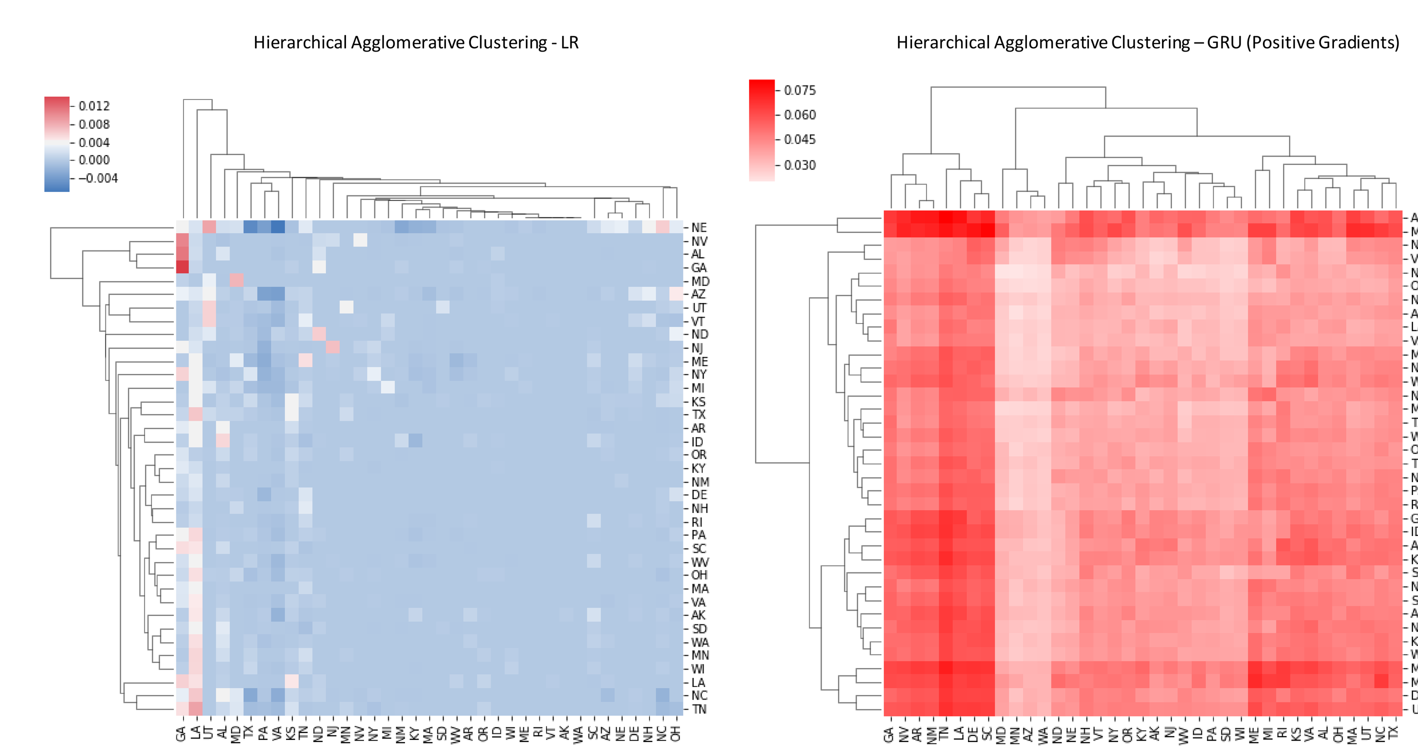


Figure 4: Comparison of HAC on Average Feature Importance for LR and GRU (4-week Forecasts)



Discussion

- Our methods pick up on intuitive temporal differences in feature importance between models that predict at short and long time horizons.
- We do not observe strong intuitive spatial relationships in feature importance. It is possible that repeating this analysis on a higher spatial resolution (ex. city-level influenza) would yield more intuitive spatial results.
- It is clear that the GRU attention takes a more holistic view of the time series than the other methods, which may partially explain why it has better performance.