FluScope: A Multi-Resolution Approach for Modeling Influenza Hospitalizations

Aditya Brahme, Achilles Emnace, Ishaan Sinha
Advisor: Prof. Nikhil Muralidhar

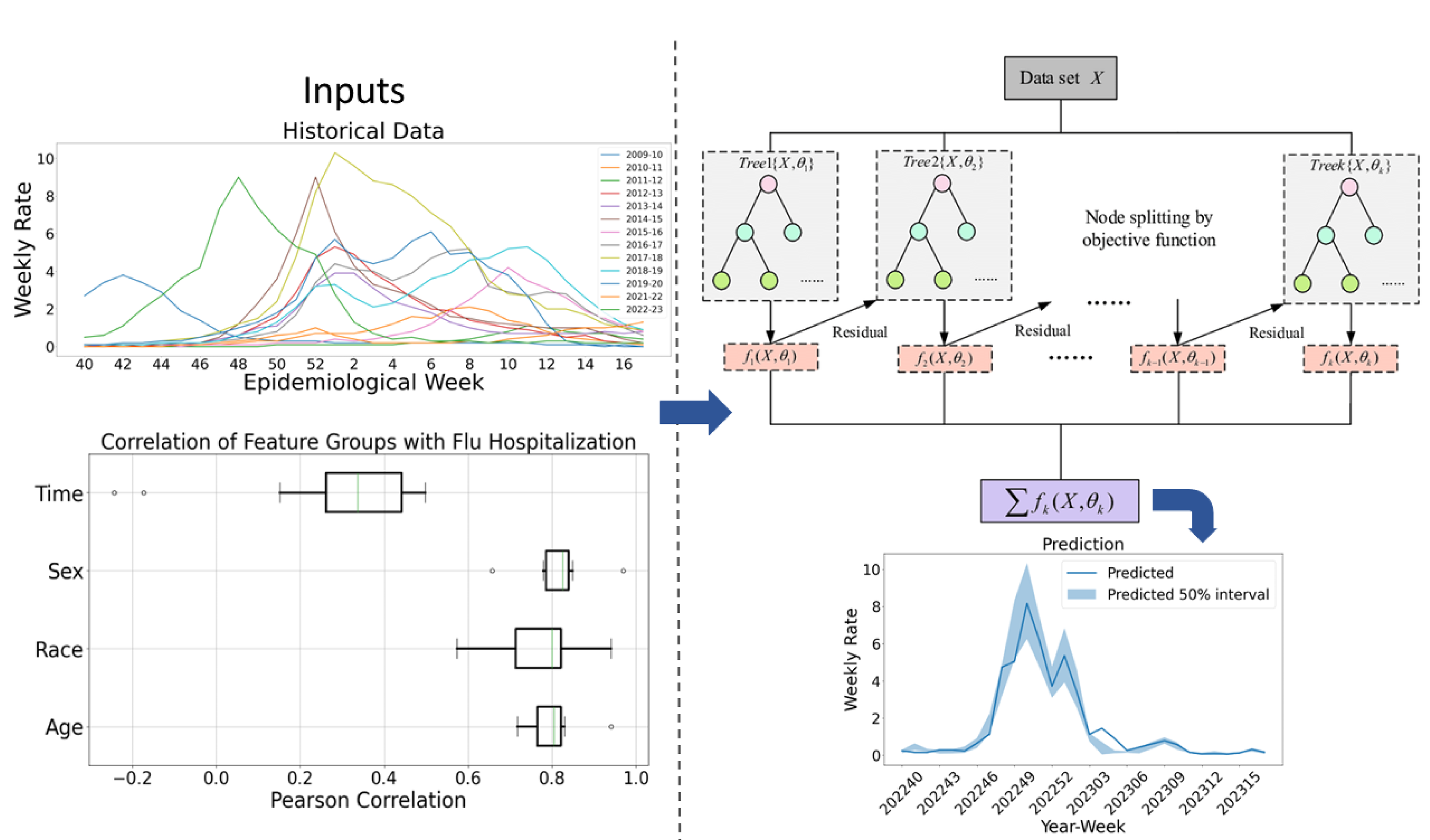
STEVENS INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY

AI Research in Summer (AIRS) Fellowship Program 2023

Introduction

- Influenza (flu) is a **contagious respiratory illness** caused by viruses that infect the nose, throat, and lungs.
- Flu is a seasonal burden and flu epidemics cause **140,000 – 710,000 hospitalizations** and **12,000 – 52,000 deaths** each year.
- Forecasting flu season dynamics, helps the CDC, healthcare facilities and policy makers **effectively develop responses** for increases in hospitalizations, **antiviral treatment allocation**, and for **informed distribution of healthcare workers and resources.**
- Forecasting influenza outcomes is inherently **challenging** due to **seasonal variation, local and global dynamics**, compounded by the continuous **evolution of flu viruses**, resulting in distinct characteristics for each flu season.
- Since 2013, the CDC has released data about each flu season and the corresponding hospitalization dynamics, from FluSurv-NET.
- In this effort, we develop a data-driven solution to forecast influenza hospitalizations in the USA. Specifically, our **multi-resolution approach** captures both *local* and *global* characteristics for modeling weekly hospitalization dynamics during the flu season.
- Specifically, each forecast of our proposed solution for a region comprises a combination of predictions from (i) a *global* model that captures **overarching hospitalization trends across the US**, (ii) a *local* (region-specific) model to capture nuances in hospitalization dynamics localized to each region.

Methods



NN (L) - Neural Network Local:
Region-specific feed-forward neural network (6 hidden layers) for predicting weekly hospitalization rate.

GB (L) - Gradient Boosting Local:
Region-specific gradient-boosting regression model for predicting weekly hospitalization rate.

GB – Gradient Boosting Global:
Single (global) model trained to forecast weekly hospitalization rates across all regions in the US.

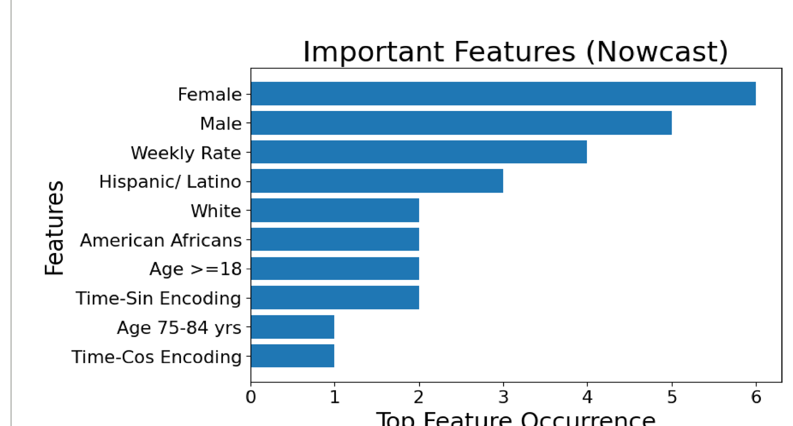
GB+NN(L) - Stacked model:
Combination of best performing *local* (NN-L) & *global* (GB) techniques. Leverages complementary strengths of each model to collectively improve predictions.

Lin-Reg – Linear Regression:
A simple linear regression baseline for predicting weekly rates of hospitalization.

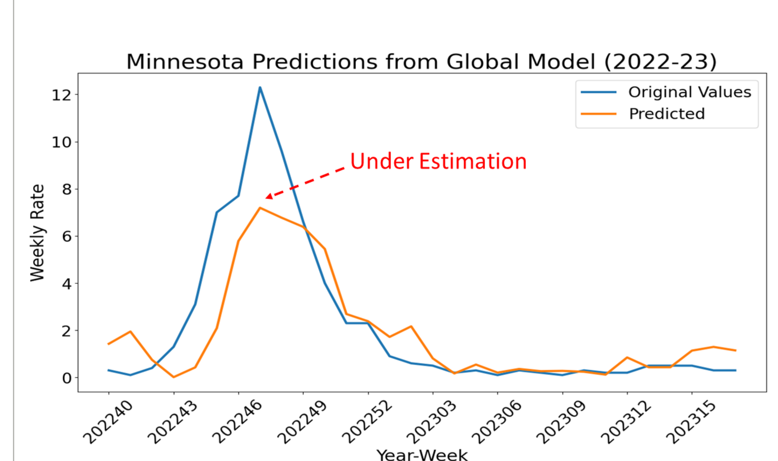
Evaluation Setup

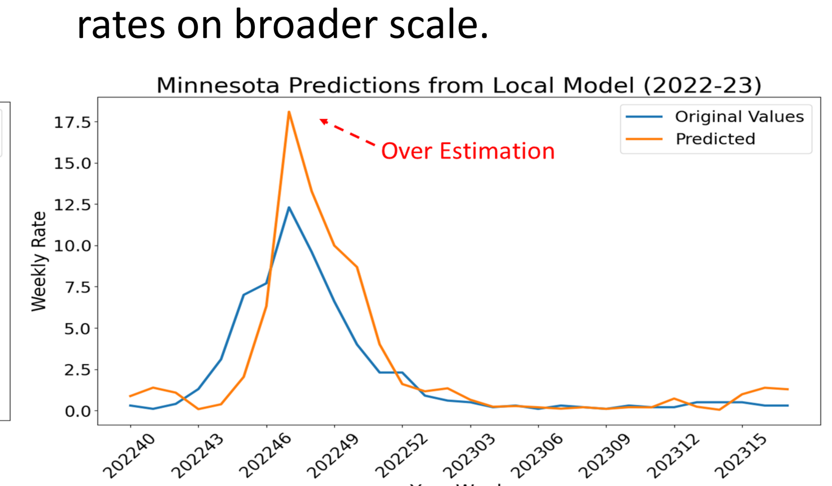
- Evaluation Metrics: MAE, MSE, R^2 Score
- Training Period: Week 40, 2009 – Week 17, 2022
- Testing Period: Week 40, 2022 – Week 17, 2023

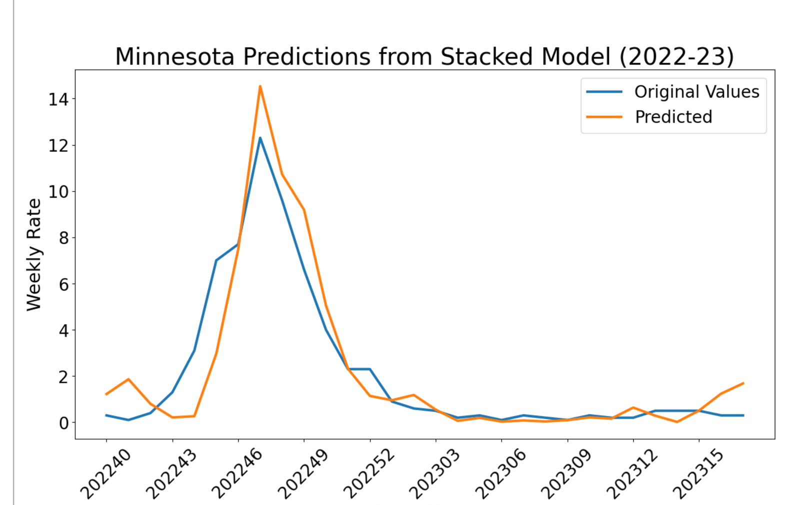
Results



Important Features (Local Model)
Hospitalization of females serves as a most indicative predictor for forecasting overall hospitalization trend While hospitalization rates stratified by race prove to be valuable indicator for assessment of weekly hospitalization rates on broader scale.







Nowcast (Minnesota)
Qualitative inspection of model predictions reveals that GB tends to underestimate hospitalization rates while NN(L) overestimates said rates, relative to the observed values. However, integration of these models into a stacked ensemble, demonstrates significant performance improvement.

Quantitative Results

	Nowcast					2-week Ahead				
	Lin-Reg	NN(L)	GB(L)	GB	GB+NN (L)	Lin-Reg	NN(L)	GB(L)	GB	GB+NN (L)
California	4.66	2.32	12.37	**1.56**	1.72	10.34	4.46	10.60	4.71	**1.51**
Colorado	2.77	2.68	1.60	**1.17**	1.56	5.68	4.34	3.12	3.88	**1.37**
Full Network	1.53	1.28	2.26	2.36	**1.24**	4.63	2.11	7.45	6.45	**1.60**
Georgia	2.39	1.22	1.13	1.35	**0.98**	3.31	2.46	2.02	6.55	**0.98**
Maryland	5.26	2.10	2.80	2.89	**1.80**	8.02	3.80	4.23	7.11	**2.58**
Minnesota	6.73	**2.20**	3.85	2.87	3.19	19.96	11.00	11.17	12.48	**2.95**
New Mexico	8.32	4.89	13.63	8.92	**4.41**	26.96	13.19	20.78	19.22	**9.86**
New York - Rochester	10.69	3.61	**2.40**	3.03	3.58	37.85	15.96	10.59	8.88	**4.19**

MSE evaluation for prediction of weeks t (*nowcasting*) and t+2 (i.e., *2-week ahead*) given data until week t-2. GB+NN(L) has good overall performance with a mean percentage improvement of **23%** over GB for nowcasting and **63.85%** for predicting t+2. The mean percentage improvement of GB+NN (L) over L-Reg is **56%** for nowcasting and **78.55%** for 2-wk ahead prediction. Stacked model had poor performance on regions that displayed two peaks, such as California and New York Albany. However, its strengths are more pronounced when predicting hospitalization trends over extended periods (2-week ahead).

Conclusions

- Influenza exhibits region-specific and global dynamics, and in this effort, we developed a multi-horizon model capable of capturing both region-specific (local) dynamics and overall (global) dynamics to model hospitalization rates across the US. Specifically, our model yields up to 2-week ahead forecasts of hospitalization rates in the US, during the 2022-23 flu season.
- Our stacked model performed better than linear and non-linear single-horizon baselines. Local models perform well in regions where specific local patterns are predominant (e.g., 'Maryland'). However global models struggle to learn these intricate patterns
- Combining global and local insights emerges as the most effective choice. It leverages the strength of both approaches to provide accurate flu predictions, ensuring adaptability to varied regional dynamics.
- It is worth highlighting that the stacked model excels in situations requiring forecasting for longer time horizons.