Steering a Historical Disease Forecasting Model Under a Pandemic: A Case of Flu and COVID-19

Authors:

- Alexander Rodriguez* Georgia Tech
- Nikhil Muralidhar* Virginia Tech
- Bijaya Adhikari The University of Iowa
- Anika Tabassum Virginia Tech
- Naren Ramakrishnan Virginia Tech
- B. Aditya Prakash Georgia Tech

*Equal contribution

AAAI Conference on Artificial Intelligence (AAAI-21)
February 2-9, 2021

Georgia Tech

Outline

- Motivation
- Problem Formulation
- Approach
- Results and discussion
- Conclusion and future work

Georgia Tech

Flu in Our Society

- Every year
  - Millions get infected
  - Hundreds of thousands get hospitalized
  - Thousands die
- Surveillance and forecasting methods are key
  - Planning
  - Designing countermeasures
- Only way to directly surveillance flu is through a virological test
  - Costly, and very few testing stations
- Instead, we use Influenza like illness (ILI) reported by hospitals
  - Symptomatic data
  - ILI = fever (temperature of 100°F [37.8°C] or greater) and a cough and/or a sore throat without a KNOWN cause other than influenza.

Flu Surveillance Systems

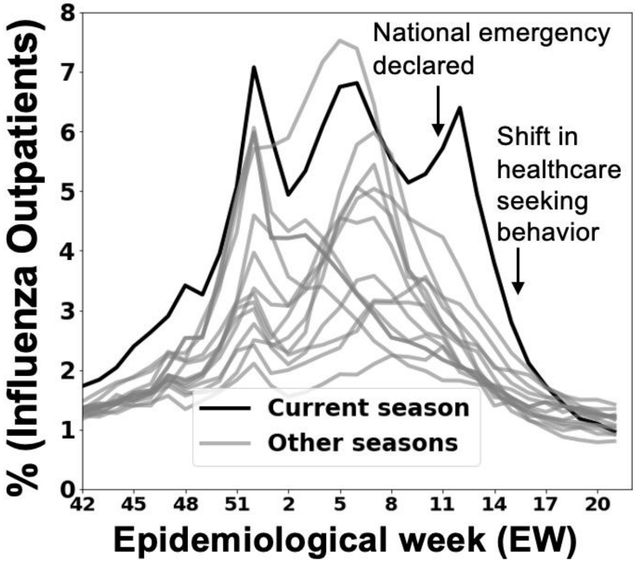
- What is weighted influenza like illness (wILI)?
  - Department of Health and Human Services (HHS) divides the country into 10 regions.
    - Each region has a separate wILI incidence count, which is a weighted average
    - 1 national region depicting overall wILI trend
- <u>Effect of COVID</u>: contamination of COVID in the flu due to symptomatic similarities
- <u>March 2020</u>: Region 2, 9, 10 initially emerged as COVID-19 hot-spots

- Region 1-Boston: Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont.
- Region 2-New York: New Jersey, New York, and the territories Puerto Rico and the Virgin Islands. (Data for territories are not included in calculations on PeriStats.)
- Region 3-Philadephia: Delaware, District of Columbia, Maryland, Pennsylvania, Virginia, and West Virginia.
- Region 4-Atlanta: Alabama, Florida, Georgia, Kentucky, Mississippi, North Carolina, South Carolina, and Tennessee.
- Region 5-Chicago: Illinois, Indiana, Michigan, Minnesota, Ohio, and Wisconsin.
- Region 6-Dallas: Arkansas, Louisiana, New Mexico, Oklahoma, and Texas.
- Region 7-Kansas City: Iowa, Kansas, Missouri, and Nebraska.
- Region 8-Denver: Colorado, Montana, North Dakota, South Dakota, Utah, and Wyoming.
- Region 9-San Francisco: Arizona, California, Hawaii, Nevada and the territories American Samoa, Commonwealth of the Northern Mariana Islands, Federated States of Micronesia, Guam, Marshall Islands, and Republic of Palau. (Data for territories are not included in calculations on PeriStats.)
- Region 10-Seattle: Alaska, Idaho, Oregon, and Washington.

Rodríguez, et al., 2020

A Novel Forecasting Setting

- Symptomatic similarities between these two illnesses and change in patient's behavior affects our current surveillance systems.
- wILI counts may be affected by
  - COVID "contamination"
  - Shift in healthcare seeking behavior during the pandemic
- This new scenario lead us a novel forecasting problem: **COVID-ILI forecasting**

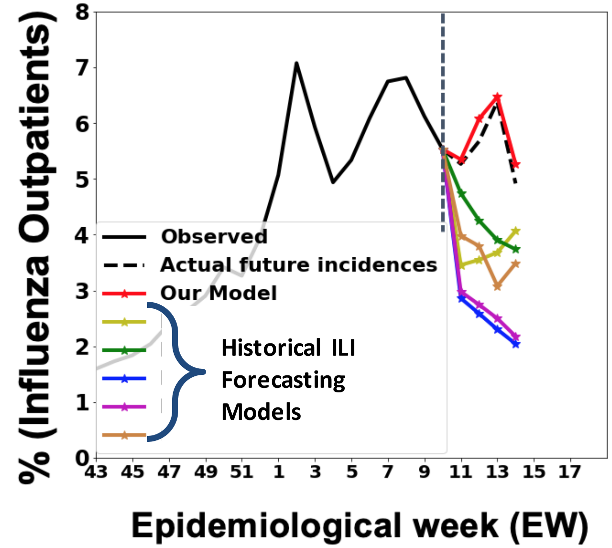


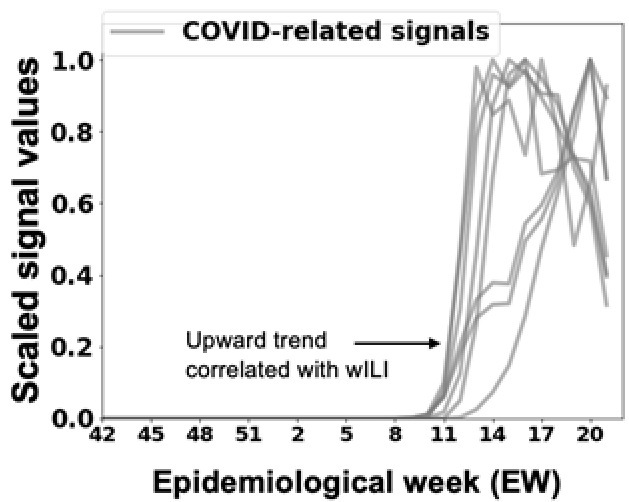
How Forecasting COVID-ILI is Useful?

- Forecasts the actual burden to hospitals
  - Helpful for resource allocation and healthcare worker deployment
- Can also be used to help with indirect COVID surveillance (Castrofino et al. 2020; Boëlle et al. 2020)
  - Especially useful at the early stages of the pandemic, when there were no well-established surveillance mechanisms for COVID
- Disambiguate trends between historical strains and new emerging strains during a flu season

Challenges

- Historical data alone is inadequate to represen t the current scenario
- Traditional ILI models are unable to adapt
- We propose to:
  - Use patterns from historical ILI
  - Leverage COVID-related signals (limited in size)





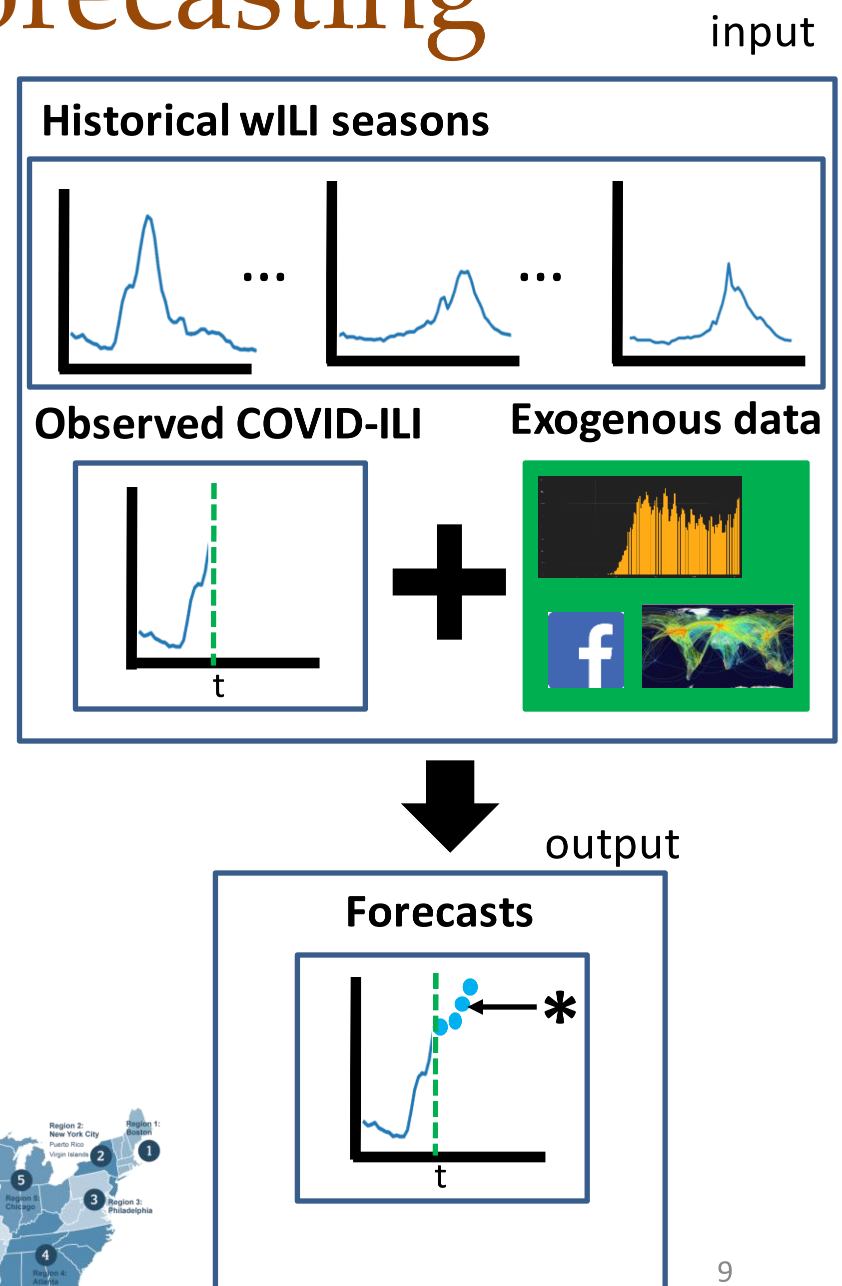
Rodríguez, et al., 2020

Outline

- Motivation
- Problem Formulation
- Approach
- Results and discussion
- Conclusion and future work

COVID-ILI Forecasting

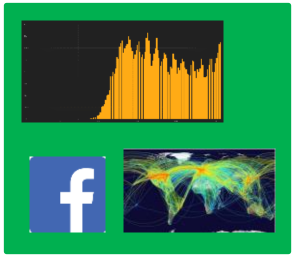
- **Given**
  - Historical wILI seasons
  - Partially observed COVID-ILI incidence curve $Y_c = \{y_c^1, y_c^2, \ldots y_c^t\}$ till week t for each region r.
  - A set of COVID-related features observed till week t for all regions.
- **Predict**
  - Future incidence for next four observations $\{y_c^i\}_{i=t+1}^{t+4}$ for each region r.



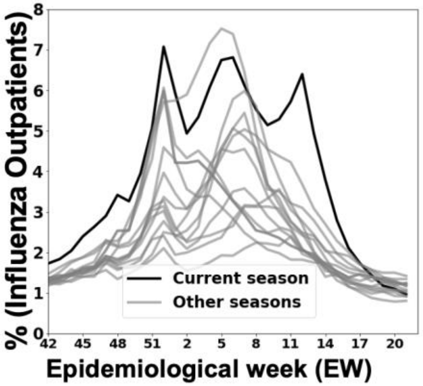
Technical Challenges

- Covid related data is very sparse
- Historical wILI data is rich but does not represent any of the effects of COVID
- How to exploit spatial correlation in exogeneous signals?
- How to leverage **both** the rich **historical wILI data** and highly **informative sparse features** for COVID-ILI forecasting?
- Historical wILI and COVID-related signals are **asynchronous**

Exogenous data



wILI (Current and Past Seasons)

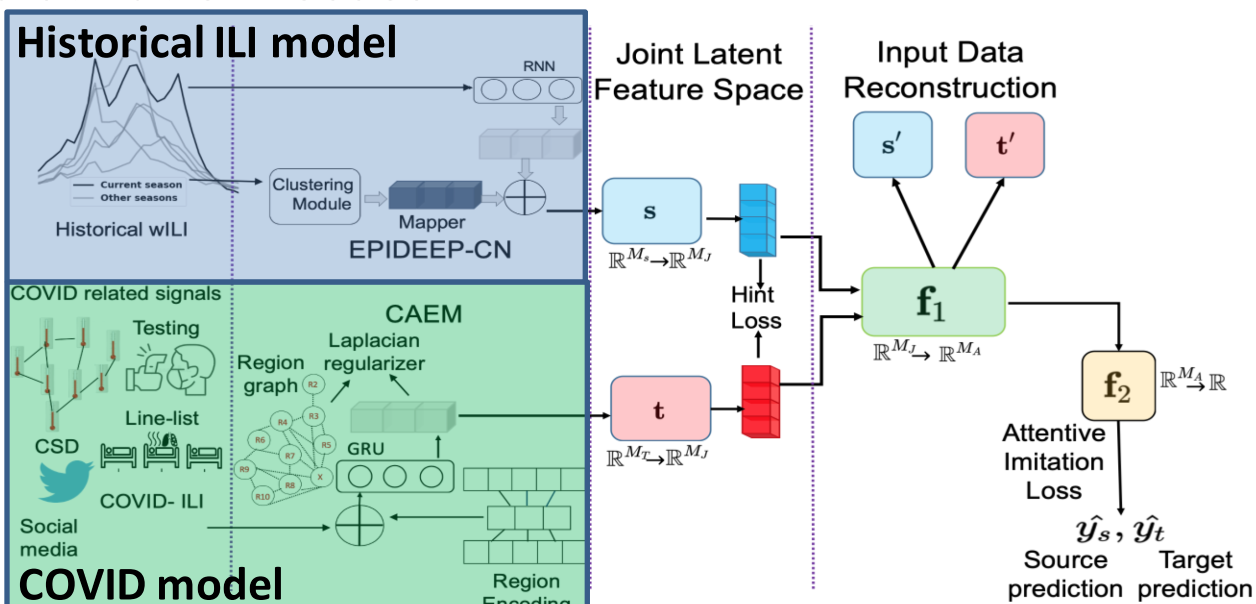


Outline

- Motivation
- Problem Formulation
- Approach
- Results and discussion
- Conclusion and future work

Covid-Augmented ILI Forecasting Network (CALI-Net)

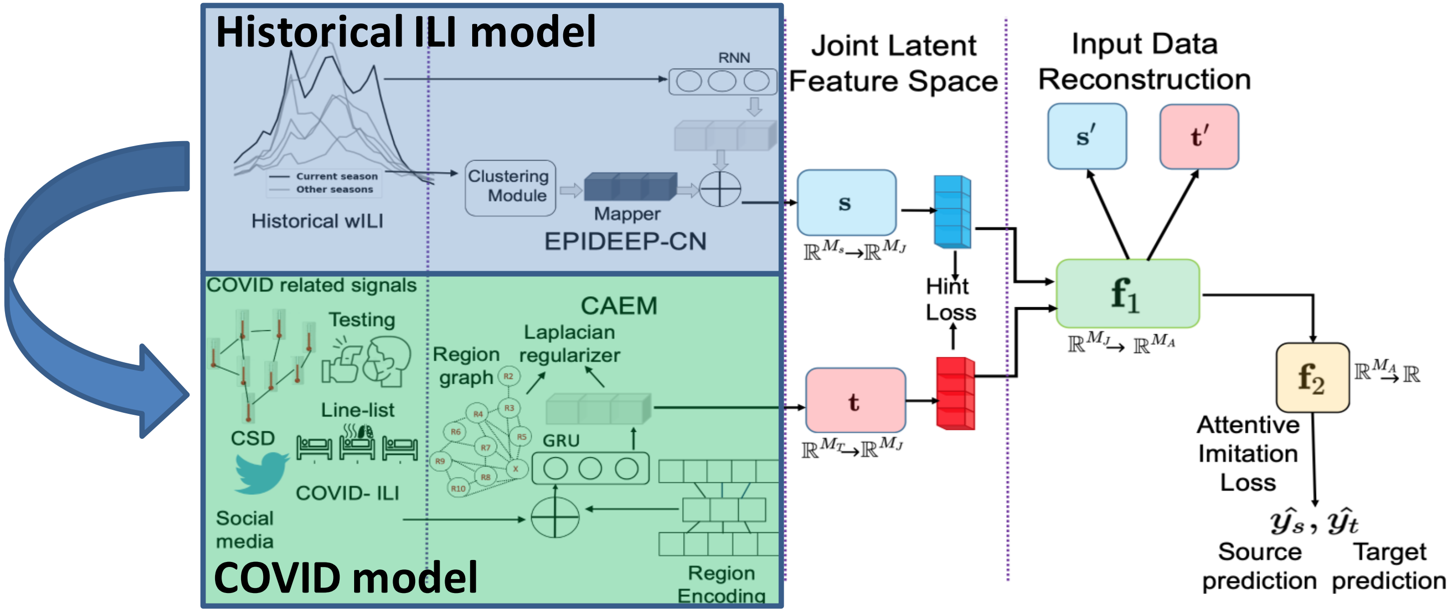
- Steer an existing historical ILI model (EpiDeep, KDD 2019) with new COVID-related signals
- Goal: enable structured knowledge transfer from our historical ILI model to a spatio-temporal COVID-ILI model
- We use heterogenous transfer learning and knowledge distillation losses



High-level abstraction

Rodríguez, et al., 2020

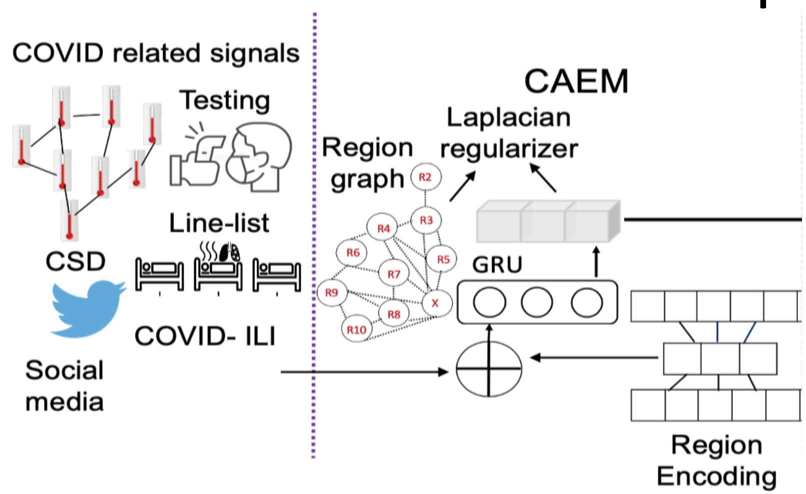
Exploiting Learned Representations from Historical wILI



- We propose to address this problem as a heterogeneous transfer learning (HTL) problem, we adapt the HTL framework of Moon and Carbonell, 2017
- Knowledge extracted from historical wILI and from COVID-related signals are projected to a shared latent space
- Use of denoising autoencoder to improve representations

COVID-Augmented Exogenous Model (CAEM)

- Global model with joint modeling of data from all regions (10 HHS regions + National)
- Region specific embeddings (one-hot encoding + autoencoder)
- Laplacian regularization exploiting regional inter-dependencies
- Recurrent architecture to model temporal evolution



Rodríguez, et al., 2020

Attentive Knowledge Distillation (KD) Losses

- To better structure the knowledge transfer, we propose to incorporate KD losses; they encourage positive transfer between the COVID model (CAEM) and the historical model (EpiDeep)
- Attention in the KD losses (Saputra et al. 2019) automatically prevent negative transfer

Student (CAEM) prediction

Student (CAEM) intermediate representation

$$\mathcal{L}_{KD} = \alpha \frac{1}{n} \sum_{i=1}^{n} \Phi_i \underbrace{\|\hat{y}_s - \hat{y}_t\|_i^2}_{\mathcal{L}_{\text{Im}}} + \Phi_i \underbrace{\|\Psi_s - \Psi_t\|_i^2}_{\mathcal{L}_{\text{Hint}}}$$

Teacher (Epideep) prediction

Teacher (Epideep) intermediate representation

Imitation loss

Hint loss

Calculating Attention in KD Loss

$$\Phi_i = \left(1 - \frac{\|\hat{y}_s - y\|_i^2}{\eta}\right)$$

$$\eta = \max(e_T) - \min(e_T)$$

$$e_s = \left\{\|y - \hat{y}_s\|_j^2 : j = 1, \ldots, N\right\}$$

Outline

- Motivation
- Problem Formulation
- Approach
- Results and discussion
- Conclusion and future work

Experiment Setup

- We divide the forecasting period in two:
  - T1: period of non-seasonal rise of wILI due to contamination by COVID-19 (EWs 9-11)
  - T2: period when COVID-ILI trend is declining more in tune with the wILI pattern (EWs 12-15)
- Metric: RMSE
- Results presented for next incidence prediction, more in appendix

Georgia Tech

Datasets

- wILI data collected by CDC and publicly available
- COVID-related signals collected from multiple public sources

Table 1: Overview of COVID-Related Exogenous Data.

Type of signal	**Description**	**Signals**	**Source**
(DS1) Line list based	They are a direct function of the disease spread	1. Confirmed cases; 2. UCI beds; 3. Hospitalizations; 4. People on ventilation; 5. Recovered; 6. Deaths; 7. Hospitalization rate; 8. ILI ER visits; 9. CLI ER visits	(COVID-Tracking 2020; CDC 2020) (JHU 2020)
(DS2) Testing based	Related to social policy and behavioral considerations	10. People tested; 11. Negative cases; 12. Emergency facilities reporting; 13. No. of providers;	(COVID-Tracking 2020; CDC 2020)
(DS3) Crowdsourced symptoms based	Crowdsourced symptomatic data from personal devices	14. Digital thermometer readings;	(Miller et al. 2018)
(DS4) Social media	Social media activity	15. Health Related Tweets	(Dredze et al. 2014)

Rodríguez, et al., 2020

Baselines

Recent historical wILI models (Reich et al. 2019):

- **Delta-Density**
  - Kernel conditional density estimation, a non-parametric statistical methodology that is a distribution-based variation on nearest-neighbors regression
- **Empirical Bayes**
  - Model past seasons' epidemic curves as smoothed versions plus noise.
  - Construct prior for the current season's epidemic curve by considering sets of transformations of past seasons'curves
- **SARIMA**
  - Autoregressive Integrated Moving Average model with seasonality

Also, **HIST**, a persistence baseline based on weekly average of the historical seasons

Experimental Questions

Transfer Learning

- Q1. Is CALI-NET able to achieve successful positive transfer to model the contamination of wILI values?
- Q2. Does CALI-NET prevent negative transfer by automatically recognizing when wILI and COVID-19 trends deviate?

Forecasting Performance

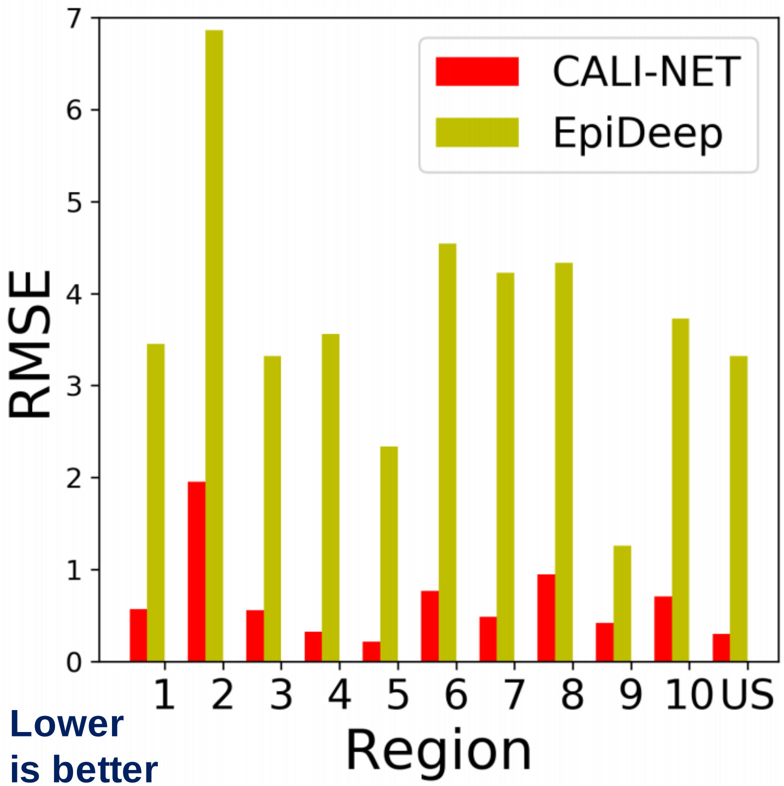
- Q3. Does CALI-NET's emphasis on transfer learning sacrifice overall performance with respect to state-of-the-art methods?

Ablation Studies

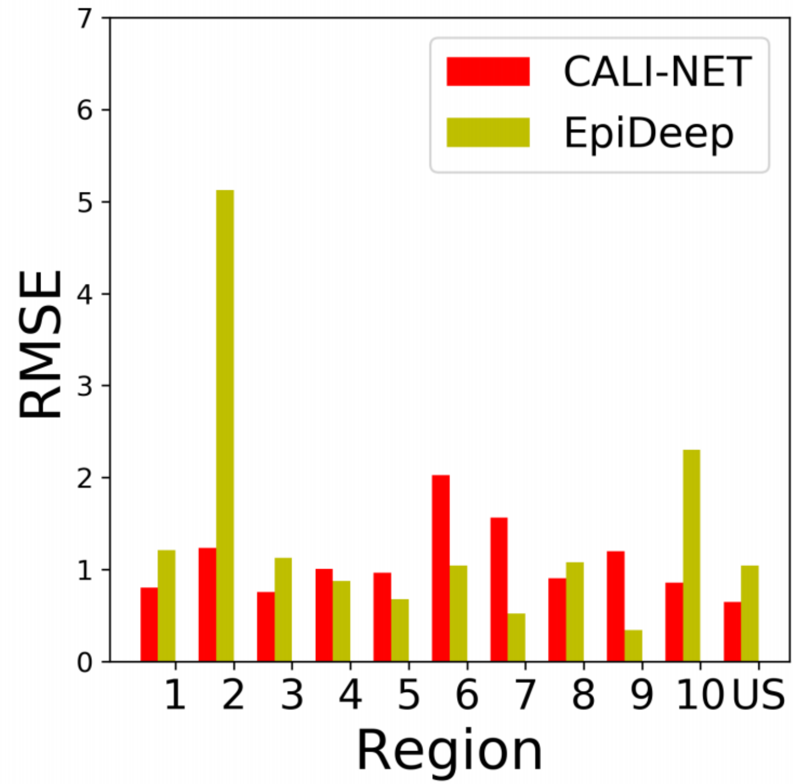
- Q4. How does each facet of CALI-NET affect COVID-ILI forecasting performance?
- Q5. What data signals are most relevant to COVID-ILI forecasting?

Transfer Learning Results (compare vs historical ILI model)

Leverage positive transfer



Prevent negative transfer



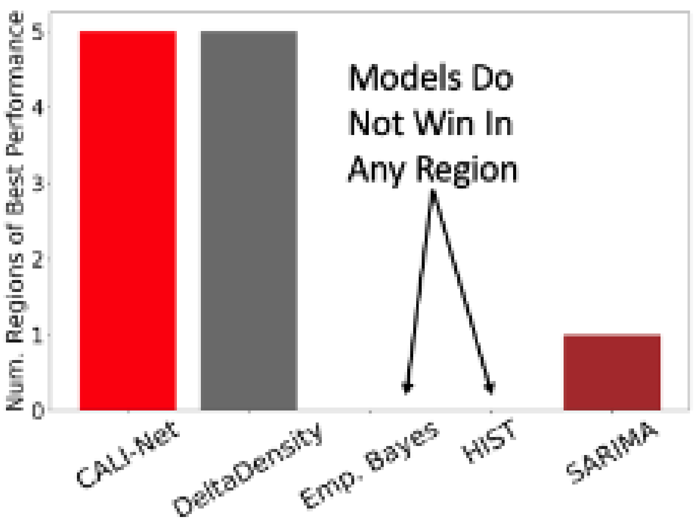
Lower is better

Forecasting performance during period of increasing COVID-ILI leading to unseasonal peak

Forecasting performance during period of declining COVID-ILI trend i.e., return to traditional dynamics.

Rodríguez, et al., 2020

Overall Performance: Emphasis in Adaptation Doesn't Compromise It



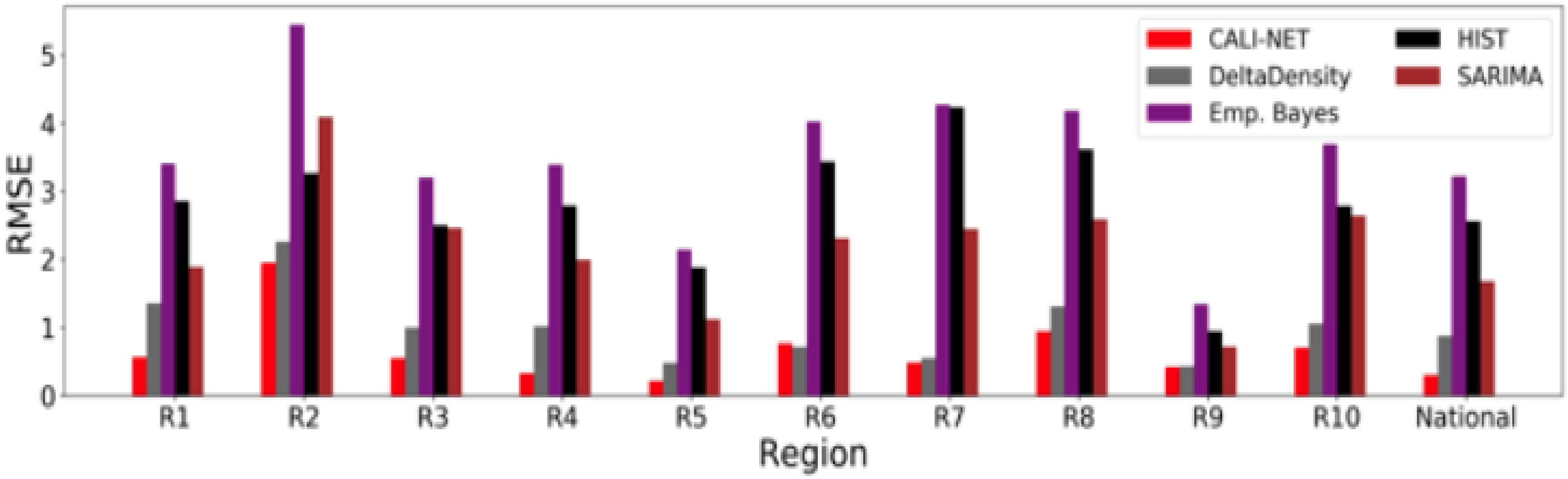
Histogram of Number of Regions where each Model is the Best Performing One

- Overall model performance across period $T_1 + T_2$
- CALI-NET outperforms other models in 5 out of 11 regions, on par with DeltaDensity, a SOTA model
- CALI-NET yields to competitive performance across the entire course $T_1 + T_2$

Rodríguez, et al., 2020

Focusing on Period T1 (uptake)

Performance Characterization in Period T1



CALI-Net outperforms all models in 9 out of 11 regions for positive transfer phase T1 where COVID-19 contamination of wILI is the greatest.

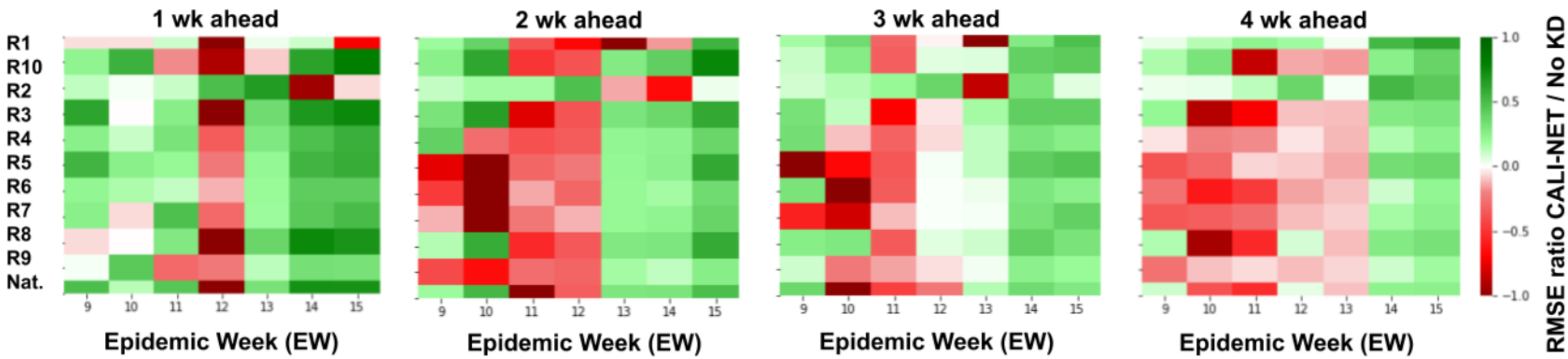
Rodríguez, et al., 2020

Model Ablation

Table 1: Per-region RMSE performance characterization of the CALI-NET model when different components of CAEM architecture are deactivated.

Regions	Our Method	CALI-NET w/o GRU	CALI-NET w/o Laplacian	CALI-NET w/o Regional Recon.
R1	0.9196	22.0574	**0.9118**	0.9161
R2	2.6869	9.2662	**2.6843**	2.6977
R3	**1.293**	13.2952	1.3647	1.2965
R4	**1.6605**	6.9054	1.7944	1.7345
R5	**1.5879**	16.1975	1.687	1.6532
R6	**2.93**	7.8045	3.0516	2.951
R7	**2.2805**	5.7593	2.4184	2.322
R8	**1.3774**	14.9026	1.4898	1.3949
R9	1.8225	4.7056	**1.8099**	1.8714
R10	**1.2069**	6.2994	1.2578	1.2262
National	**1.3308**	9.9319	1.4597	1.4141

Effect of Knowledge Distillation



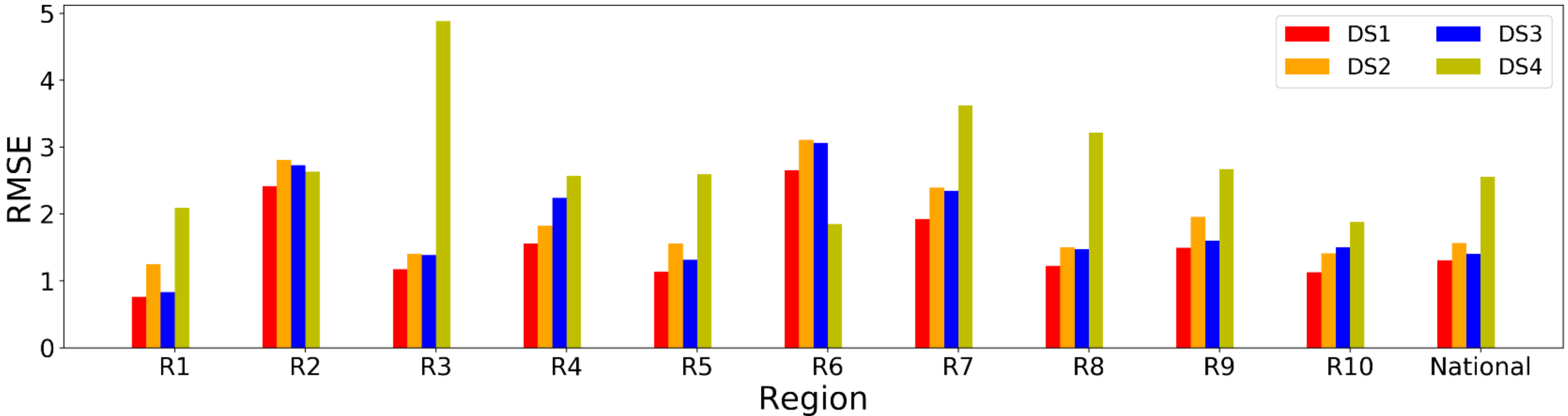
Ratio of RMSE of CALI-NET with vs without knowledge distillation losses

KD helps improve predictions

KD is not helping

Knowledge distillation (KD) is helpful in most of the regions/weeks, especially in short term forecasting and in T2 (i.e., helping to prevent negative transfer)

Data Ablation



DS1: Line list data
DS2: Testing data
DS3: Crowdsourced symptomatic data
DS4: Social media

Line list data is the most helpful, followed by crowdsourced and testing. Social media is the least helpful

Rodríguez, et al., 2020

Outline

- Motivation
- Problem Formulation
- Approach
- Results and discussion
- Conclusion and future work

Conclusions and Future Work

- We proposed CALI-Net, a novel framework for principled **transfer of relevant knowledge** from an existing forecasting model (based on rich historical data) to a one relying on relevant but limited recent exogenous signals

- **Characterized CALI-Net performance** at different stages of the wILI season and showcase effectiveness of its transfer learning capabilities

- **Compared CALI-Net to SOTA** and showcase comparable (and in many cases superior) performance of CALI-Net

- Moving forward, we wish to:
  - Automatically differentiate outbreaks of COVID and flu

Thanks!

Pre-print:
https://arxiv.org/abs/2009.11407

Code:
https://github.com/AdityaLab/CALI-Net

Contact:
Alexander Rodríguez
arodriguezc@gatech.edu