COVID-19 SYMPTOM DATA CHALLENGE PHASE 2 - WHITE PAPER

Team Name: DeepOutbreak

Name of Analytic Approach: DeepOutbreak

APPROACH

Patients suffering from COVID-19 and influenza-like illnesses (ILI) are a significant and immediate burden on the healthcare system. It is important to forecast the dynamics of each of these co-evolving diseases to design effective healthcare policies and minimize devastating effects.

Forecasting the spread of COVID-19 is complex due to insufficient representative data and inconsistency in dynamics due to policy effects. On the other hand, forecasting ILI (co-evolving with COVID-19) is also challenging due to its symptomatic similarities with COVID-19 (Zipfel & Bansal 2020) and the change in healthcare seeking behavior of people afflicted with ILI. In fact, during the current pandemic, ILI counts have been useful for estimating unreported COVID cases in the US (Kou et al., 2020) also demonstrated in other countries, e.g. (Castrofino et al., 2020) in Italy, (Boëlle et al.,2020) in France. For the current season, the ILI curve is clearly been affected by the shift in healthcare seeking behavior of outpatients and it is not clear if it will capture the actual influenza activity; however, forecasting this metric gives us important information about the expected burden to the healthcare system, which is relevant for resource planning in these trying times. Thus, to improve situational awareness regarding these two diseases, we propose *DeepOutbreak*, a framework for concurrently forecasting both COVID-19 and ILI activity during the pandemic.

To motivate research into epidemic forecasting, the CDC has been organizing epidemic forecasting challenges in recent years. The Flusight challenge for influenza and COVID-19 Forecast hub for COVID are the two challenges relevant to our work. The Flusight challenge seeks point and probabilistic forecasts for various targets including future incidence and peak value. Similarly, the COVID-19 Forecast hub seeks point and probabilistic predictions for COVID induced mortality and hospitalizations. Both challenges seek forecasts at multiple geographic resolutions including the US national, state-level, and HHS regional level. We now describe the specific sub-tasks addressed en route to addressing each aforementioned challenge.

1. <u>COVID-19 Forecast Hub target tasks:</u>
    a. **Incidence and cumulative weekly deaths.** Specifically, it involves forecasting newly reported COVID induced deaths and the cumulative deaths for all US states and the US national region. Here, the data reported by Johns Hopkins University (JHU) serves as the gold standard for the CDC.
    b. **Daily Hospitalizations.** Here the goal is to forecast the reported new hospitalizations for US states and the US overall. CDC recently released a ground truth dataset for this.

Given these two targets, the problem we solve for COVID-19 Forecast hub for a specific geography can be formally stated as follows.
Covid-19 Forecast hub problem statement: *Given*: an observed multivariate time series of COVID-related signals X for N weeks and corresponding values for the forecasting target Y for the

same N weeks. *Predict*: next k values of forecasting target, i.e. Y[N+1: N+k] where k=4 for the first target (four weeks) and k=28 for the second target (28 days ahead).

Approach Summary: We solve the problem above by designing a deep neural approach with autoregressive inputs which trains on bootstrap samples of the input data. We leverage several data signals including symptomatic survey data, mobility, and testing data to name a few. We describe our methods in more detail in the following section.

2. **Weighted influenza-like-illness (wILI) forecasting target tasks:**
    a. On the other hand, the Flusight challenge has targets in terms of wILI (weighted ILI) counts. The wILI counts are a proxy to the influenza incidence (total of influenza cases in any given region). One of the key targets in the challenge is to forecast the next four wILI values.

wILI forecasting problem statement: *Given:* the wILI incidence \`Y\` (also our forecasting target) till week N in the form of a multivariate time series of wILI (and in our case COVID-related signals X), *predict:* wILI values for weeks N+1 to N+k, where k = 4.

Approach Summary: Predicting wILI values in the presence of COVID is a challenging task. Due to symptomatic similarities between influenza and COVID and changes in health-care seeking behavior of the general public, the wILI values become 'contaminated' due to COVID. This contamination results in the wILI curve exhibiting unseasonal dynamics and peaking at unexpected times. Hence, leveraging an out-of-the-box wILI forecasting approach trained on historical wILI data does not work well. In an effort to model this novel unseasonal behavior, we propose to \`steer\` a historical wILI forecasting model with covid related signals by leveraging transfer learning. In addition to the historical wILI values, we use several types of data signals such as symptomatic surveillance, mobility, line-lists and so on to train our model. The approach is described in more detail below.

METHODS

Our approach is based on two forecasting modules:

1. **DeepCOVID: Covid-19 forecasting using Covid-related signals.**

    i. Forecasting tasks: Covid-related mortality and hospitalizations
    ii. Challenges: Coping with heterogeneous, scarce and noisy data.

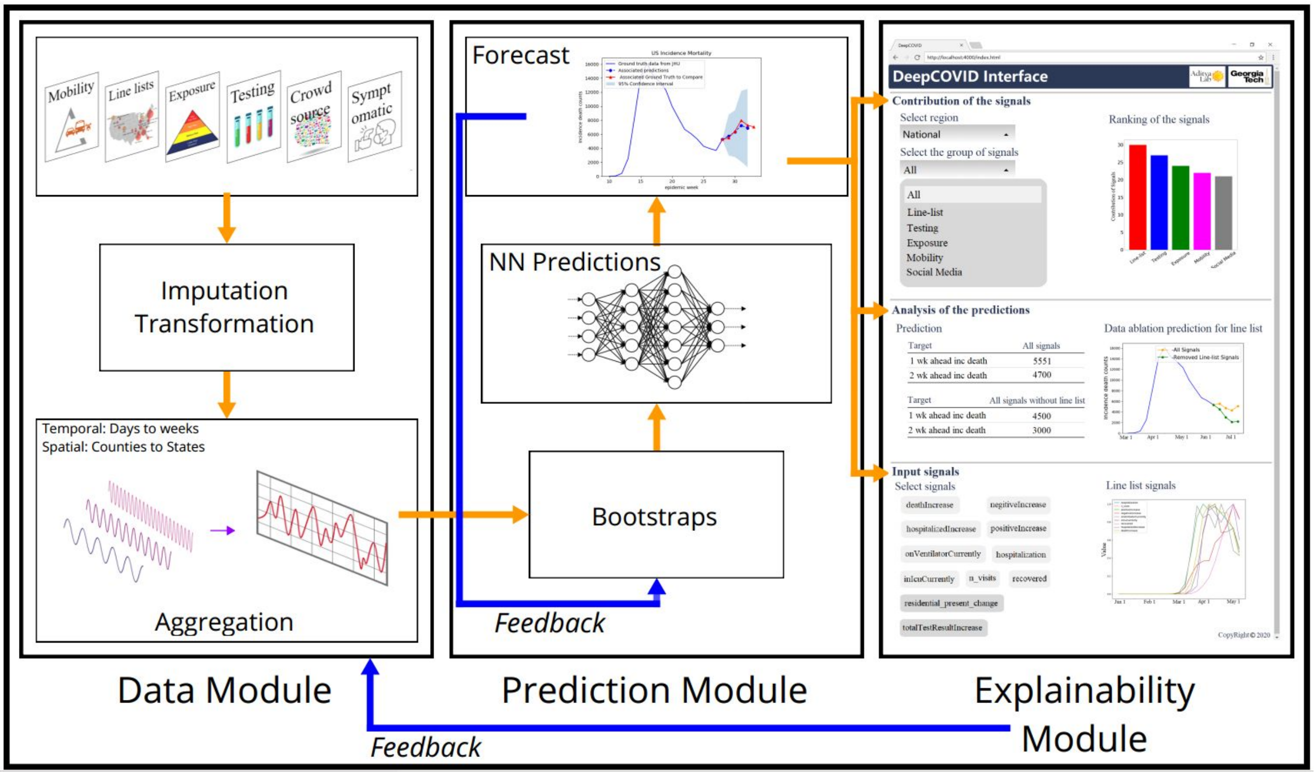


Figure 1: Schematic of DeepCOVID framework for real-time COVID-19 forecasting. The data module is dedicated to pre-processing including imputation of missing values and aggregating at the right temporal and spatial resolution. The prediction module generates probabilistic forecasts based on the curated data. Finally, the explainability module (with interface) allows both the real-time and retrospective analysis of forecasts to build an intuitive explanation of forecasts.

iii. Solution: A deep learning (DL)-based model can ingest many heterogeneous signals that are more sensitive to what is happening on the ground, without laborious feature engineering. To fully take advantage of this, our framework is designed with careful consideration of data and modeling challenges faced in robust real-time forecasting with principled uncertainty estimation. Secondly, uncertainty estimation also enables effective communication of predictions to domain-experts by giving explanations for model forecasts, which are very important for communication and interpretation by both the public and decision makers.
DeepCOVID (architecture depicted in Fig. 1) is an operational DL driven framework for real-time COVID forecasting, whose predictions have been submitted to the CDC via the hub on a weekly basis since April 2020. Our method exhibits interpretability, encouraging short-term and trend performance, principled uncertainty estimation, correlation between forecasts, and ingestion of several data sources despite the chaotic and fast-moving pandemic scenario which naturally brings with it several modeling and data challenges. This was the first purely data driven and deep learning approach to be submitted to the COVID-19 Forecasts hub.

2. **CALI-Net: Covid-ILI forecasting using historical influenza data + Covid-related signals**
   i. Forecasting task: wILI incidence forecast
   ii. Challenges: The first challenge is to capture the atypical trends that occur as a result of COVID-19 contamination of wILI. In modeling this novel wILI trend, we propose to leverage

historical knowledge and recent COVID-related data signals. Hence, the second challenge is how to effectively model the COVID-ILI (wILI contaminated by COVID) curve by appropriately leveraging both historical and recent data. Note that these COVID-related data signals are not available for historical wILI seasons. How do we address the imbalance in data to leverage both of these data sources? Further, as the contaminated COVID-ILI is a very new phenomenon which suddenly emerged, there is limited data regarding the same from external signals and hence, a significant challenge is also to learn to model it effectively under data paucity.

CALI-Net: Covid-ILI forecasting using historical influenza data + Covid-related signals

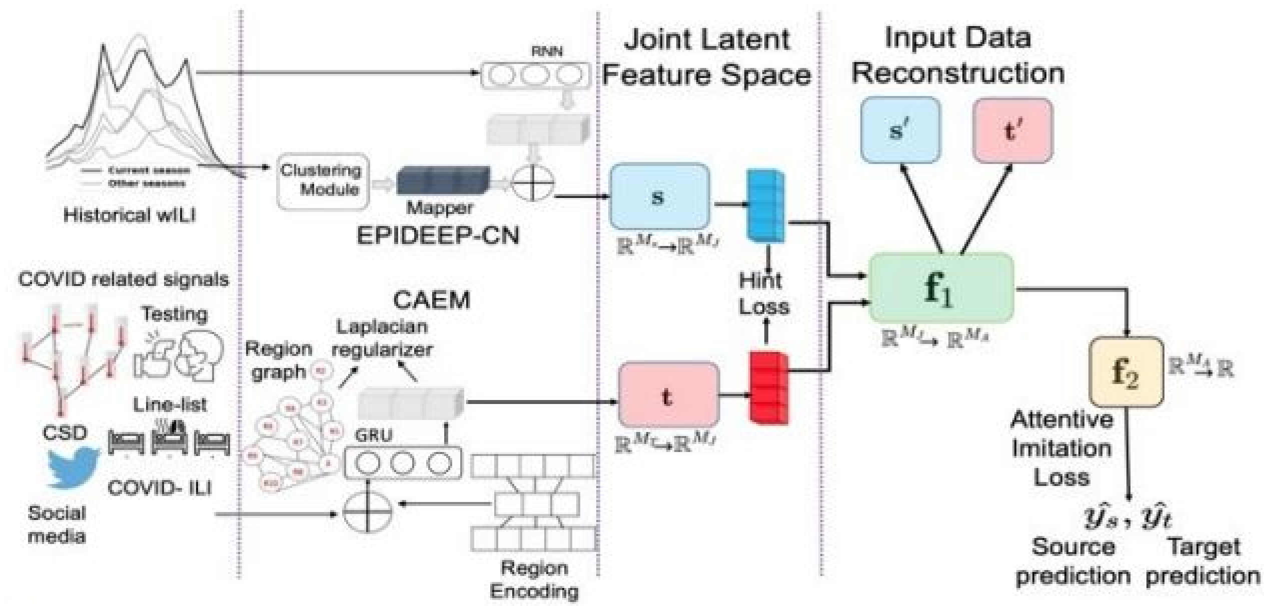


Figure 2: Our heterogeneous transfer learning architecture is designed to transfer knowledge from EpiDeep-CN (source model trained on historical wILI data) about historical wILI trends to the CAEM module (target model trained on COVID-related exogenous signals) for COVID-ILI forecasting while addressing the challenges of data paucity.

iii. **Solution**: As mentioned above, one of the challenges we face is the atypical nature of the current influenza season because of contamination by COVID related dynamics resulting in the COVID-ILI trends. Since this feature is exclusive to the current introduction of the pandemic into the wILI season, using only the historical wILI data is insufficient. Hence we propose to leverage external COVID-related signals such as confirmed cases, hospitalizations, and emergency room visits as well and propose a novel transfer learning framework with novel loss functions and specifically design DL model architectures to deal with the data paucity issue. We propose CALI-Net (COVID Augmented ILI deep Network), a principled way to 'steer' flu-forecasting models to adapt to new scenarios where flu and COVID co-exist. The full CALI-Net architecture is detailed in Fig. 2. We employ transfer learning and knowledge distillation approaches to ensure effective transfer of knowledge of historical wILI trends. We incorporate multiple COVID-related data signals all of which help capture the complex data contamination process showcased by COVID-ILI. Finally, in order to alleviate the data paucity issue, we train a single global architecture with explicit spatial constraints to model COVID-ILI trends of all regions as opposed to previous approaches which have modeled each region separately leading to a superior forecasting performance.

Research questions:

Our goal for the challenge is to investigate different facets of the contribution of survey signals in our COVID and wILI (aka. COVID-ILI) predictions.

Q1: Should survey signals be used in conjunction with other data or by itself?
Q2: Are survey signals confirming or orthogonal when used in conjunction with others?
Q3: Do survey signals help in forecasting trend changes with some weeks of anticipation?
Q4: Do survey signals capture important differences between geographical regions?

RESULTS

Summarize your results:

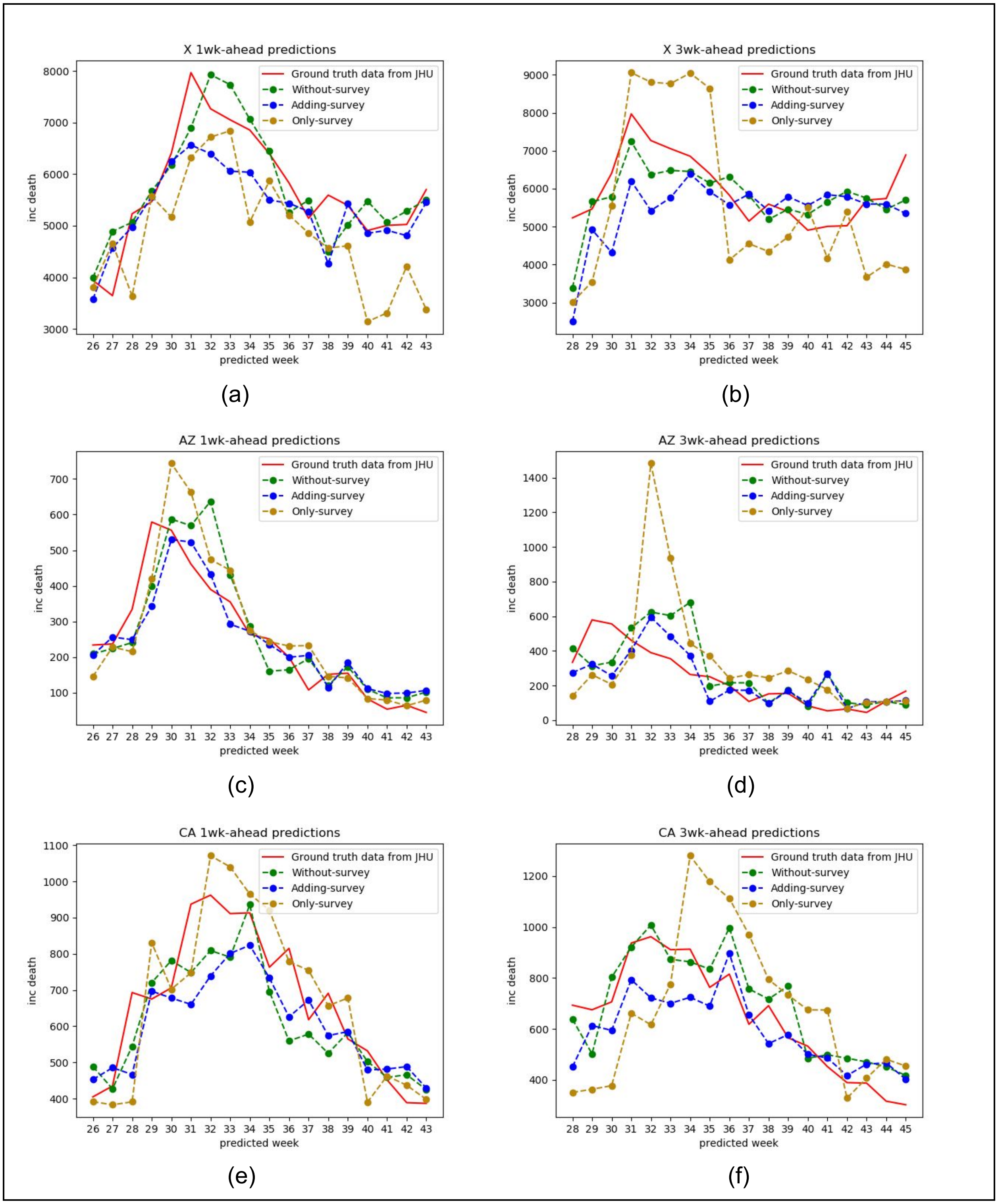
In this section we first focus on the research questions introduced in the previous section to emphasize the contribution of symptom survey signals; later showcase the competitiveness of our method in the two forecasting tasks.

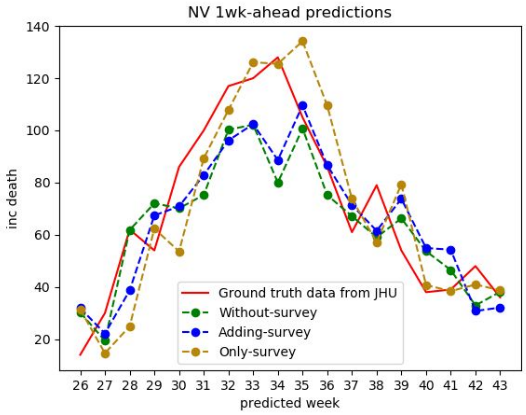
For results addressing research questions from above, our forecasts start at epidemic week 25 (June) to week 42 (October); therefore, these results are considering predictions over a period of 4 months. The predictions labeled as “Without-survey” are our actual weekly submissions to the CDC, while “Adding-survey” and “Only-survey” were recently performed emulating a real-time forecasting scenario (using only data available until the prediction week).

The specific survey signals we added were the raw_cli and raw_wili, both available in COVIDCast.

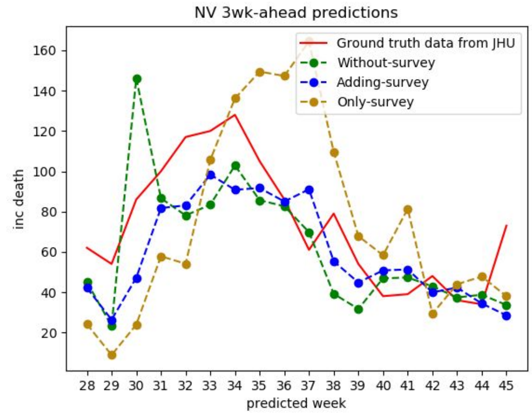
Q1: Should survey signals be used in conjunction with other data or by itself?

> For this question, we tested the usefulness of symptom survey data independently and in conjunction with other signals for COVID-19 mortality forecasting and found that symptom survey data is more effective when used in conjunction with other signals. In short-term forecasting cases (1- 2-week ahead forecasting), models using only symptom survey data performed comparably well with respect to models using symptom survey signals in conjunction with other signals as observed in the left column of Fig. 3. However, we noticed a pattern of overestimation in longer-term forecasting of COVID-19 mortality (see right column of Fig. 3). This suggests that we should use the symptom survey signals in conjunction with others.
>
> It is useful to observe that short-term forecasts from only using survey data might be especially useful when there is an uptrend occurring. We found that they may lead to a better estimation than when using them in conjunction, as it happens in California and Nevada (see Fig. 3e and 3g).
>
> **Conclusion:** In general, survey signals should be used in conjunction with other data, but their standalone use should be considered when there are uptrends as it may lead to better <u>short-term</u> predictions of the disease evolution during this critical stage.


X 1wk-ahead predictions
Ground truth data from JHU
Without-survey
Adding-survey
Only-survey
inc death
predicted week
(a)
X 3wk-ahead predictions
Ground truth data from JHU
Without-survey
Adding-survey
Only-survey
inc death
predicted week
(b)
AZ 1wk-ahead predictions
Ground truth data from JHU
Without-survey
Adding-survey
Only-survey
inc death
predicted week
(c)
AZ 3wk-ahead predictions
Ground truth data from JHU
Without-survey
Adding-survey
Only-survey
inc death
predicted week
(d)
CA 1wk-ahead predictions
Ground truth data from JHU
Without-survey
Adding-survey
Only-survey
inc death
predicted week
(e)
CA 3wk-ahead predictions
Ground truth data from JHU
Without-survey
Adding-survey
Only-survey
inc death
predicted week
(f)



(g)



(h)

Figure 3: Symptom survey data employed independently and in conjunction with other signals for COVID-19 mortality forecasting in the US National region. Left is 1-week ahead forecasting, and right is 3-weeks ahead. We found using it independently exhibits good short-term performance but usually leads to overestimation of longer-term predictions. From top to bottom: US National (X), Arizona (AZ), California (CA), and Nevada (NV).

Q2: Are survey signals confirming or orthogonal when used in conjunction with others?

For this question, we compared our CDC submissions for COVID forecasting against the predictions of our model as we would have included the symptom survey signals during the forecasting period (i.e. “Adding-survey”). At US National level, we found that including the survey signals led to a decrease in performance by 39%, however, at state-level, the results are more diverse, as noted in Fig. 4. In total, predictions on 29 of the 51 regions (US National + 50 states) were benefited by the incorporation of survey signals.

Conclusion: When used in conjunction with others, survey signals have a different role per geographical region. Nevertheless, in general, they are orthogonal and can benefit forecasting in the majority of regions.

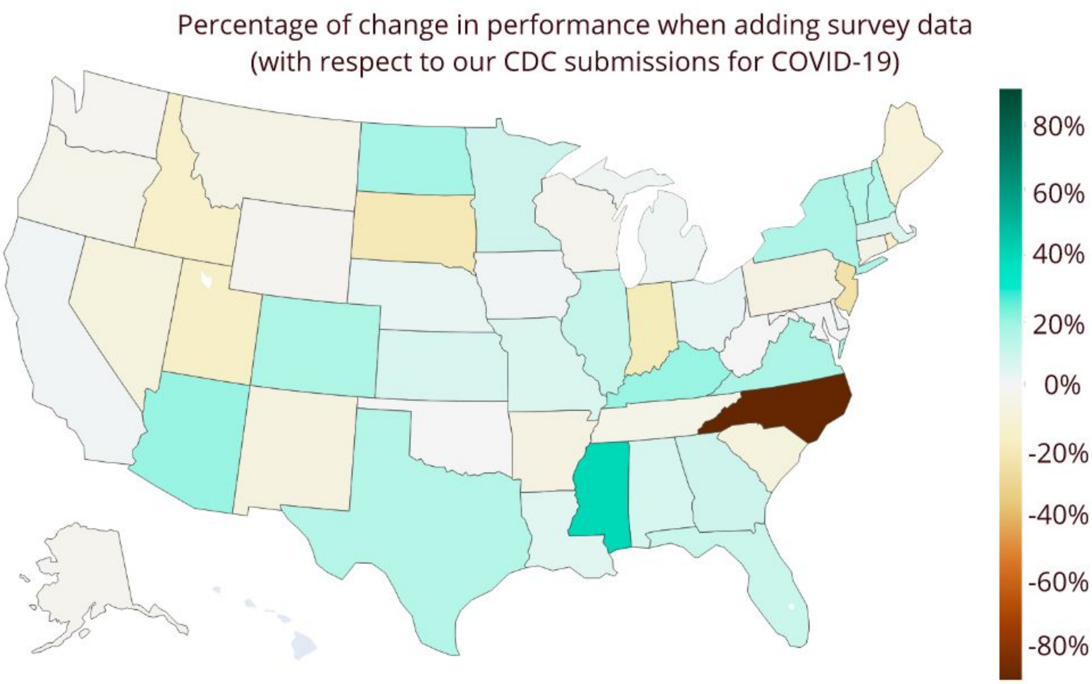


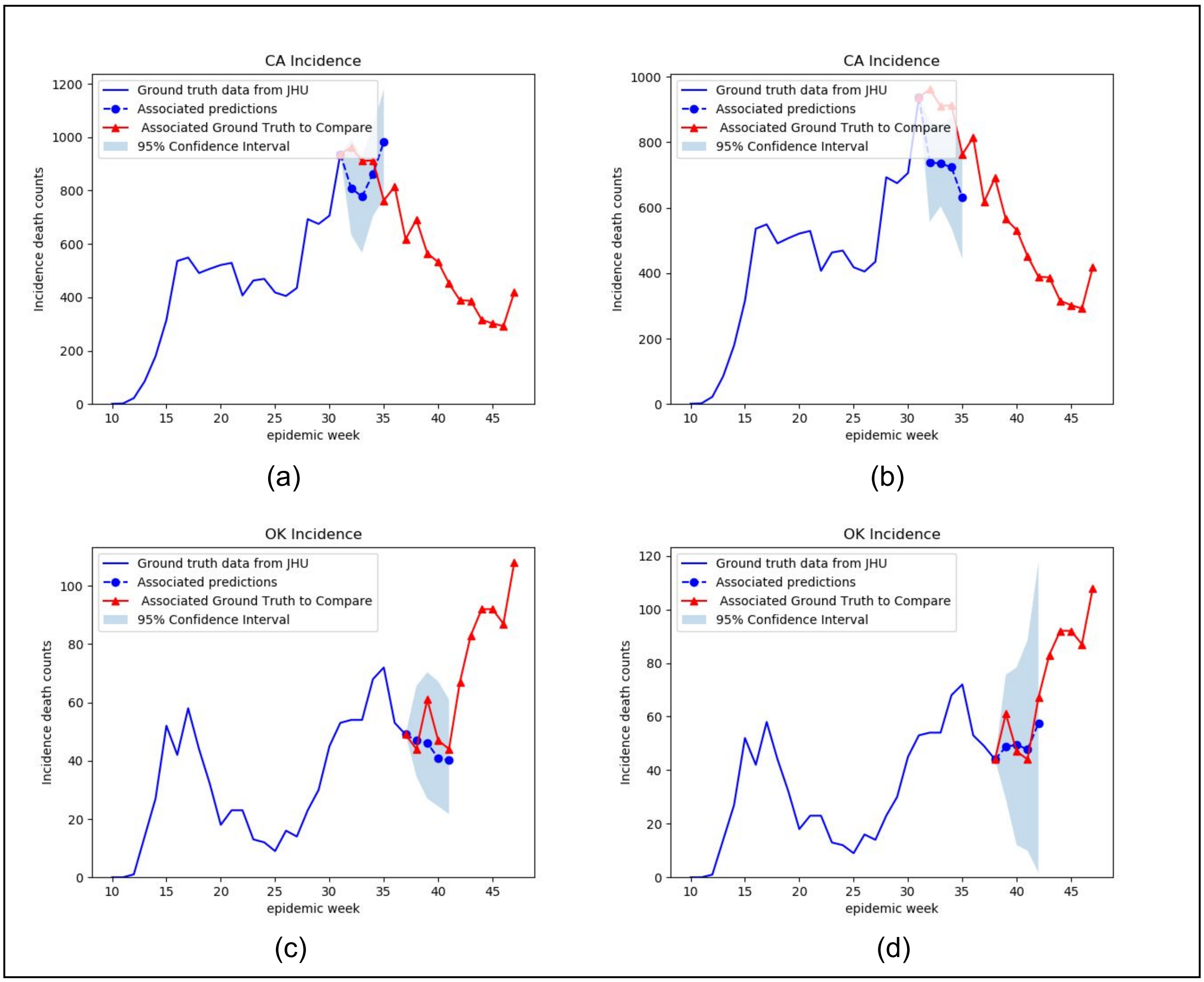
Figure 4: Percentage of change in performance per state (measured with MAE) after adding symptom survey data (i.e. using survey data in conjunction with other data signals). Green (positive) represents increase in performance; brown (negative) decrease.

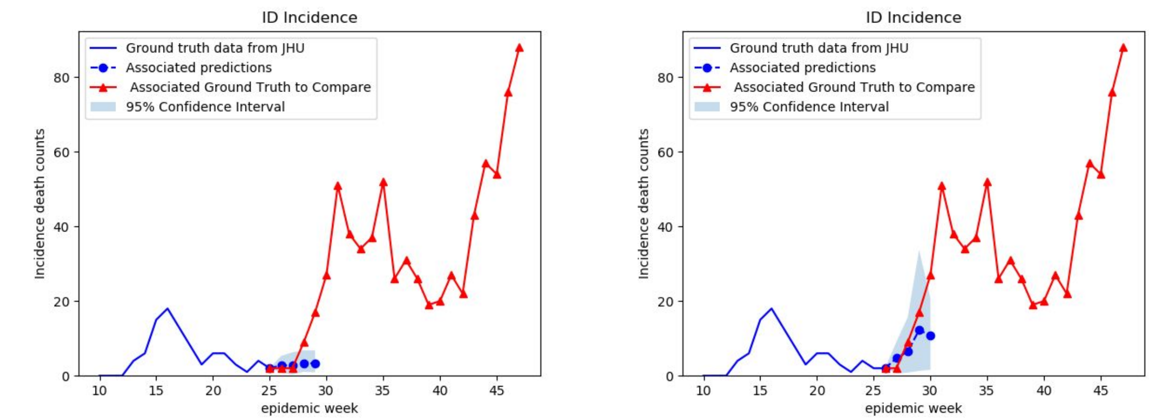
Q3: Do survey signals help in forecasting trend changes with some weeks of anticipation?

For this question our main focus is in predictions with and without survey signals, but we also found some interesting observations when using only survey signals.

Positive findings:

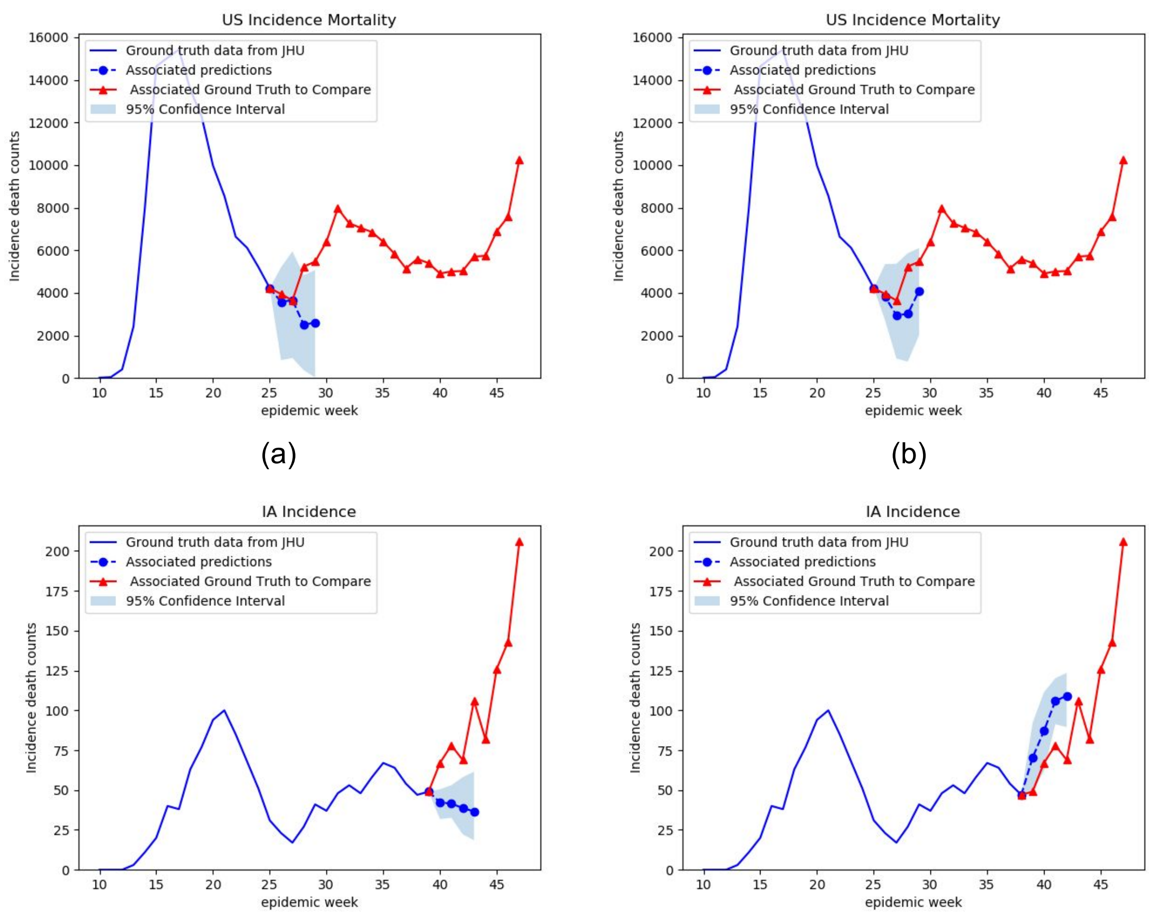
In COVID forecasting, we found several cases where adding survey signals (i.e. using them in conjunction with other data signals) has been beneficial to predict important trend changes as the ones in Fig. 5.





(e) (f)

Figure 5: Left: Without survey signals; right: adding survey signals. We can appreciate that the contribution of survey signals is also reflected helping to anticipate changes in trend. From top to bottom: California (CA), Oklahoma (OK) and Idaho (ID).

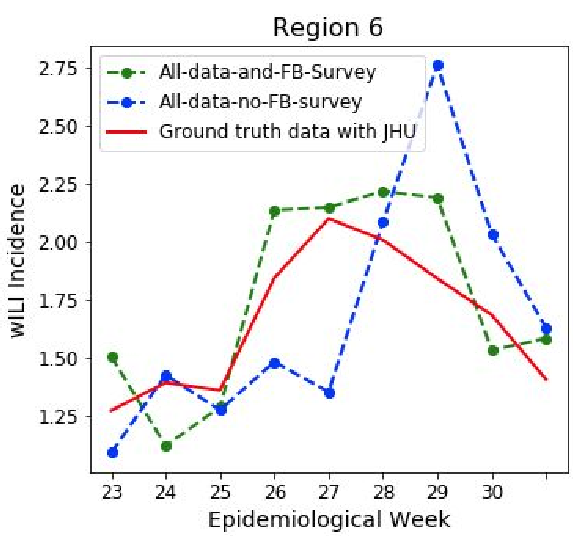


(a) (b)

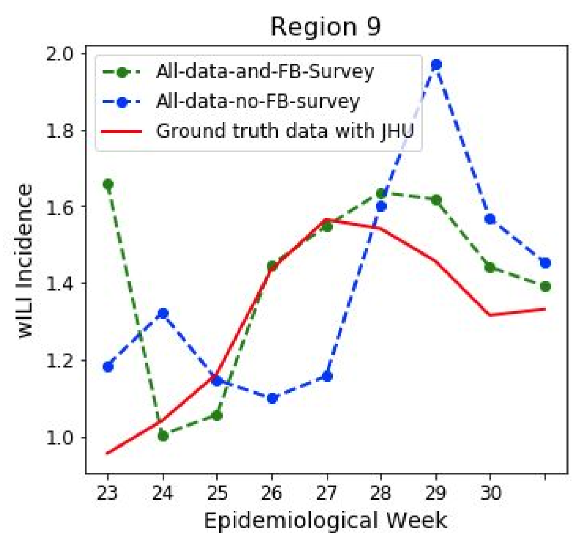
(c) (d)

Figure 6: Left: adding survey signals; right: only survey signals. We also found that, in some cases, some important trend predictions that can be made with only survey signals are hindered by the other signals in our dataset. From top to bottom: US National and Iowa (IA).

In ILI forecasting, in Fig. 7, we notice that in each case, the symptom survey data helps yield better forecasts compared to the model trained without the survey data which is found to either overestimate or underestimate the trends in the ground truth ILI curve. The model trained with the survey data on the other hand is able to anticipate trend changes (Region 6 1-week ahead) and also able to adapt quickly to the rise and fall in ILI dynamics (Region 9 1-week ahead). Models trained without the survey data also tend to overestimate peaks in addition to generally underestimating dynamics for ILI forecasting. Interestingly we also notice that models trained with the survey data are able to adapt quickly to changes in trend (rising, falling) of the ILI dynamics compared to the models trained sans the survey data.



(a) Region 6, 1-week ahead



(b) Region 9, 1-week ahead

Figure 7: We notice in figure 8a. that the model trained with FB survey data is able to anticipate trend changes (like the peak) in region 6 a week in advance utilizing signals in FB survey data. In both figure 8a, 8b, we notice that the model trained with FB survey data is able to adapt much better to the decreasing wILI trend starting around epidemiological week 27.

Negative findings: In our CDC submission, our model predicted the second peak value and time for US National three weeks early (Fig. 8a). However, when adding the survey signals (Fig. 8b), our estimation of the peak height suffered because the model underestimated the peak. Other examples of adversarial effects are illustrated in Fig. 8.

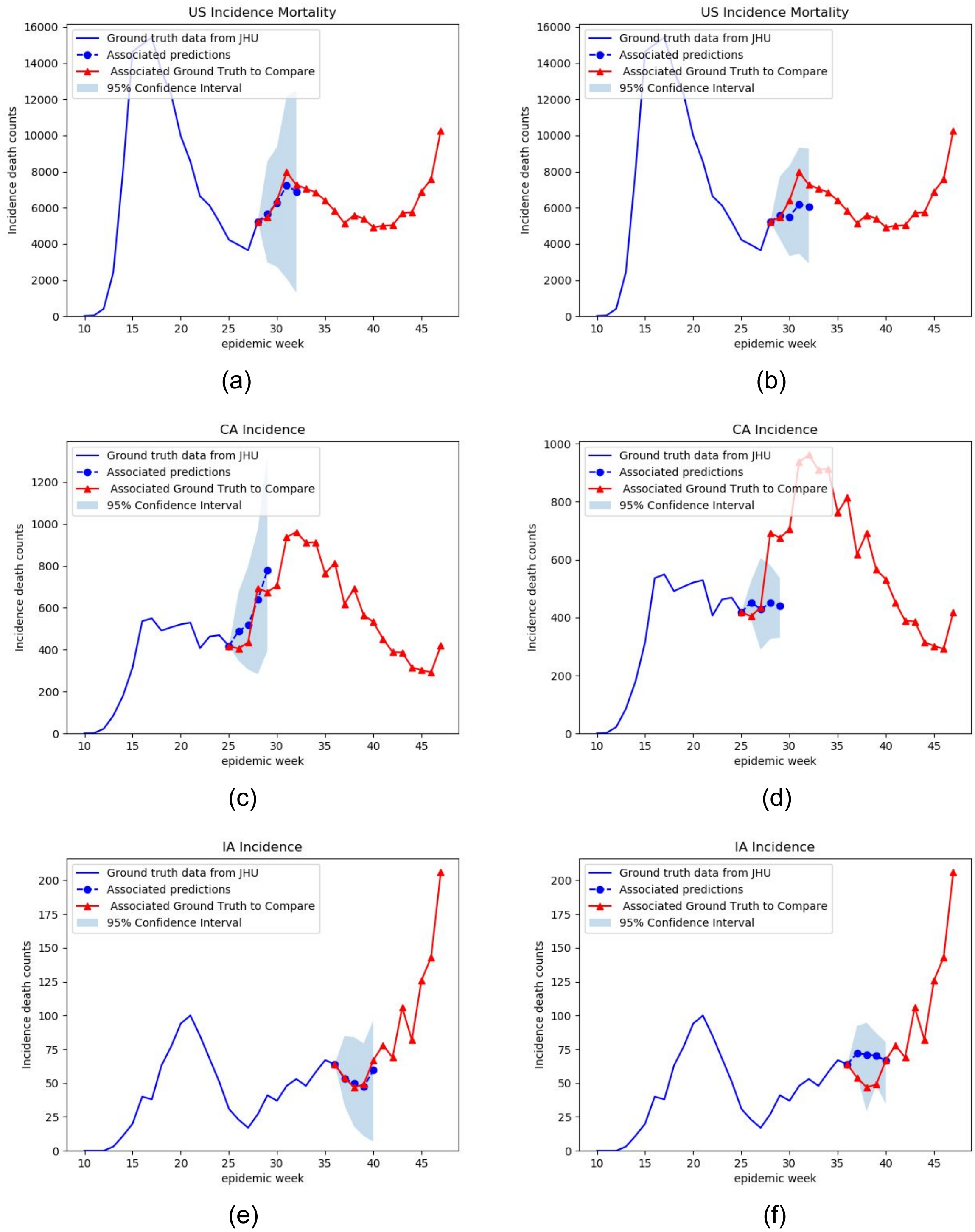


Figure 8: Left: without survey signals; right: adding survey signals. We found a few cases where survey signals hindered the trend prediction made by the model without the survey signals. From top to bottom: US National, California (CA), and Iowa (IA).

Conclusion: We have demonstrated the capability of the survey signals to help in anticipating important trend changes in the epidemic curve. However, there are also some negative cases that may appear when using it. It is reasonable to believe there might be some learnable patterns that can warn us from these negative cases, for which we may require more history of data to capture.

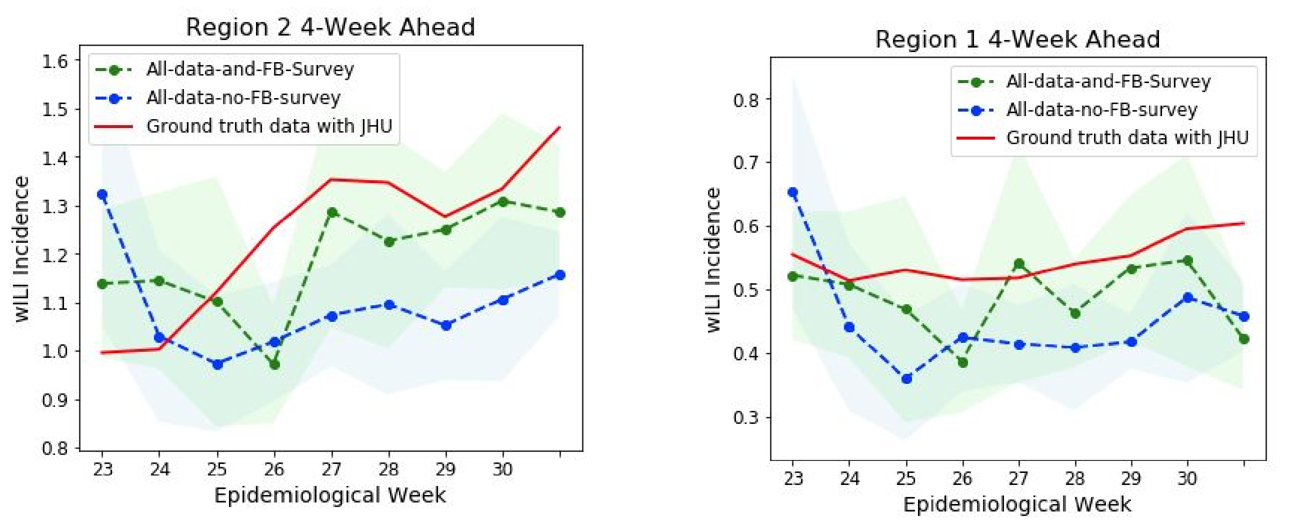
Q4: Do survey signals capture important differences between geographical regions?

For COVID, this ability was showcased with the map presented for Question 2. The positive contribution of survey signals is dispersed across several states in the US. The variety of these states goes from some with lower-than-average epidemic activity as Kentucky to states with high epidemic activity as Florida.

In ILI forecasting, we found that models employing the survey data for forecasting ILI dynamics yield better forecasts compared to models trained without the survey data which tend to underestimate the true ILI progression dynamics. Results shown for HSS Region 1* (Fig. 9b) and HSS Region 2** (Fig. 9a) which are regions with different levels of epidemic activity (thereby different levels of ILI contamination). This showcases the usefulness of the survey data across regions with differences in demographic and epidemic activity.

*HSS Region 1 contains Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont.
**HSS Region 2 contains New Jersey, New York, Puerto Rico, and the Virgin Islands.

Conclusion: Yes, they do capture and help us in forecasting in regions with important differences such as epidemic activity.



(a) Region 2: Medium epidemic activity.

(b) Region 1: Low Epidemic activity

Figure 9: We notice that in the aforementioned two regions showcased, with varying levels of pandemic activity affecting the ILI dynamics, the models with symptom survey data are able to better forecast the ILI dynamics than the models that do not employ the survey data which usually tend to underestimate the future ILI trends.

Competitiveness and distinctive features of our approach with respect to other state-of-the-art models

Excellence in short-term COVID forecasting: We compare against the official ensemble of all contributing models in the COVID-19 Forecast Hub (including ours). The ensemble has been regarded as one of the best performing models by different independent assessments published on the Web. Needless to say, national-level forecasts are crucial for federal decision makers and are the most visible forecasts in national media. Our approach DeepCOVID clearly outperforms this very strong baseline in 1- and 2-week ahead across three months (see Fig. 10).

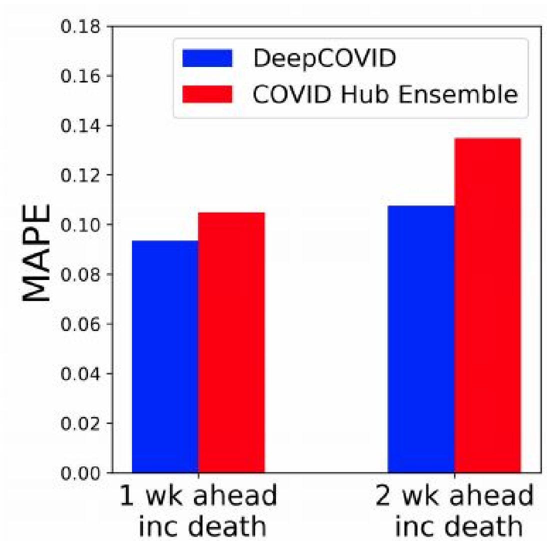


Figure 10: DeepCOVID outperforms the official ensemble in US National short-term (1-2-week ahead) forecasting measured using MAPE.

Epidemiological insights: In contrast to other COVID forecasting models, we designed our approach with an interpretability module also allows us to communicate to domain-experts about which signals are helping the most. For instance, we noticed a high contribution of mobility on our forecasts during May to June, when stay-at-home orders were lifted in most states, businesses reopened, and mobility signals increased in both US National and California. However, this contribution was lower or non-existing during July to August, a period when most mobility signals have already stabilized.

Successful adaptation to wILI forecasting in times of COVID: The effect of contamination from COVID was the most pronounced during March, leading to the COVID-ILI curve exhibiting uncharacteristic non-trivial progression dynamics. Our method was able to adapt our historical ILI model to this new scenario while other methods had difficulties (see Fig. 11).

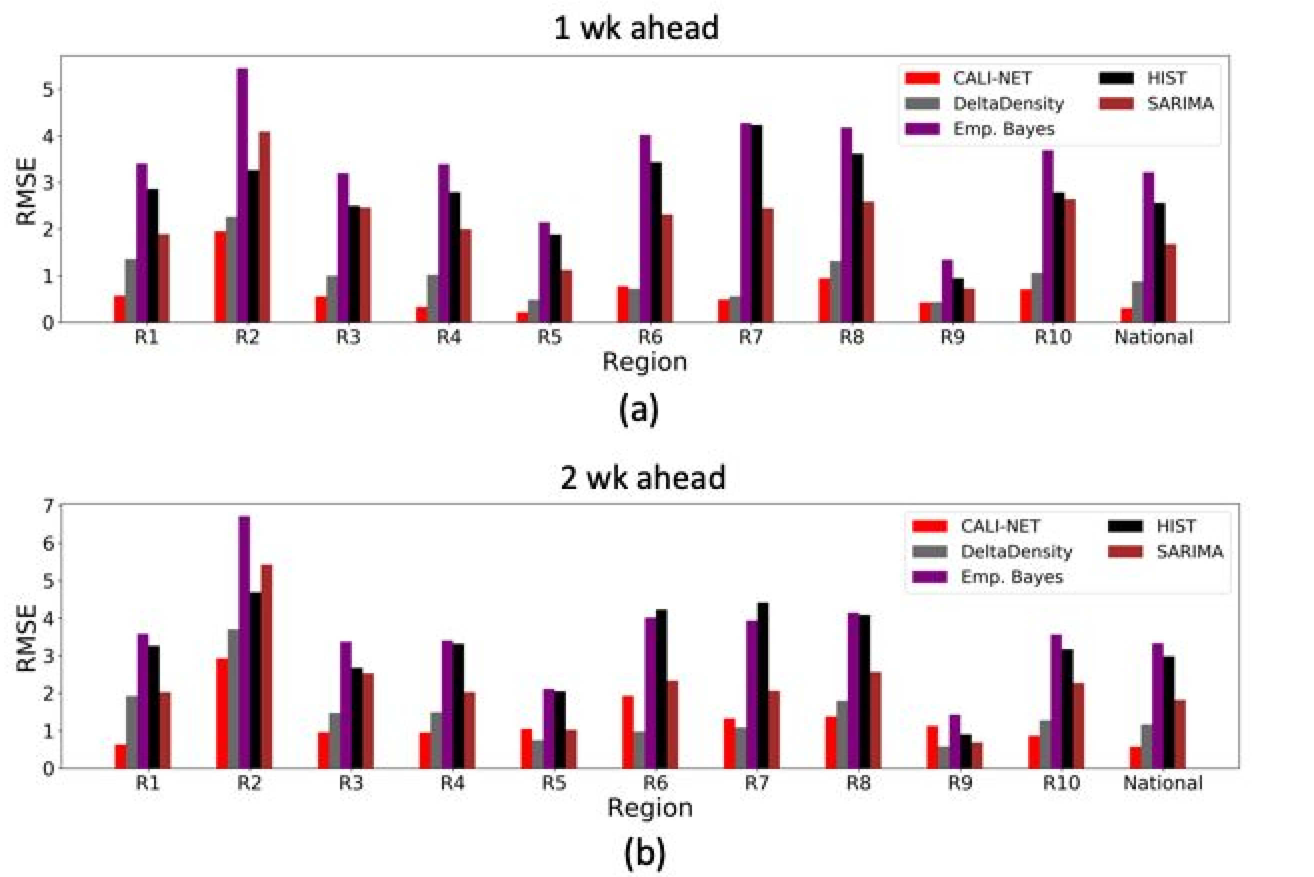


Figure 11: Our method, CALI-Net, outperforms state-of-the-art baselines in adapting to the new scenario where flu and COVID co-exists.

More recently, we also found that in some regions there was an unusual increase in wILI values during the summer which led to a small peak (see. Fig. 12). As we can see in Fig. 13 (showcasing results for HHS region 4), our model was able to adapt to this better than the historical ILI model EpiDeep (published at the premier conference SIGKDD 2019).

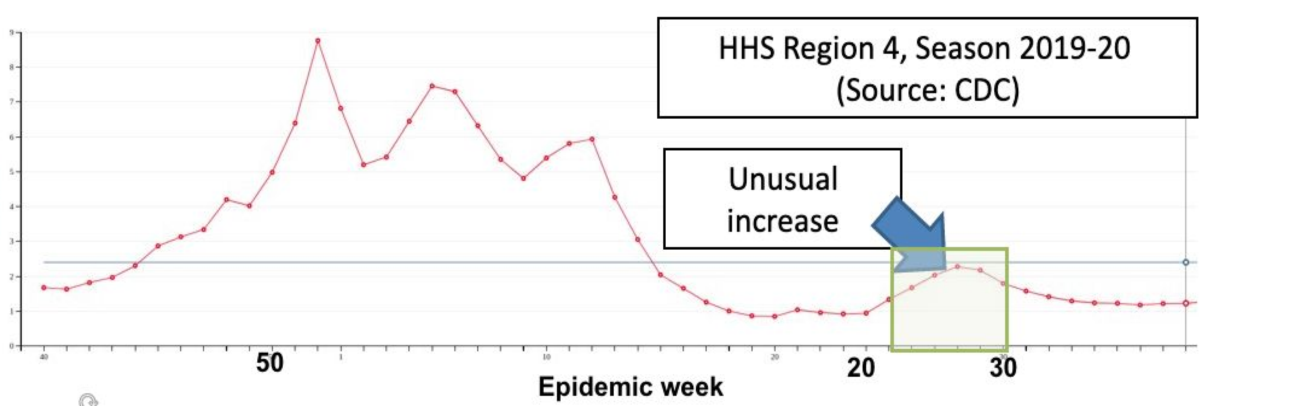
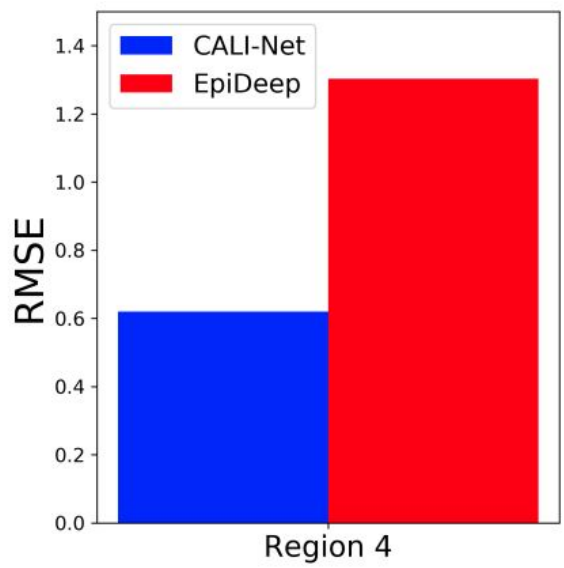


Figure 12. Region 4 wILI Season 2019-20 depicting unusual increase around weeks 24 - 30.



CALI-Net performs better than EpiDeep (a historical ILI model)

Figure 13: Symptom Survey data helps CALI-Net yield better forecasts (e.g., HHS Region 4), measured using RMSE.

These results suggest that we are effectively adapting our historical ILI model to the novel scenario where Covid and influenza coexist. Our approach transfers knowledge to the historical model from the Covid signal when appropriate (positive transfer), and prevents 'negative' transfer when it is better to rely on the historical data. Performance results in the aforementioned figures were obtained during epidemiological weeks 23 to 30.

We also found that the symptom survey data was helpful in achieving good forecasting performance (Fig. 14).

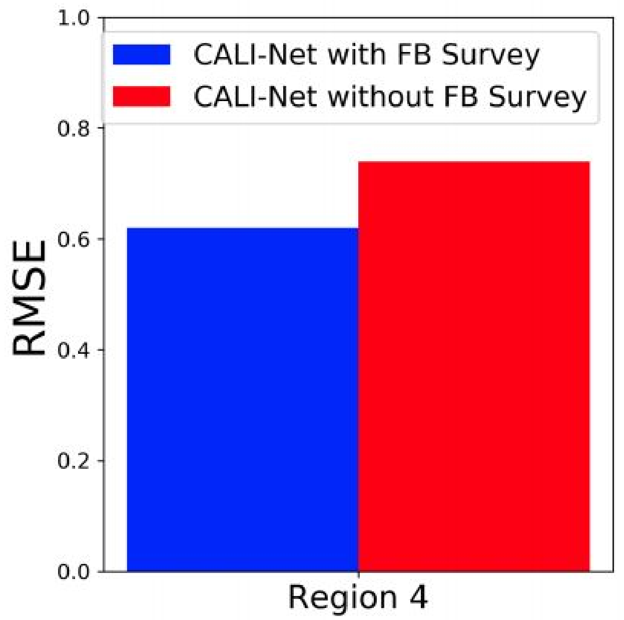


Figure 14: Symptom Survey data helps CALI-Net (our model) yield better forecasts (e.g., HHS Region 4), measured using RMSE.

DISCUSSION

Discussion and implications of findings: We have presented a forecasting framework that aims to improve the situational awareness of two infectious diseases with symptomatic similarities during a pandemic. We emphasized our findings on the usefulness of the symptom survey data, which helps both of our models yield better forecasts.

Symptom Survey Data Usage Notable Highlights:

- In general, survey signals should be used in conjunction with others; however, we found a few interesting cases when they alone offer a different and more accurate forecasting perspective.
- In general, survey signals are orthogonal to other available signals that we included in our models. We found them useful to improve our performance in the majority of geographical regions.
- We showed that survey signals help guide our forecasts to effectively anticipate future trends, which is the general case; however, there are some cases where it may lead to hinder some good trend predictions.
- Survey signals capture and help us in forecasting in regions with important differences such as epidemic activity. In particular, we found that in ILI forecasting, not using symptom survey data may lead to underestimating the epidemic curve.

LIMITATIONS AND FURTHER WORK

We presented an analysis describing different facets of the contribution of survey signals in forecasting COVID-19 and wILI. It is reasonable to expect that some conclusions of this analysis may only hold for the models described, as every model may be designed to emphasize a different aspect of modeling, and/or the data signals included may affect the output. Still, it is important to emphasize that we have shown that similar contributions have been displayed in both models (which are different architectures). Therefore, we have good evidence to believe that other purely data-driven models as ours may reach similar findings.

On the other hand, there are still open questions in forecasting with purely data-driven models. We found that forecasting in lower-level granularities, where some signals are more bursty, may lead to problems in predictions. This is also a good example of where mechanistic models can help. Models trained with symptom survey data, in general did not yield significant improvements with respect to better uncertainty quantification when compared to models that do not employ symptom survey data hence exploring avenues to better incorporate symptom survey data to aid uncertainty estimation can also be considered interesting future work.