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

Alexander Rodríguez[1][§], Nikhil Muralidhar[2][§], Bijaya Adhikari[3], Anika Tabassum[2,1],
Naren Ramakrishnan[2], B. Aditya Prakash[1]

[1]College of Computing, Georgia Institute of Technology, Atlanta, USA
[2]Department of Computer Science, Virginia Tech, Arlington, USA
[3]Department of Computer Science, University of Iowa, Iowa City, USA
{arodriguezc, badityap}@gatech.edu, {nik90, anikat1, naren}@cs.vt.edu, bijaya-adhikari@uiowa.edu

Abstract

Forecasting influenza in a timely manner aids health organizations and policymakers in adequate preparation and decision making. However, effective influenza forecasting still remains a challenge despite increasing research interest. It is even more challenging amidst the COVID pandemic, when the influenza-like illness (ILI) counts are affected by various factors such as symptomatic similarities with COVID-19 and shift in healthcare seeking patterns of the general population. Under the current pandemic, historical influenza models carry valuable expertise about the disease dynamics but face difficulties adapting. Therefore, we propose CALI-NET, a neural transfer learning architecture which allows us to 'steer' a historical disease forecasting model to new scenarios where flu and COVID co-exist. Our framework enables this adaptation by automatically learning when it should emphasize learning from COVID-related signals and when it should learn from the historical model. Thus, we exploit representations learned from historical ILI data as well as the limited COVID-related signals. Our experiments demonstrate that our approach is successful in adapting a historical forecasting model to the current pandemic. In addition, we show that success in our primary goal, adaptation, does not sacrifice overall performance as compared with state-of-the-art influenza forecasting approaches.

1 Introduction

Influenza is a seasonal virus which affects 9–45 million people annually in the United States alone resulting in between 12,000–61,000 deaths. Forecasting flu outbreak progression each year is an important and non-trivial task due to many confounding social, biological, and demographic factors. Accurate forecasts of the onset, peak and incidence can all aid significantly toward informing personalized policy roll out to minimize the effects of the flu season. To this end, the Centers for Disease Control and Prevention (CDC) has been organizing the FluSight challenge for the past several years, where the goal is to predict weighted influenza-like-illness counts (wILI) throughout the flu season in the United States (Biggerstaff et al. 2016). wILI measures the percentage of healthcare

seekers who show influenza like symptoms. Estimating various measures related to the progression of a flu season (such as future incidence) gives policymakers valuable lead time to plan interventions and optimize supply chain decisions.

Moreover, the world has also been experiencing the devastating impacts of the COVID-19 pandemic which has sharply illustrated our enormous vulnerability to emerging infectious diseases. Hence in addition to being affected by various biological and demographic factors, wILI counts may now get further 'contaminated' in the current (and possibly future) influenza seasons, in part due to symptomatic similarities with COVID. Such miscounting manifests as significant changes in wILI seasonal progression as observed in Fig. 1a. Here the wILI curve for the current season 2019-2020 (contaminated by COVID, bold black) clearly shows a very different pattern compared to the previous seasons (in grey). To capture the deviance of the current wILI season and to forecast it in presence of COVID, we require a novel approach. Accurate forecasts of these unexpected trends in the current season are very helpful for resource allocation and healthcare worker deployment. Additionally, predicting wILI 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. Finally, it is widely believed that COVID may be in circulation for a long period of time. Additionally, wILI itself models a mix of flu strains (CDC 2020). Hence, more generally, such a method can be used to disambiguate trends between historical strains and new emerging strains during a flu season.

There has been a recent spate of work on flu forecasting using statistical approaches usually trained on historical wILI data (Adhikari et al. 2019). However, this new forecasting problem of adapting to a new emerging pandemic scenario is complex and cannot be addressed by traditional historical wILI methods alone. See Fig. 1b; current methods based on historical wILI cannot predict the uptrend, while our method (in red) can. The atypical nature of our 'COVID-ILI' season may be caused by multiple co-occurring phenomena, e.g., the actual COVID-19 infections, the corresponding shutdowns and societal lock-downs and also changes in the healthcare-seeking behaviors of the public. This leads to the peak in
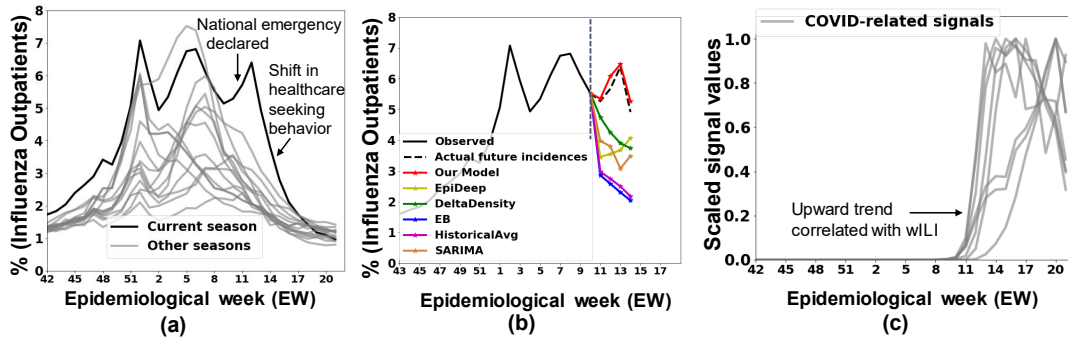
Figure 1: (a) A novel forecasting scenario due to an emerging pandemic. Note the difference between the current (2019-20) and past seasons. (b) Established flu forecasting methods are not able to adapt to uptrend caused by COVID. (c) Exogenous COVID related signals correlate better with wILI trend changes (due to contamination), which we exploit for more accurate forecasting.

COVID-ILI cases to be "out of step" with traditional wILI seasonal trends. As this feature is exclusive to this season, capturing the new trend is a major challenge.

Note that using only the historical wILI seasons is not sufficient to overcome it. Hence for this novel problem, we propose to leverage external COVID-related signals such as confirmed cases, hospitalizations, and emergency room visits as well. This leads us to the second challenge, viz. how to effectively model the COVID-ILI curve with new COVID-related signals, while also leveraging past prior knowledge present in the previous wILI seasons? However, note that these external 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 has 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.

To address these challenges, 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. We employ transfer learning and knowledge distillation approaches to ensure effective transfer of knowledge of the historical wILI trends. We incorporate multiple COVID-related data signals all of which help capture the complex data contamination process showcased by COVID-ILI. As shown in Fig. 1c, these exogenous signals correlate better with the anomalous trends caused by COVID. 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 (See Sec. 5). Our contributions are as follows:

• We develop CALI-NET, a novel heterogeneous transfer learning framework to adapt a flu forecasting historical model into the new scenario of COVID-ILI forecasting.

• We embed CALI-NET with a recurrent neural network including domain-informed spatial constraints to capture the spatiotemporal dynamics across different wILI regions.

• We also employ a Knowledge Distillation scheme to explicitly transfer historical wILI knowledge to our target model

in CALI-NET, thereby alleviating the effect of paucity of COVID-ILI data.

• Finally, we show how CALI-NET succeeds in adaptation, and also perform a rigorous performance comparison of CALI-NET with several state-of-the-art wILI forecasting baselines. In addition, we perform several quantitative and qualitative experiments to understand the effects of various components of CALI-NET.

Overall, more broadly, our work is geared towards adapting a historical model to an emerging disease scenario, and we specifically demonstrate the effectiveness of our approach in the context of wILI forecasting in the COVID-19 emerging disease scenario. Appendix, code, and other resources can be found online[1].

2   Related Work

To summarize, we are the first to address the problem of adapting to shifting trends using transfer learning and knowledge distillation in an epidemic forecasting setting, leveraging exogenous signals as well as historical models. Our research draws from multiple lines of work.

Epidemic Forecasting: Several approaches for epidemic forecasting have been proposed including statistical (Tizzoni et al. 2012; Adhikari et al. 2019; Osthus et al. 2019; Brooks et al. 2018), mechanistic (Shaman and Karspeck 2012; Zhang et al. 2017), and ensemble (Reich et al. 2019b) approaches. Several approaches rely on external signals such as environmental conditions and weather reports (Shaman, Goldstein, and Lipsitch 2010; Tamerius et al. 2013; Volkova et al. 2017), social media (Chen et al. 2016; Lee, Agrawal, and Choudhary 2013), search engine data (Ginsberg et al. 2009; Yuan et al. 2013), and a combination of multiple sources (Chakraborty et al. 2014). Recently, there has been increasing interest in deep learning for epidemic forecasting (Adhikari et al. 2019; Wang, Chen, and Marathe 2019; Rodríguez et al. 2020). These methods typically exploit intra and inter seasonal trends. Other approaches like (Venna et al. 2018) are limited to specific situations, e.g., for military populations. However, to the best of our knowledge, there has been no

---

[1]Resources website: https://github.com/AdityaLab/CALI-Net

work on developing deep architectures for adapting to trend shifts using exogenous data.

Time Series Analysis: There are several data driven, statistical and model-based approaches that have been developed for time series forecasting such as auto-regression, Kalman-filters and groups/panels (Box et al. 2015; Sapankevych and Sankar 2009). Recently, deep recurrent architectures (Hochreiter and Schmidhuber 1997) have shown great promise in learning good representations of temporal evolution (Fu, Zhang, and Li 2016; Muralidhar, Muthiah, and Ramakrishnan 2019; Connor, Martin, and Atlas 1994).

Transfer Learning within heterogeneous domains: This challenging setting of transfer learning with heterogeneous domains (different feature spaces) aims to leverage knowledge extracted from a *source* domain to a different but related *target* domain. (Moon and Carbonell 2017) proposed to learn feature mappings in a common-subspace, and then apply shared neural layers where the transfer would occur. (Li et al. 2019) proposed transfer learning via deep matrix completion. (Yan et al. 2018) formulated this problem as an optimal transport problem using the entropic Gromov-Wasserstein discrepancy. We adapt the classification method in (Moon and Carbonell 2017) to our regression setting, for effectively transferring knowledge from the source to the target model in our COVID-ILI forecasting task. Knowledge Distillation (KD) is also a popular transfer learning method, to develop shallow neural networks capable of yielding performance similar to deeper models by learning to "mimic" their behavior (Ba and Caruana 2014; Hinton, Vinyals, and Dean 2015). (Saputra et al. 2019) inspects KD for deep pose regression. The authors propose two regression specific losses, namely the hint loss and the imitation loss which we adapt in this work for COVID-ILI forecasting. Unlike our paper, most KD work has been applied to classification and efforts for adapting KD for regression have been sparse (Saputra et al. 2019; Takamoto, Morishita, and Imaoka 2020).

3 Background

COVID-ILI forecasting task: Here we consider a short term forecasting task of predicting the next k wILI incidence given the data till week $t - 1$ for each US HHS region and the national region. This corresponds to predicting the wILI values for week $\{t, t+1, \ldots, t+k\}$ at week t (matching the exact real-time setting of the CDC tasks) for each region.

We are given a set of historical annual wILI time-series, $Y_i = \{y_i^1, y_i^2, \ldots, y_i^{t-1}\}$ for each region i. The wILI values have been contaminated by COVID-19 for all weeks $t \geq w$. We also have various COVID-related exogenous data signals $X_i = \{\mathbf{x}_i^w, \mathbf{x}_i^{w+1}, \ldots, \mathbf{x}_i^{t-1}\}$, where each feature vector $\mathbf{x}_i^j$ is constructed using various signals such as COVID line list data, test availability, crowd-sourced symptomatic data, and social media. Our task is to forecast the next k wILI incidence for all regions $i \in I$. Specifically, our novel problem is:

Problem 1 *COVID-ILI Forecasting Problem*
Given: *a set of historical annual wILI time-series* $Y_i = \{y_i^1, y_i^2, \ldots, y_i^{t-1}\}$ *for regions* $i \in I$ *and the set of COVID-related exogenous signals* $X_i = \{\mathbf{x}_i^w, \mathbf{x}_i^{w+1}, \ldots, \mathbf{x}_i^{t-1}\}$ *for*

the current season and regions $i \in I$.
Predict: *next k wILI incidences* $\forall_{j=t}^{t+k} y_i^j$ *for each region* $i \in I$.

Epideep for wILI forecasting. EPIDEEP (Adhikari et al. 2019) is a deep neural architecture designed specifically for wILI forecasting. The core idea behind EPIDEEP is to leverage the seasonal similarity between the current season and historical seasons to forecast various metrics of interest such as next incidence values, onset of current season, the seasonal peak value and peak time. To infer the seasonal similarity between current and the historical season, it employs a deep clustering module which learns latent low dimensional embeddings of the seasons.

4 Our Approach

In this section, we describe our method CALI-NET, which models COVID-ILI by incorporating historical wILI knowledge as well as the limited new COVID-related exogenous signals. Next we give an overview of how our approach uses heterogeneous transfer learning (HTL), overcomes data paucity issues, and controls the transfer of only useful knowledge and avoids negative transfer.

4.1 Exploiting Learned Representations from Historical wILI via HTL

We leverage recent advances in HTL to incorporate the rich historical wILI data. To that end, we use the EPIDEEP model as our base model. EPIDEEP was designed to learn representations from historical wILI that embed seasonal and temporal patterns. Here, we adapt the CTHL framework (Moon and Carbonell 2017) to transfer knowledge from EPIDEEP.

In our HTL setting, a modified version of the EPIDEEP model is the source model and we design a feature module (discussed in Sec. 4.2) to be the target model. As depicted in Fig. 2, the embeddings of the source and target are each transformed by modules $\mathbf{s}$ and $\mathbf{t}$ respectively, such that the latent embeddings of the source and target model are placed into a common feature space. In this way, we are projecting knowledge extracted from both heterogeneous feature spaces into a shared latent space. Formally, the transformations may be expressed as $\mathbf{s} : \mathbb{R}^{M_S} \to \mathbb{R}^{M_J}$ and $\mathbf{t} : \mathbb{R}^{M_T} \to \mathbb{R}^{M_J}$, where M_S and M_T are the dimensions of the source and target embeddings, respectively, and M_J is the dimension of the joint latent feature space. After projecting representations from the source and the target models into a joint latent feature space, a sequence of shared transformations $\mathbf{f}_1 : \mathbb{R}^{M_J} \to \mathbb{R}^{M_A}$ and $\mathbf{f}_2 : \mathbb{R}^{M_J} \to \mathbb{R}$ is applied on them, thereby transporting them both into the same latent space. On top of this architecture, we employ a denoising autoencoder to reconstruct the input data as we find it improves our latent representations. These modules are depicted as $\mathbf{s}'$ and $\mathbf{t}'$.

4.2 COVID-Augmented Exogenous Model (CAEM)

Our target model from Sec. 4.1 could be a simple feedforward network. Instead, to alleviate the data paucity that exists for the COVID-related exogenous data, we develop the COVID-Augmented Exogenous Model (CAEM) which *jointly* models
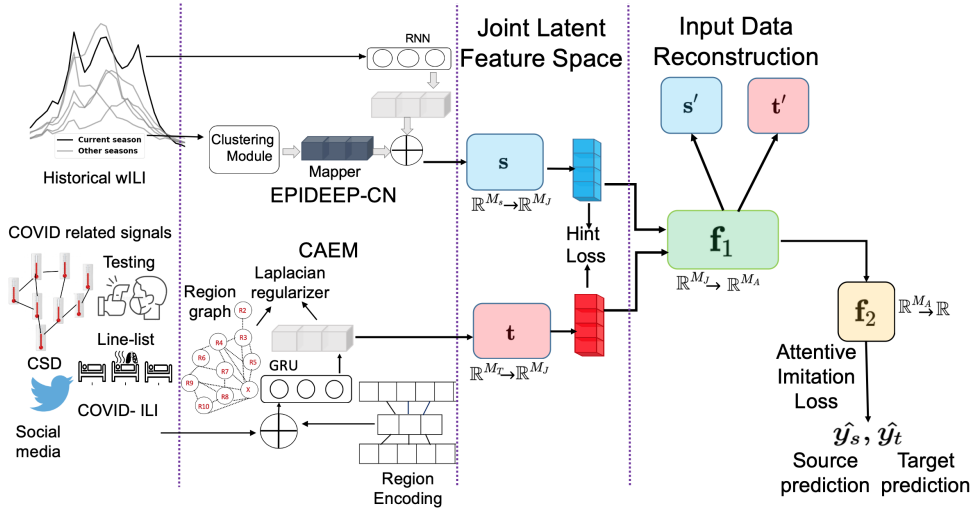
Figure 2: Our proposed model CALI-NET. Our heterogeneous transfer learning architecture is designed to transfer knowledge from EPIDEEP-CN about historical wILI trends to the CAEM module (using exogenous signals) for COVID-ILI forecasting while addressing the challenges of negative transfer, spatial consistency and data paucity.

all regions exploiting regional interplay characteristics. Such an approach allows us to extract the most out of our limited training data enabling us to employ more sophisticated sequential architectures. To enable model awareness of multiple regional patterns, we explicitly encode each *region embedding* $\mathbf{r} \in \mathbb{R}^{1 \times h_r}$, and pass to CAEM along with the exogenous input data of the region for a particular week. The region embeddings are produced by an autoencoder whose task is to reconstruct one-hot encodings of each region.

The data we consider exhibit sequential dependencies. In order to model these dependencies, we employ the popular GRU recurrent neural architecture (Cho et al. 2014). The GRU is trained to encode temporal dependencies using data from week $t - W$ to week $t - 1$ and predict values for week $t + k$. At each step of recurrence, the GRU receives as input, exogenous data signals $\mathbf{x}_i^{t-\lambda} \in \mathbb{R}^{1 \times l}$ (for week $t - \lambda$ with $\lambda \in \{W, W - 1, \dots, 1\}$ and region i) and the region embedding r_i, both concatenated to form the full GRU input. For simplicity, henceforth, we consider $\mathbf{x}_i^{t-\lambda} \in \mathbb{R}^{1 \times l + h_r}$ to represent this concatenated input to the GRU (l is the number of different data signals we employ and h_r is the dimension of the latent region embedding obtained from the CAEM Region Embedding autoencoder).

Laplacian Regularization: Infectious diseases like COVID and flu naturally also show strong spatial correlations and to capture this aspect of the wILI season evolution across different regions effectively, we incorporate spatial constraints using Laplacian Regularization (Belkin, Matveeva, and Niyogi 2004) and predict COVID-ILI values for all regions jointly. Let us consider the region graph $G(V, E)$ where V (vertices) indicates the number of regions (11 in our case including the national region) and E indicates edges between the vertices. Two regions are considered to have an edge between them if they are bordering each other. We construct G based on region demarcations provided by the HHS/CDC and connect

the national region to all other regions.

The optimization objective for CAEM is as follows:

$$\min_{\Theta_{RE}\Theta_F} ||F(\mathbf{X}^{t-W:t-1}; \Theta_F) - Y^t||_2^2 + \\ RE(\mathbf{E}; \Theta_{RE}) + Tr(\mathbf{h}^T \mathbf{L} \mathbf{h}) \tag{1}$$

In Eq. 1, Θ_{RE}, Θ_F represent the model parameters for *region embedding* (RE) function and the recurrent *forecasting* (F) function respectively. The input to F, $\mathbf{X}^{t-W:t-1} \in R^{|V| \times l + h_r}$, is the historical COVID-related exogenous data for the past W weeks for all 11 regions (along with the regional embedding for each). The output of F, $Y^t \in R^{|V| \times 1}$ for week t includes forecasts for all 11 regions. The region embedding is generated using an autoencoder which accepts the one-hot encoding for all regions $\mathbf{E} \in \mathbb{B}^{|V| \times |V|}$. Finally, $\mathbf{h} \in \mathbb{R}^{1 \times h_r}$ is the hidden representation of the input sequence generated by the forecasting model F at the end of the recurrence which is used to enforce regional representation similarity governed by the normalized Laplacian ($\mathbf{L}$) of graph G. Laplacian regularization has been shown to systematically enforce regional similarity, effectively capturing spatial correlations (Subbian and Banerjee 2013). Both the RE and F modules of CAEM are jointly trained coupled with Laplacian regularization. It must be noted that when integrated into CALI-NET, the function F includes the GRU parameters and the parameters for transformations $\mathbf{t}, \mathbf{f}_1, \mathbf{f}_2$ employed to yield the final k-week ahead predictions.

4.3 Attentive Knowledge Distillation Loss

A mechanism for the target model to exercise control over knowledge transfer and prevent negative transfer is necessary in our setting to avoid the transfer of possibly erroneous predictions made by EPIDEEP for the atypical portions of the current influenza season. To enable this, we employ attentive knowledge distillation (KD) techniques. Recently, (Saputra et al. 2019) has employed KD in deep pose estimation. We

noticed that our modification to this method is capable of not only transferring knowledge from EPIDEEP (our source model) to CAEM (our target model) but also showcases how to selectively transfer knowledge based on the quality of source model predictions.

Adapting EPIDEEP: To achieve effective transfer of knowledge from EPIDEEP to CAEM, we modify the existing EPIDEEP architecture. EPIDEEP, by design, requires a different model to be trained for each week in the season which is not ideal for effective knowledge transfer. Hence, to prevent this, we modify EPIDEEP into EPIDEEP-CN (EpiDeep-CALI-NET) which incrementally re-trains the same model for each week in the season, thereby allowing the KD losses to be applied to the same set of EPIDEEP-CN parameters, enabling more efficient knowledge transfer to CAEM. Specifics about EPIDEEP-CN are in our appendix.

The training datasets for EPIDEEP-CN and CAEM are not the same, however they share an overlapping subset of data, from January 2020 onward when the COVID pandemic started. Therefore, we enforce the *KD loss* only on this subset of the training data. Our KD loss is composed of two terms: imitation loss $\mathcal{L}_{\text{Im}}$ and hint loss $\mathcal{L}_{\text{Hint}}$, described mathematically as follows,

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

where $\Phi_i = \left(1 - \frac{\|\hat{y}_s - y\|_i^2}{\eta}\right)$, and Ψ_s and Ψ_t are the output embeddings of **s** and **t**, respectively; $i \in \{1, \ldots, n\}$ is the index for each training observation, and n the batch size; $\eta = \max(e_s) - \min(e_s)$ is a normalizing factor (i.e range of squared error losses of source model) and $e_s = \left\{\|y - \hat{y}_s\|_j^2 : j = 1, \ldots, N\right\}$, the actual set of squared errors between the source predictions and ground truth. N is the total number of observations in the overlap training data; Φ_i is the attention weight assigned to the ith training observation. The attention is a function of how well the source model is able to predict $(\hat{y}_s)$ a particular ground truth target y. The attention weights are applied over the imitation loss between the source predictions $(\hat{y}_s)$ and the ground truth $(\hat{y}_t)$ as well as over the hint loss between the latent output embeddings Ψ_s and Ψ_t to ensure transfer of knowledge at multiple levels in the architecture from the source EPIDEEP-CN to the target CAEM. The goal of KD is to enforce a unidirectional transfer from the source (EPIDEEP-CN) to the target (CAEM) model. Hence KD losses do not affect the representations learned by EPIDEEP-CN and module **s**.

5 Experiments

Setup. All experiments are conducted using a 4 Xeon E7-4850 CPU with 512GB of 1066 Mhz main memory. Our method implemented in PyTorch (implementation details in the appendix) is very fast, training for one predictive task in about 3 mins. Here, we present our results for next incidence prediction (i.e. $k = 1$). We present results for next-two incidence predictions in the appendix, which are similar. Note that we define T_1 as the period of non-seasonal rise of wILI

due to contamination by COVID-19 related issues (EWs 9-11), T_2 as the time period when COVID-ILI trend is declining more in tune with the wILI pattern (EWs 12-15), and T as the entire course (EWs 9-15).

Data. We use the historical weighted Influenza-like Illness (wILI) data released by the CDC which collects it through the Outpatient Influenza-like Illness Surveillance Network (ILINet). ILINet consists of more than 3,500 outpatient healthcare providers all over the US. We refer to wILI from June 2004 until Dec 2019 as **historical wILI**, and wILI from January 2020 as **COVID-wILI**. Next, Table 1 details the various signal types we employ for COVID-related exogenous data. A more detailed description of each data signal can be found in the appendix. All datasets are publicly available and were collected in May 2020.

Goals. In our experiments we aim to demonstrate that our method CALI-NET can systematically steer a historical model to the new COVID-ILI scenario by enabling it to learn from Covid-related signals, when appropriate. We are interested in determining whether our model can transfer useful information from the historical model (i.e. EPIDEEP) when required and if it can prevent transfer of detrimental information. Specifically, our questions are:

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 forecasting?

Q1 and Q2, which are about transfer learning, are aligned with the main goal of this paper. In Q3, we are interested in determining whether CALI-NET sacrifices any overall forecasting performance, as compared to the state-of-the art (SOTA) baselines, by being too focused on balancing the transfer of knowledge? In Q4 and Q5, we analyze the importance of different components and data signals to performance.

Training and Optimization. For training CALI-NET, we do the following. We found that, for practical purposes, it is convenient to pre-train EPIDEEP-CN, and then remove its last feedforward layers (decoder). Hence the concatenated output of RNN encoder and embedding mapper are input to module **s**. During joint training, we do not modify EPIDEEP-CN's pre-trained parameters. As recommended in (Moon and Carbonell 2017), we train the HTL architecture in an alternating fashion.

Baselines. We use traditional historical wILI forecasting methods used in literature (Reich et al. 2019a): EPIDEEP, extended DELTA DENSITY from Delphi Group (Brooks et al. 2018) (which is SOTA as the top performing method in recent CDC influenza forecasting challenges (Reich et al. 2019a)), SARIMA from ReichLab (Ray et al. 2017) (seasonal autoregressive method, top performing in recent CDC challenges), and EMPIRICAL BAYES (Brooks et al. 2015) (which lever-

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 Project[2], CDC[3] JHU CSSE[4]
(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 Project[2], CDC[3]
(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)

Table 1: Overview of COVID-Related Exogenous Data.

ages transformation of historical seasons to forecast the current season), and HIST (common persistence baseline which forecasts based on weekly average of the historical seasons).

5.1 Q1: Leveraging Positive Transfer for Covid-contaminated wILI

The effect of contamination is most pronounced in T_1, leading to the COVID-ILI curve exhibiting uncharacteristic non-trivial progression dynamics. Therefore, to effectively steer our historical model, CALI-NET automatically leverages positive transfer of Covid-related signals into EPIDEEP. The effect of this automatic positive transfer is shown in Fig. 3a where we see that CALI-NET significantly outperforms EPIDEEP across *all regions*, thanks to our architecture.
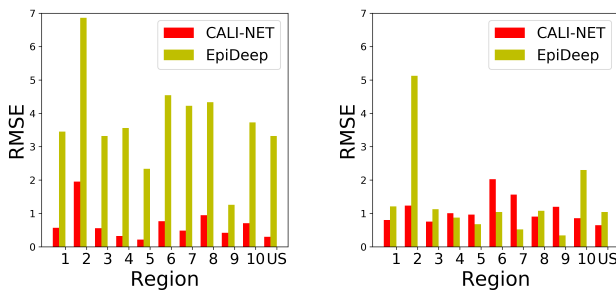
(a) Positive transfer stage   (b) Negative transfer stage

Figure 3: (a) CALI-NET achieves good forecasts of the uncharacteristic trend in period T_1 by steering our influenza forecasting historical model EPIDEEP with knowledge learned from Covid-related signals. (b) Shows forecasting errors from period T_2, when the COVID-ILI trend is declining more in tune with the traditional wILI pattern. We notice that CALI-NET is competitive with EPIDEEP, also outperforming it in 6 out of 11 regions while remaining competitive in the rest of the regions.

5.2 Q2: Does CALI-NET Prevent Negative Transfer Automatically?

Having showcased the adaptation of CALI-NET in T_1, we now show in Fig. 3b that our method is effective at prevent-

ing negative transfer when wILI is no longer aligned with the exogenous COVID signals (i.e., period $T2$). In the first place, in some regions the wILI trajectory was never significantly affected by COVID as confirmed COVID cases started to increase significantly only once the influenza season ended. Second, COVID-affected wILI trajectories of regions displayed a subsequent downtrend after a few weeks. This may be due to the change in care-seeking behavior of outpatients (Kou et al. 2020). In this stage, preventing negative transfer from COVID-related signals is needed, such that our model displays more characteristics of traditional influenza models. From Fig. 3b, we see that CALI-NET is better than EPIDEEP in a majority of the regions indicating that it is able to effectively stop knowledge from misaligned COVID signals from adversely affecting forecasting accuracy thereby effectively preventing negative transfer.
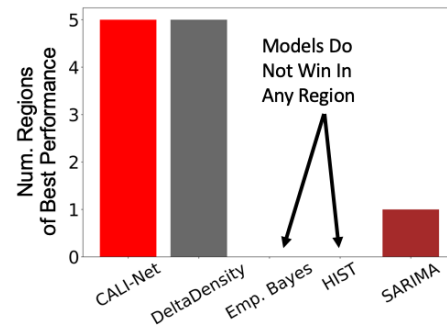

Figure 4: The number of regions in which each model yields best performance (models performing within 1% of best model per region are considered equivalent best performers). CALI-NET outperforms other models in 5 out of 11 regions, on par with DeltaDensity which also yields best performance in 5 other regions; SARIMA is the best only in Region 9.

5.3 Q3: Does CALI-NET Sacrifice Overall Performance?

Sec 5.1 and 5.2 show that CALI-NET successfully achieves the main goal of the paper i.e. steering a historical model in a novel scenario. We now study if we sacrifice any performance in this process. To this end, we compare CALI-NET with the traditional SOTA wILI forecasting approaches for the

---

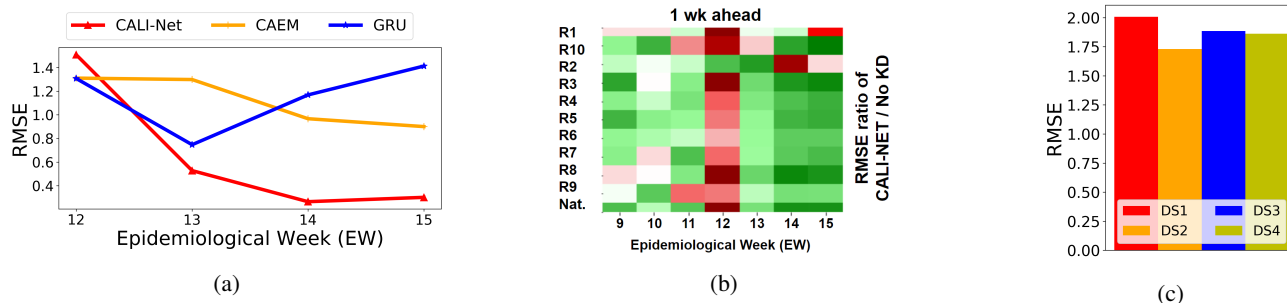[2]covidtracking.com [3]cdc.gov/coronavirus/2019-ncov/ [4]coronavirus.jhu.edu

Figure 5: (a) RMSE in period T_2 in Nat. (b) 1-week ahead RMSE ratio $\frac{\text{CALI-NET}}{\text{CALI-NET-No KD}}$. (c) Data bucket ablation.

entire course T. Specifically, we quantify the number of regions (among all 11), where each method outperforms all others. Fig. 4 showcases our findings. Overall, CALI-NET is able to match the performance of the SOTA historical wILI forecasting models in forecasts for the entire course T and is the top performer in 5 of 11 regions and is one of top 2 best models in 10 out of 11 regions. Note that the traditional wILI baselines do not capture the non-seasonal rise of wILI due to COVID contamination in T_1 (see Fig. 1 and the appendix). Hence, we note that CALI-NET is the best-suited approach for real-time forecasting in a novel scenario as it captures the non-seasonal patterns while maintaining overall performance. Moreover, we also noticed that CALI-NET outperforms all baselines in regions worst affected by COVID (see appendix).

5.4 Q4 and Q5: Module, Data and Parameter Importance and Sensitivity

Justification for CAEM Architecture. We conducted an ablation study testing the three components of CAEM: (a) regional reconstruction, (b) Laplacian regularization, and (c) the recurrent model. We found that removing each of them degrades performance showing their individual effectiveness.
Module Performance Analysis. Fig. 5a shows RMSE evolution over period T_2 for the national region for CALI-NET and sub-models of CALI-NET that do not have transfer learning capability. Both GRU (standard gated recurrent unit model) and the standalone CAEM model use exogenous data as CALI-NET does. CALI-NET is the only model able to adapt quickly to downtrend in period T_2, due to the effect of HTL framework which prevents negative transfer of knowledge from COVID related signals, while other models fail to adapt and, in fact, predict rising or flat wILI forecasts. From Sec. 5.1 - 5.3, we see CALI-NET is the *only* method able to capture both initial uptrend of COVID-ILI and subsequent decline effectively, showing its usefulness for emerging diseases.
Effect of KD. We perform an ablation study to understand the contribution of the proposed KD losses. Often the usefulness of source (EPIDEEP-CN) and target (CAEM) modules vary depending on the usefulness of the historical and exogenous data sources. Our attentive KD distillation losses provide structure and balance to the transferred knowledge.

We compare 1-week ahead forecasting performance of CALI-NET and a variant of CALI-NET with KD losses removed. See Fig. 5b; each box is colored by the ratio of

RMSE of CALI-NET and its variant (capped at -1 and 1 to help visualization). Green cells indicate CALI-NET does better while red cells indicate that CALI-NET w/o KD losses is better. We notice for 1-week ahead forecasting, structuring the knowledge transferred from EPIDEEP proves valuable for most EWs. However, for long-term forecasting, KD losses seem to downgrade guidance of EPIDEEP-CN (results in appendix). This may be because the season reverts to typical behavior in the timeframe predicted in long-term forecasts.

Contribution of Exogenous Signals. In Fig. 5c, we can see the average overall RMSE obtained when a single data bucket was removed during the training of CALI-NET. We noticed that line list based data (DS1) is very helpful in COVID-ILI forecasting while the effectiveness of testing (DS2) and crowdsourced based (DS3) data is slightly more varied across regions, an observation that resolves Q5. This also suggests that data closer to the disease is more reliable. More detailed results (regional breakdown) are in the appendix.

Hyper-Parameter Sensitivity. See appendix.

6 Discussion

We introduced the challenging COVID-ILI forecasting task and proposed our novel approach CALI-NET. We show usefulness of a principled method to transfer relevant knowledge from an existing deep flu forecasting model (based on rich historical data) to one relying on relevant but limited recent COVID-related exogenous signals. Our method is based on carefully designed components to avoid negative transfer (by attentive KD losses), promote spatial consistency (via Laplacian losses in a novel recurrent architecture CAEM), and also handle data paucity (via the global nature of CAEM and other aspects). CALI-NET effectively captures non-trivial atypical trends in COVID-ILI evolution whereas other models and baselines do not. We also demonstrate how each of our components and data signals is important and useful for performance. These results provide guidance for steering forecasting models in an emerging disease scenario. In future, we believe our techniques can be applied to other source models (in addition to EPIDEEP-CN), as well as designing more sophisticated architectures for the target CAEM model. We can also explore adding interpretability to our forecasts for additional insights.

Acknowledgments

References

Adhikari, B.; Xu, X.; Ramakrishnan, N.; and Prakash, B. A. 2019. Epideep: Exploiting embeddings for epidemic forecasting. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 577–586.

Ba, J.; and Caruana, R. 2014. Do Deep Nets Really Need to Be Deep? In *Advances in Neural Information Processing Systems*, 2654–2662.

Belkin, M.; Matveeva, I.; and Niyogi, P. 2004. Regularization and semi-supervised learning on large graphs. In *International Conference on Computational Learning Theory*, 624–638. Springer.

Biggerstaff, M.; Alper, D.; Dredze, M.; Fox, S.; Fung, I. C.-H.; Hickmann, K. S.; Lewis, B.; Rosenfeld, R.; Shaman, J.; Tsou, M.-H.; et al. 2016. Results from the centers for disease control and prevention's predict the 2013–2014 Influenza Season Challenge. *BMC infectious diseases* 16(1): 357.

Boëlle, P.-Y.; Souty, C.; Launay, T.; Guerrisi, C.; Turbelin, C.; Behillil, S.; Enouf, V.; Poletto, C.; Lina, B.; van der Werf, S.; et al. 2020. Excess cases of influenza-like illnesses synchronous with coronavirus disease (COVID-19) epidemic, France, March 2020. *Eurosurveillance* 25(14): 2000326.

Box, G. E.; Jenkins, G. M.; Reinsel, G. C.; and Ljung, G. M. 2015. *Time series analysis: forecasting and control*. John Wiley & Sons.

Brooks, L. C.; Farrow, D. C.; Hyun, S.; Tibshirani, R. J.; and Rosenfeld, R. 2015. Flexible modeling of epidemics with an empirical Bayes framework. *PLoS computational biology* 11(8): e1004382.

Brooks, L. C.; Farrow, D. C.; Hyun, S.; Tibshirani, R. J.; and Rosenfeld, R. 2018. Nonmechanistic forecasts of seasonal influenza with iterative one-week-ahead distributions. *PLoS computational biology* 14(6): e1006134.

Castrofino, A.; Del Castillo, G.; Grosso, F.; Barone, A.; Gramegna, M.; Galli, C.; Tirani, M.; Castaldi, S.; Pariani, E.; and Cereda, D. 2020. Influenza surveillance system and Covid-19. *European Journal of Public Health* 30(Supplement_5): ckaa165–354.

CDC. 2020. Weekly U.S. Influenza Surveillance Report (FluView). URL https://www.cdc.gov/flu/weekly/. Accessed: 2020-09-09.

Chakraborty, P.; Khadivi, P.; Lewis, B.; Mahendiran, A.; Chen, J.; Butler, P.; Nsoesie, E. O.; Mekaru, S. R.; Brownstein, J. S.; Marathe, M. V.; et al. 2014. Forecasting a moving target: Ensemble models for ILI case count predictions. In *Proceedings of the 2014 SIAM international conference on data mining*, 262–270. SIAM.

Chen, L.; Hossain, K. T.; Butler, P.; Ramakrishnan, N.; and Prakash, B. A. 2016. Syndromic surveillance of Flu on Twitter using weakly supervised temporal topic models. *Data mining and knowledge discovery* 30(3): 681–710.

Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; and Bengio, Y. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

Connor, J. T.; Martin, R. D.; and Atlas, L. E. 1994. Recurrent neural networks and robust time series prediction. *IEEE transactions on neural networks* 5(2): 240–254.

Dredze, M.; Cheng, R.; Paul, M. J.; and Broniatowski, D. 2014. HealthTweets. org: a platform for public health surveillance using Twitter. In *Workshops at the Twenty-Eighth AAAI Conference on Artificial Intelligence*.

Fu, R.; Zhang, Z.; and Li, L. 2016. Using LSTM and GRU neural network methods for traffic flow prediction. In *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, 324–328. IEEE.

Ginsberg, J.; Mohebbi, M. H.; Patel, R. S.; Brammer, L.; Smolinski, M. S.; and Brilliant, L. 2009. Detecting influenza epidemics using search engine query data. *Nature* 457(7232): 1012.

Hinton, G.; Vinyals, O.; and Dean, J. 2015. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531* .

Hochreiter, S.; and Schmidhuber, J. 1997. Long short-term memory. *Neural computation* 9(8): 1735–1780.

Kou, S.; Yang, S.; Chang, C.-J.; Ho, T.-H.; and Graver, L. 2020. Unmasking the Actual COVID-19 Case Count. *Clinical Infectious Diseases* .

Lee, K.; Agrawal, A.; and Choudhary, A. 2013. Real-time disease surveillance using twitter data: demonstration on flu and cancer. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1474–1477. ACM.

Li, H.; Pan, S. J.; Wan, R.; and Kot, A. C. 2019. Heterogeneous transfer learning via deep matrix completion with adversarial kernel embedding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 8602–8609.

Miller, A. C.; Singh, I.; Koehler, E.; and Polgreen, P. M. 2018. A smartphone-driven thermometer application for real-time population-and individual-level influenza surveillance. *Clinical Infectious Diseases* 67(3): 388–397.

Moon, S.; and Carbonell, J. G. 2017. Completely Heterogeneous Transfer Learning with Attention: What and What Not to Transfer. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, 2508–2514.

Muralidhar, N.; Muthiah, S.; and Ramakrishnan, N. 2019. DyAt nets: dynamic attention networks for state forecasting in cyber-physical systems. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, 3180–3186. AAAI Press.

Osthus, D.; Gattiker, J.; Priedhorsky, R.; Del Valle, S. Y.; et al. 2019. Dynamic Bayesian influenza forecasting in the United States with hierarchical discrepancy (with discussion). *Bayesian Analysis* 14(1): 261–312.

Ray, E. L.; Sakrejda, K.; Lauer, S. A.; Johansson, M. A.; and Reich, N. G. 2017. Infectious disease prediction with kernel conditional density estimation. *Statistics in medicine* 36(30): 4908–4929.

Reich, N. G.; Brooks, L. C.; Fox, S. J.; Kandula, S.; Mc-Gowan, C. J.; Moore, E.; Osthus, D.; Ray, E. L.; Tushar, A.; Yamana, T. K.; et al. 2019a. A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the United States. *Proceedings of the National Academy of Sciences* 116(8): 3146–3154.

Reich, N. G.; McGowan, C. J.; Yamana, T. K.; Tushar, A.; Ray, E. L.; Osthus, D.; Kandula, S.; Brooks, L. C.; Crawford-Crudell, W.; Gibson, G. C.; et al. 2019b. Accuracy of real-time multi-model ensemble forecasts for seasonal influenza in the US. *PLoS computational biology* 15(11).

Rodríguez, A.; Tabassum, A.; Cui, J.; Xie, J.; Ho, J.; Agarwal, P.; Adhikari, B.; and Prakash, B. A. 2020. DeepCOVID: An Operational Deep Learning-driven Framework for Explainable Real-time COVID-19 Forecasting. *medRxiv* .

Sapankevych, N. I.; and Sankar, R. 2009. Time series prediction using support vector machines: a survey. *IEEE Computational Intelligence Magazine* 4(2).

Saputra, M. R. U.; de Gusmao, P. P.; Almalioglu, Y.; Markham, A.; and Trigoni, N. 2019. Distilling knowledge from a deep pose regressor network. In *Proceedings of the IEEE International Conference on Computer Vision*, 263–272.

Shaman, J.; Goldstein, E.; and Lipsitch, M. 2010. Absolute humidity and pandemic versus epidemic influenza. *American journal of epidemiology* 173(2): 127–135.

Shaman, J.; and Karspeck, A. 2012. Forecasting seasonal outbreaks of influenza. *Proceedings of the National Academy of Sciences* 109(50): 20425–20430.

Subbian, K.; and Banerjee, A. 2013. Climate multi-model regression using spatial smoothing. In *Proceedings of the 2013 SIAM International Conference on Data Mining*, 324–332. SIAM.

Takamoto, M.; Morishita, Y.; and Imaoka, H. 2020. An Efficient Method of Training Small Models for Regression Problems with Knowledge Distillation. *arXiv preprint arXiv:2002.12597* .

Tamerius, J. D.; Shaman, J.; Alonso, W. J.; Bloom-Feshbach, K.; Uejio, C. K.; Comrie, A.; and Viboud, C. 2013. Environmental predictors of seasonal influenza epidemics across temperate and tropical climates. *PLoS pathogens* 9(3): e1003194.

Tizzoni, M.; Bajardi, P.; Poletto, C.; Ramasco, J. J.; Balcan, D.; Gonçalves, B.; Perra, N.; Colizza, V.; and Vespignani, A. 2012. Real-time numerical forecast of global epidemic spreading: case study of 2009 A/H1N1pdm. *BMC medicine* 10(1): 165.

Venna, S. R.; Tavanaei, A.; Gottumukkala, R. N.; Raghavan, V. V.; Maida, A. S.; and Nichols, S. 2018. A novel data-driven model for real-time influenza forecasting. *IEEE Access* 7: 7691–7701.

Volkova, S.; Ayton, E.; Porterfield, K.; and Corley, C. D. 2017. Forecasting influenza-like illness dynamics for military populations using neural networks and social media. *PloS one* 12(12): e0188941.

Wang, L.; Chen, J.; and Marathe, M. 2019. DEFSI: Deep learning based epidemic forecasting with synthetic information. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 9607–9612.

Yan, Y.; Li, W.; Wu, H.; Min, H.; Tan, M.; and Wu, Q. 2018. Semi-Supervised Optimal Transport for Heterogeneous Domain Adaptation. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, 2969–2975.

Yuan, Q.; Nsoesie, E. O.; Lv, B.; Peng, G.; Chunara, R.; and Brownstein, J. S. 2013. Monitoring influenza epidemics in china with search query from baidu. *PloS one* 8(5): e64323.

Zhang, Q.; Perra, N.; Perrotta, D.; Tizzoni, M.; Paolotti, D.; and Vespignani, A. 2017. Forecasting seasonal influenza fusing digital indicators and a mechanistic disease model. In *Proceedings of the 26th International Conference on World Wide Web*, 311–319.