DEFSI: Deep Learning Based Epidemic Forecasting with Synthetic Information

LIJING WANG, Computer Science, Virginia Tech
JIANGZHUO CHEN, Biocomplexity Institute & Initiative, University of Virginia
MADHAV MARATHE, Computer Science, University of Virginia

Influenza-like illness (ILI) is among the most common diseases worldwide. Producing timely, well-informed, and reliable forecasts for ILI is crucial for preparedness and optimal interventions. In this work, we focus on short-term but high resolution forecasting and propose DEFSI (*D*eep Learning Based *E*pidemic *F*orecasting with *S*ynthetic *I*nformation), an epidemic forecasting framework that integrates the strengths of artificial neural networks and causal methods. In DEFSI, we builds a two-branch neural network structure to take both within-season observations and between-season observations as features. The model is trained on geographically high-resolution synthetic data. It enables detailed forecasting when high-resolution surveillance data is not available. Our method achieves comparable/better performance than state-of-the-art methods for short-term ILI forecasting at the state level. For high-resolution forecasting at the county level, DEFSI significantly outperforms the other methods.

Additional Key Words and Phrases: Epidemic forecasting, deep neural network, LSTM, causal method, synthetic information

1 INTRODUCTION

Influenza-like illness (ILI) poses a serious threat to global public health. Worldwide seasonal influenza causes about three to five million cases of severe illness and about 290,000 to 650,000 deaths [36]. In the United States, the seasonal influenza has resulted in between 9.2 million and 35.6 million illnesses, between 140,000 and 710,000 hospitalizations, between 12,000 and 56,000 deaths, and is responsible for about $87.1 billion economic lost annually since 2010 [7, 25]. Traditionally, ILI surveillance data from the Centers for Disease Control and Prevention (CDC) has been used as reference data to predict future ILI incidence. The surveillance data is usually updated regularly but often delayed by 1 to 4 weeks and is provided with a coarse resolution. For example, in the USA it has been provided previously at HHS region level and recently at state level. Considering the heterogeneity between different subregion locations and populations, accurate predictions with a finer resolution, e.g. at county level in the USA, is crucial for local public health decision making, optimal mitigation resource allocation among subregions, as well as household or individual level preventive actions informed by neighboring prevalence. We focus on the problem of **high-resolution ILI incidence forecasting** based on ILI surveillance data of coarse resolution.

In this paper we use *flat-resolution* forecasting to denote the forecasting of ILI incidence with the same resolution as the surveillance data; and *high-resolution* forecasting to denote the forecasting with a higher geographical resolution than provided in surveillance data. To be concrete, in this paper flat-resolution means state level while high-resolution means county level, since the highest resolution CDC surveillance data is at state level.

One of the challenges to high-resolution ILI incidence forecasting is the lack of surveillance data at a finer spatial scale. The data driven methods including statistical methods such as ARIMA [5], ARGO [39, 40] and artificial neural networks (ANN) methods such as Long Short Term Memory (LSTM) [33, 34, 37] for ILI forecasting suffer from this limitation. Even for a few states where county level surveillance data is available, training ANN methods for them is difficult due to the small size of

the data. Causal methods are recently introduced to enable high-resolution forecasting [27, 41, 42]. They estimate the parameters of the underlying disease model from the surveillance data. Then ILI incidence prediction is made from the output of simulations using the identified disease model. Depending on the resolution of the causal model, the causal methods can make predictions of various resolutions, even down to individual level in case of a detailed agent-based model. They face challenges including the high dimension of parameter space, which makes searching for globally optimal parameters difficult, as well as the computational complexity of the causal model, especially when it is an agent-based model.

To address the above challenges, we propose a novel epidemic forecasting framework, called **D**eep Learning Based **E**pidemic **F**orecasting with **S**ynthetic **I**nformation (**DEFSI**). It combines **multi-agent system** and **deep neural network** techniques from artificial intelligence (AI). The idea is to model the non-linear relationship between the past higher level (state) ILI incidences and the future higher (state) and lower level (county) ILI incidences with a deep neural network. The success of this idea depends on large amount of realistic training data. The novelty of our approach is to generate the training data using a multi-agent simulation that is based on a synthetic population and contact network, where agent heterogeneities and unstructured interactions among agents are modeled. With a multi-agent model, individual or household level behavior can also be modeled as well as the public health intervention measures, which changes the disease dynamics. The purpose is to use simulations to create training data as similar as possible to the surveillance data observed in the real world. Training data generation is the most compute intensive part of DEFSI, but it can be pre-computed and once the neural networks are trained, they can be applied for forecasting in the whole season with minimal computation.

To the best of our knowledge, DEFSI is the first to combine a realistic multi-agent model with deep learning for epidemic forecasting. **Our major contributions are as follows**:

- DEFSI enables accurate high-resolution forecasting with flat-resolution observations as inputs.
- DEFSI proposes a two-branch neural network model for ILI forecasting. It combines within-season observations (observed data points of the current season that characterize the ongoing epidemic) and between-season historical observations (observed data points from similar weeks of the past seasons that characterize general trends around the current week).
- DEFSI constructs region-specific training dataset at multiple spatially fine-grained scale with low costs. We initialize region-specific simulations with realistic parameter settings learned from the corresponding surveillance data.
- Extensive experiments on ILI incidence forecasting for two states of the USA show that DEFSI achieves comparable/better performance than the state-of-the-art methods at state level. For high-resolution forecasting at county level, DEFSI significantly outperforms the comparison methods.

2 RELATED WORK

Large amounts of research work have made outstanding contributions to the field of infectious disease forecasting. In this section, we review related work that is most relevant to ours. Interested readers can refer to [1, 12, 28] for more details. We discuss the existing ILI forecasting methods categorized based on their underlying methodologies: causal methods, statistical methods, and artificial neural network methods.

2.0.1 Causal Methods. In epidemiology, infectious disease models for within-host progression include: susceptible-infectious-recovered (SIR), susceptible-exposed-infectious-recovered (SEIR),

susceptible-infectious-recovered-susceptible (SIRS), and their extensions [2, 21]. Forecasting methods employing these models are called causal methods because they describe the causal mechanisms of infectious diseases. In these methods the underlying epidemic model can be either a compartmental model (CM) [15, 22, 23] or an agent-based model (ABM) [11, 29]. In a compartmental model, a population is divided into compartments (e.g. S, E, I, R). The differential equations characterize the change of the sizes of each compartments due to disease propagation and progression. In an agent-based model, disease spreads among heterogeneous agents through an unstructured network. The individual level details in an agent-based model can be easily aggregated to obtain epidemic data of any resolution, e.g. number of newly infected people in a county in a specific week. To get county level epidemics in a compartmental model, however, one needs to create compartments in each county, where county population sizes and between county travel data become crucial. Researchers of many scientific community developed innovative methods to predict influenza activity. Shaman et al. [30] developed a framework for initializing real-time forecasts of seasonal influenza outbreaks, using a data assimilation technique commonly applied in numerical weather prediction. Tuite et al. [32] used an SIR CM to estimate parameters and morbidity in pandemic H1N1. Yang et al. [41] applied various filter methods to model and forecast influenza activity using an SIRS CM. In [27], authors proposed an simulation optimization approach based on the SEIR ABM for epidemic forecasting. Hua et al. [19] infer the parameters of the SEIR ABM from social media data for ILI forecasting. *Limitations: There are plenty of parameters within a causal model and each parameter often has extremely high dimensionality. Thus, the optimal estimate is quite difficult in terms of searching and computing. Especially, methods employing agent-based models are computationally intensive in real-time forecasting.*

2.0.2 Statistical methods. Statistical methods employ statistical and time-series based methodologies to learn patterns in historical epidemic outbreaks and leverage those patterns for forecasting. Popular statistical methods for ILI forecasting include e.g. generalized linear models (GLM), autoregressive integrated moving average (ARIMA), and generalized autoregressive moving average (GARMA) [3, 5, 14]. Wang et al. [35] proposed a dynamic Poisson autoregressive model with exogenous input variables (DPARX) for flu forecasting. Yang et al. [40] proposed ARGO, an autoregressive-based influenza tracking model for nowcasting that incorporated CDC ILI data and Google search data. The extensive work based on ARGO are discussed in [39]. *Limitations: Although existing work has made significant achievements in short-term forecasting, none of them could make high-resolution forecasting due to the lack of high-resolution observations for most of regions around the world. In addition, in statistical methods, we often assume a linear relationship between the inputs and outputs, which may not be true for real cases.*

2.0.3 Artificial neural network methods. Artificial neural networks (ANN) have gained increased prominence recently in epidemic forecasting due to their self-learning ability without prior knowledge. Xu et al. [38] first introduced feed-forward neural networks (FNN) into surveillance of infectious diseases and investigated its predictive utility using CDC ILI data, Google search data, and meteorological data. Recurrent Neural Network (RNN) has been demonstrated to be able to capture dynamic temporal behavior of a time sequence. In [34] Volkova et al. built an LSTM model for short-term ILI forecasting using CDC ILI and twitter data. Venna et al. [33] proposed LSTM based method that integrates the impacts of climatic factors and geographical proximity to achieve better forecasting performance. Wu et al. [37] constructed a deep learning structure combining RNN and convolutional neural network to fuse information from different sources. *Limitations: Similar to statistical methods, they could not make high-resolution forecasting without corresponding observations. Moreover, overfitting often happens due to the small training data points available in epidemic history.*

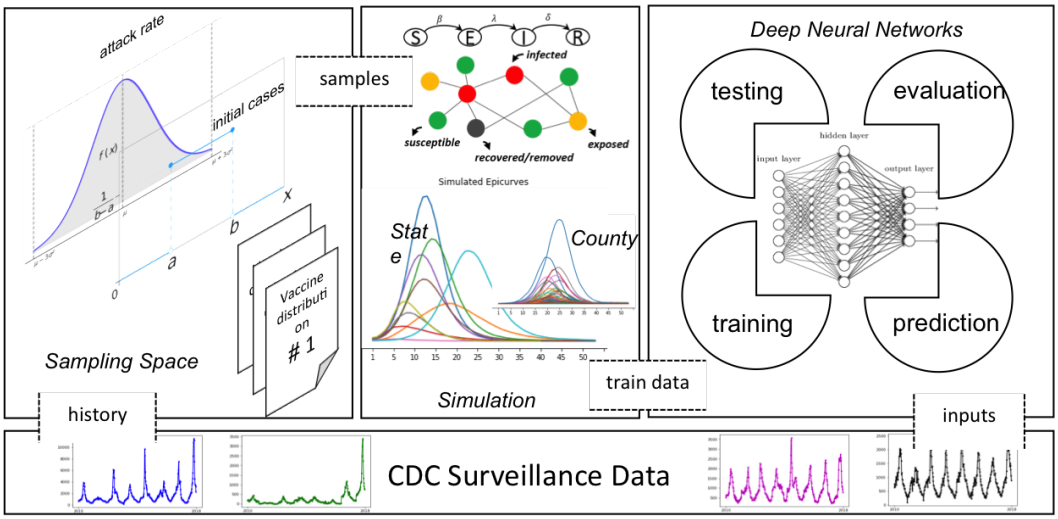


Fig. 1. DEFSI framework. In this framework, a region-specific disease parameter space is constructed based on surveillance data. Then the synthetic training data consisting of both state level and county level weekly ILI incidence curves is generated by simulations parameterized by samples from the parameter space. A two-branch deep neural network model is trained on the synthetic data. The trained model makes forecasting by taking surveillance data as the input.

3 PROBLEM SETUP

Given an observed time-series of weekly ILI incidence for a specific region, we focus on predicting ILI incidence for both the region and its subregions in short-term. Without loss of generality, in this paper we consider making predictions for a state of the United States and all counties in the state, using observations only from CDC state level ILI incidence data [8]. In this setting, state level forecasting is flat-resolution, while county level forecasting is high-resolution. The proposed framework is not limited to this setting and can be generalized for subregion forecasting in any region, e.g. state level forecasting in a country where only national level surveillance data is available. Our proposed method is different from traditional ILI incidence forecasting methods on that the model is trained on synthetic ILI incidence curves but makes forecasting by taking ILI surveillance data as inputs.

Let $\mathbf{y} = \langle y_1, y_2, \cdots, y_T, \cdots \rangle$ denote the sequence of weekly state level ILI incidence, where T is the last week of which the ILI incidence is given. Similarly, $\mathbf{y}^C = \langle y_1^C, y_2^C, \cdots, y_T^C, \cdots \rangle$ denotes the sequence of weekly ILI incidence for a particular county C within the state. Assume that there are K counties $\mathcal{D} = \{C_1, C_2, \cdots, C_K\}$ in the state. Let $\mathbf{y}_t^{\mathcal{D}} = \{y_t^C | C \in \mathcal{D}\}$ denote ILI incidence of all counties in the state at week t. The consistency constraint on county-level incidence is $y_t = \sum_{C \in \mathcal{D}} y_t^C$. The problem is defined as predicting both state level and county level incidence at week t, where $t = T + 1$, denoted as $\mathbf{z}_t = (y_t, \mathbf{y}_t^{\mathcal{D}})$, given historical state level incidence.

4 DEFSI

4.1 Framework

DEFSI framework consists of three major components (shown in Fig. 1): (i) *Disease model parameter space construction*: Given an existing disease simulator, we estimate a marginal distribution for each

model parameter based on the state-specific surveillance data; (ii) *Synthetic training data generation*: We generate a synthetic training dataset at both flat-resolution and high-resolution scales for a state by running simulations parameterized from the corresponding parameter space; (iii) *Deep neural network training and forecasting*: We design a two-branch deep neural network model. It is trained on the synthetic training dataset and makes predictions using surveillance data as its inputs. We will elaborate the details in the following subsections.

4.2 SEIR-based Epidemic Simulation

The SEIR disease model is widely used for ILI diseases [21]. Each person is in one of the following four health states at any time: susceptible (S), exposed (E), infectious (I), recovered or removed (R). A person v is in the susceptible state until he becomes exposed. If v becomes exposed, he remains so for $p_E(v)$ days, which is called the incubation period, during which he is not infectious. Then he becomes infectious and remains so for $p_I(v)$ days, which is called the infectious period. Finally he becomes removed (or recovered) and remains so permanently. While the SEIR model characterizes within-host disease progression, between-host disease propagation is modeled by transmission from person to person with a probability parameter τ, through either complete mixing or heterogeneous connections among people.

In this work, we adopt an agent-based simulator EpiFast [6]. The outputs are individual level infected cases with the infected days of a simulated season. They can be aggregated to any temporal and spatial scale, such as daily (weekly) state (county) level ILI incidence. Vaccine intervention I_V (i.e. quantity of vaccines applied to the population and the timing of the application) is applied in our simulations. A distribution on the parameter space $\mathcal{P}(p_E, p_I, \tau, N_I, I_V)$ is estimated from CDC historical data, where N_I denotes the number of infections at the beginning of a flu season.

4.3 Disease Model Parameter Space

For clarity, we define an epidemiological week in a calendar year as **ew**, and a season week in a flu season as **sw**. Note that ew is only meaningful in real flu seasons (i.e. surveillance data) while sw is meaningful for both surveillance and simulated flu seasons. The historical time series of CDC surveillance data used to construct parameter space is segmented into seasons by cutting at $ew(40)$ of each year (i.e. $ew(40)$ of a calendar year corresponds to $sw(1)$ of a flu season). Among $\mathcal{P}$, (p_E, p_I) are known from literature [24]. We assume (τ, N_I, I_V) follows distributions that can be estimated from historical data.

Firstly, for a given state, we collect observations of each parameter value by the following ways:

- **Transmissibility (τ)**: We compute season attack rate ar (i.e. fraction of population getting infected in the season) for the target state and its neighbor states (i.e. geographically contiguous states) of each historical season. We calibrate a transmissibility value for each of ar as the solution to $\min_\tau |AR(PS(\tau)) - ar|$, where $AR(\cdot)$ computes attack rate from the output of $PS(\cdot)$.
- **Initial Case Number (N_I)**: This is the number of infections at the beginning of a flu season. We collect the ILI incidence of the first week of each season for the target state and its neighbors.
- **Vaccine Intervention (I_V)**: Collect 6 schedules extracted from the past six influenza seasons in the United States [9]. Each schedule consists of timing and percentage coverage of vaccine application throughout the season. We assume that the state level vaccine schedule is the same as the nationwide schedule.

Secondly, for τ and N_I, we fit the collected samples to several distributions including normal, exponential, gamma, and uniform. Then we run KS-test to choose a distribution with highest

ALGORITHM 1: Training Dataset Generation for DEFSI

Input: Simulator PS, and Parameter space $\mathcal{P}$.
Output: Simulated epicurves $\Omega = \{(\mathbf{y}_{(i)}, \mathbf{y}^{\mathcal{D}}_{(i)}) | i = 1, 2, \cdots, r\}$.

$\Omega = \emptyset$;
for $i = 1$ *to* r **do**
  P = Sample($\mathcal{P}$);
  $(\mathbf{y}_{(i)}, \mathbf{y}^{\mathcal{D}}_{(i)})$ = PS(P);
  $\Omega = \Omega \cup (\mathbf{y}_{(i)}, \mathbf{y}^{\mathcal{D}}_{(i)})$
end

significance (refer to [31] for more details). For I_V, we assume the 6 vaccination schedules follow a discrete uniform distribution. In this way, a region-specific parameter space $\mathcal{P}$ is constructed.

Any intervention actions taken during a flu season will obviously affect the disease spread. Thus, a $\mathcal{P}$ including realistic interventions will lead to a more realistic synthetic training dataset, which will improve forecasting performance of our method. We will discuss the significance of I_V for good forecasting performance of DEFSI in The Significance of I_V in $\mathcal{P}$

4.4 Training Dataset from Simulations

For each run of a simulation, a specific parameter setting is sampled from $\mathcal{P}$, and the simulator is called to generate individual level outcomes (i.e. who and when gets infected). These outcomes will be aggregated to state and county level weekly incidence according to the demographics of the synthetic population, called *synthetic epicurve*. Week 1 in the synthetic epicurve corresponds to $sw(1)$ of a flu season. Large volumes of high-resolution synthetic data are generated by repeating the sampling and simulating process. Let us denote $\Omega = \{(\mathbf{y}_{(i)}, \mathbf{y}^{\mathcal{D}}_{(i)}) \in \mathbb{R}^{\ell \times (K+1)} | i = 1, 2, \cdots, r\}$ as all simulated epicurves, where ℓ is the length of an epicurve (number of weeks), K is the number of counties in the state, and r is the total number of simulation runs. Algorithm 1 describes the generating process.

Compared with CDC surveillance data, the training dataset Ω is prominent in two aspects: (i) it includes high-resolution information; (ii) the large volume of realistic training data reduce the possibility of overfitting when training a deep learning model. Thus the trained model is likely to have a lower error rate on new unseen seasons.

4.5 DEFSI Neural Network Model

In traditional time series modeling problem, ILI incidences of the few previous weeks are used as the observations for the prediction of the current week.

In DEFSI, we use two kinds of observations: (1) ***Within-season observations***, denoted as $\mathbf{x1} = \langle y_{t-a}, \cdots, y_{t-1} \rangle$, are ILI incidence from previous a weeks of the current season. They are used as the main observations to follow the weekly trend. (2) ***Between-season observations***, denoted as $\mathbf{x2} = \langle y_{t-\ell*b}, \cdots, y_{t-\ell*1} \rangle$, are ILI incidences of the same sw from the past b seasons. They are used as the surrogate information to improve forecasting performance. As shown in Fig. 2, for example, there are 4 seasons ordered by sw. The within-season observations are ILI incidence of previous $a = 3$ weeks in current season. The between-season observations are ILI incidence of the same $sw(t)$ from the past $b = 3$ seasons.

The Long Short Term Memory (LSTM) network [18] is adopted in our neural network architecture to capture the dynamic temporal behavior of the observations. An LSTM layer consists of a sequence of $a(b)$ cells, of which the current cell takes one ILI incidence from $\mathbf{x1}(\mathbf{x2})$ as well as the output and

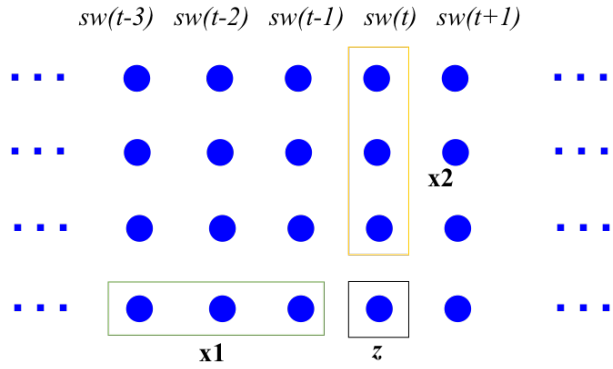


Fig. 2. Within-season and between-season observations for forecasting. In this graph, there are four flu seasons (rows). Nodes in each row denote weekly ILI incidence in each season, which are ordered by sw. For a target week $sw(t)$ (black square), the model observes two kinds of information: (1) within-season observations **x1** - the ILI incidence from previous weeks of the current season (green rectangular); (2) between-season observations **x2** - the historical ILI incidence from similar weeks of the past seasons (yellow rectangular).

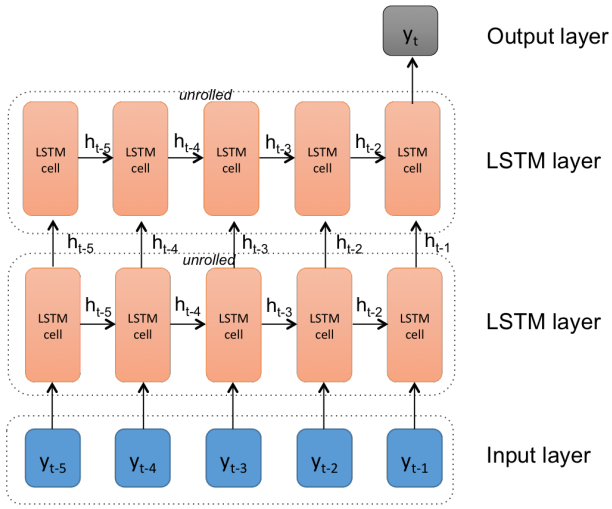


Fig. 3. Stacked LSTM layers. There are two LSTM layers. Each layer consists of 5 cells, which means it takes observations of previous 5 time points and outputs predictions for the current time point. For example, the LSTM layers designed for the input **x1** in Fig. 2 should consist of 3 cells, and **x2** is in a similar way.

the cell state of its previous cell as inputs. Let $\mathbf{i}_{t-1}, \mathbf{f}_{t-1}, \mathbf{o}_{t-1}, \mathbf{C}_{t-1}, \mathbf{h}_{t-1}$ denote input gate, forget

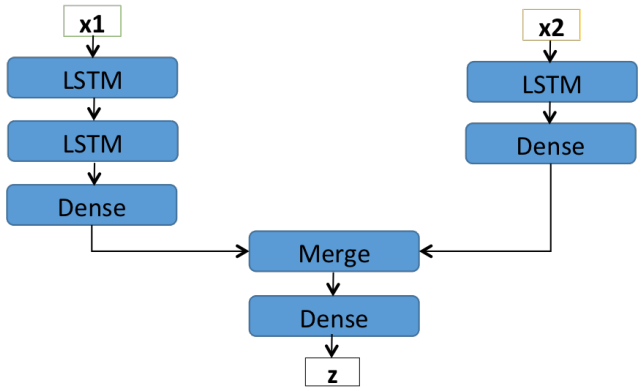


Fig. 4. DEFSI neural network architecture. This architecture consists of two branches. The left branch consists of stacked LSTM layers that encode within-season observations **x1**, and the right branch is designed to be LSTM based layers that encodes between-season observations **x1**. A merge layer is added to combine two branches and the output **z** is state-level and county-level predictions.

gate, output gate, cell state, and output of the last cell in the layer. It is described mathematically as:

$$
\begin{aligned}
\mathbf{i}_{t-1} &= \sigma(\mathbf{W}_i \cdot y_{t-1} + \mathbf{U}_i \cdot \mathbf{h}_{t-2} + \mathbf{b}_i) \\
\mathbf{f}_{t-1} &= \sigma(\mathbf{W}_f \cdot y_{t-1} + \mathbf{U}_f \cdot \mathbf{h}_{t-2} + \mathbf{b}_f) \\
\mathbf{o}_{t-1} &= \sigma(\mathbf{W}_o \cdot y_{t-1} + \mathbf{U}_o \cdot \mathbf{h}_{t-2} + \mathbf{b}_o) \\
\widetilde{C}_{t-1} &= tanh(\mathbf{W}_C \cdot y_{t-1} + \mathbf{U}_C \cdot \mathbf{h}_{t-2} + \mathbf{b}_C) \\
\mathbf{C}_{t-1} &= \mathbf{f}_{t-1} \circ \mathbf{C}_{t-2} + \mathbf{i}_{t-1} \circ \widetilde{C}_{t-1} \\
\mathbf{h}_{t-1} &= \mathbf{o}_{t-1} \circ \mathbf{C}_{t-1}
\end{aligned}
\tag{1}
$$

where σ and $tanh$ are sigmoid and tanh activation functions. $\mathbf{W}, \mathbf{U}$, and $\mathbf{b}$ are adaptive weights and bias. y_{t-1} is the ILI incidence at time point $t-1$. $\mathbf{C}_{t-2}, \mathbf{h}_{t-2}$ are the cell state and output of the previous cell. Operator $\circ$ denotes element wise product (Hadamard product). Fig. 2 shows how the cells work in LSTM layers.

In DEFSI model, we design a two-branch LSTM based deep neural network model to capture temporal dynamics of within-season observations and between-season observations. As shown in Fig. 4, the left branch consists of stacked LSTM layers that encode within-season observations $\mathbf{x1} = \langle y_{t-a}, \cdots, y_{t-1} \rangle$. The right branch is a single LSTM layer that encodes between-season observations $\mathbf{x2} = \langle y_{t-\ell * b}, \cdots, y_{t-\ell * 1} \rangle$. A merge layer is added to combine the outputs of two branches. The final output dimension is the same as $\mathbf{z}_t$ which is state-level and county-level predictions (as defined in Problem Setup). This LSTM based deep learning model is able to connect historical ILI incidence information to the current prediction. It also allows long-term dependency learning without gradient vanishing problem.

We are interested in a predictor f, which predicts the current week's state level and county level incidence $\mathbf{z}_t$ based on the previous a weeks of within-season state level ILI incidence **x1** and the previous b seasons of between-season state level ILI incidence **x2**:

$$
\hat{\mathbf{z}}_t = f([\mathbf{x1}, \mathbf{x2}]_t, \theta) \tag{2}
$$

where θ denotes parameters of the predictor, $\hat{\mathbf{z}}_t$ denotes the prediction of $\mathbf{z}_t$. Note that **the output of f is always one week ahead forecast** in our model.

The loss function $\mathcal{L}$ is defined as mean-square-error, with consistency constraint on outputs:

$$\min_{\theta} \mathcal{L}(\theta) = \sum_{t} ||\mathbf{z}_t - f([\mathbf{x1}, \mathbf{x2}]_t, \theta)||^2,$$
$$\text{with consistency constraint } \hat{y}_t = \sum_{C \in \mathcal{D}} \hat{y}_t^C \tag{3}$$

Adam optimization algorithm [20] is used to learn θ. An activity regularizer is added to $\hat{\mathbf{z}}_t$ for consistency constraint.

Training and forecasting: In the training process, we use synthetic training data Ω to train DEFSI model. The historical surveillance data is only used for constructing the disease model parameter space $\mathcal{P}$. In the predicting process, the trained model takes state-level surveillance observations as the inputs and makes one week ahead forecasting on both state-level and county-level. All region specific DEFSI models are trained once before the target flu season start, then can be used for forecasting throughout the season.

4.6 Variants of DEFSI Model

The two-branch neural network architecture is flexible for multiple variants: (1) ***DEFSI***: Two-branch neural network as shown in Fig. 4. (2) ***DEFSI-L***: Only the left branch is used to take within-season observations. (3) ***DEFSI-RDENSE***: Changing the LSTM layer of the right branch with Dense layers, which means that the model do not care about the temporal relationship between between-season data points. We will discuss the results of different variants in Experiments.

4.7 Multi-step Forecasting

In practical situations, we are interested in making predictions for several weeks ahead. In DEFSI, the left branch of the model appends the most recent state level prediction to the input for predicting the target at the upcoming week, and the right branch uses the state level ILI incidences from the past seasons with sw equals to the upcoming week number.

5 EXPERIMENTS

5.1 Datasets

CDC ILI incidence [8]: The CDC surveillance data used in the experiments is the weekly ILI incidence at state level from 2010 $ew(40)$ to 2018 $ew(18)$. **ILI Lab tested flu positive counts of New Jersey** [13]: To evaluate the county-level forecasting performance, we collect state-level and county-level ILI Lab tested flu positive counts of season 2017-2018 in NJ. The data is available from $ew(40)$ to the next year's $ew(20)$. We use it as the ground truth when evaluating county-level forecasting. **Google data** [16, 17]: The Google correlate terms of each state are queried. Then the Google Health Trends of each correlated terms for each state is collected and aggregated weekly from 2010 $ew(40)$ to 2018 $ew(18)$. **Weather data** [10]: We download daily weather data (including max temperature, min temperature, precipitation) from Climate Data Online (CDO) for each state and aggregate the daily data to average weekly data from 2010 $ew(40)$ to 2018 $ew(18)$. Google data and weather data are used as surrogate information in comparison methods (described in Comparison Methods).

5.2 Comparison Methods

Our method is compared with 5 state-of-the-art methods from ANN methods, statistical methods, and causal methods, respectively. They are *LSTM* (single layer LSTM) [18] and *AdapLSTM* (CDC + Weather data) [33] from artificial neural network methods; *ARIMA* (classic ARIMA) [5] and *ARGO* (CDC + Google data) [40] from statistical methods; and *EpiFast* [4] from agent-based causal models.

AdapLSTM, LSTM, ARGO, and ARIMA are used for state level forecasting. EpiFast is applied for both state level and county level forecasting.

5.3 Experiment Setup

Our experiments are performed on two states: Virginia (VA) and New Jersey (NJ). For each state, we separate the time sequence into 8 flu seasons from 2010-2011 to 2017-2018 by cutting at $ew(40)$. We use the dataset of 2010-2011 to 2016-2017 seasons as training dataset; and use the dataset of 2017-2018 season for testing. In DEFSI, the training dataset is used to estimate disease parameter space, while for other comparison methods, it is used for training forecasting model directly. The collected county-level ILI data of NJ is only used for evaluating. At each time step in the testing season, each model makes predictions 5 weeks ahead, i.e. $horizon = \{1, 2, 3, 4, 5\}$. More detailed settings (including estimated parameter space in DEFSI, parameter settings for comparison methods) are elaborated in Appendix.

5.4 Performance Metrics

The metrics used to evaluate the forecasting performance are: *root mean squared error (RMSE), mean absolute percentage error (MAPE), Pearson correlation (PCORR).*

Root mean squared error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{4}$$

Mean absolute percentage error (MAPE):

$$MAPE = (\frac{1}{n}\sum_{i=1}^{n}|\frac{y_i - \hat{y}_i}{y_i}|) * 100 \tag{5}$$

Pearson correlation (PCORR):

$$PCORR = \frac{cov(\mathbf{y}, \hat{\mathbf{y}})}{\sigma_{\mathbf{y}}\sigma_{\hat{\mathbf{y}}}} \tag{6}$$

where $cov(\mathbf{y}, \hat{\mathbf{y}})$ is the covariance of $\mathbf{y}$ and $\hat{\mathbf{y}}$, and σ is the standard deviation.

Among these metrics, RMSE and MAPE evaluate ILI incidence prediction accuracy, PCORR evaluates linear correlation between the true curve and the predicted curve.

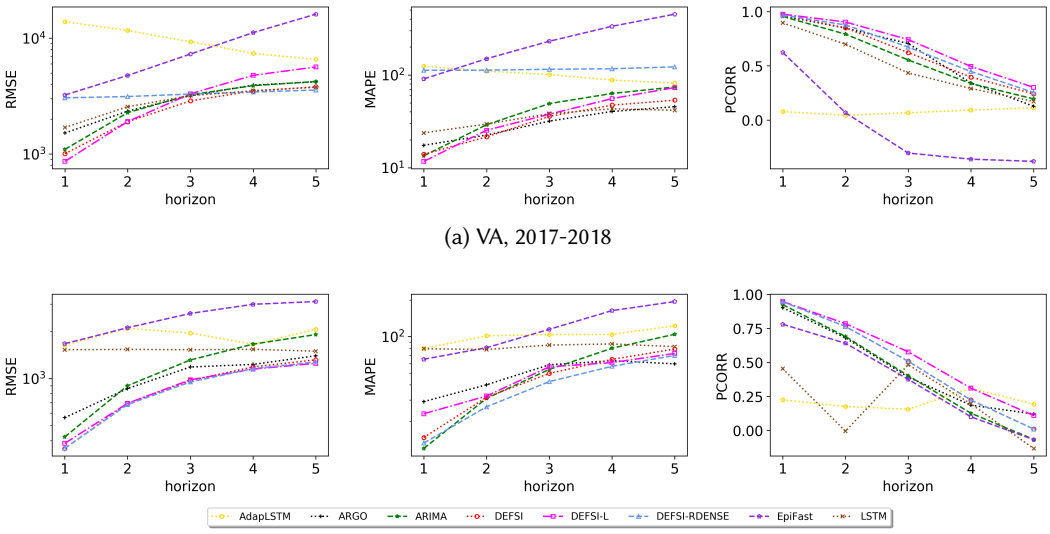
5.5 Performance of Flat-resolution Forecasting

We forecast state-level ILI incidence for VA, 2017-2018 and NJ, 2017-2018. Fig. 5 shows the forecasting performance on RMSE, MAPE, PCORR.

(1) *Performance on RMSE (left column of Fig. 5)*: In VA, DEFSI, DEFSI-L, DEFSI-RDENSE, ARIMA, LSTM achieve similar performance that are better than EpiFast, AdapLSTM. In NJ, DEFSI and its variants consistently outperform others across the horizon, which on average have 15%-64% improvement on RMSE. LSTM is not robust to different states.

(2) *Performance on MAPE (middle column of Fig. 5)*: In VA, ARGO, DEFSI, DEFSI-L, ARIMA, and LSTM have similar performance that is better than AdapLSTM, DEFSI-RDENSE, and EpiFast. In NJ, DEFSI-RDENSE achieves the best performance closely followed by DEFSI. And DEFSI-L, ARIMA, and ARGO achieve comparable performance with the best one that is better than LSTM, AdapLSTM, and EpiFast. LSTM and DEFSI-RDENSE is not robust to different states.

(3) *Performance on PCORR (right column of Fig. 5)*: In both VA and NJ, DEFSI-L outperforms others on pearson correlation (i.e. around 0.96 with horizon 1) which on average have 20%-100%



(a) VA, 2017-2018

(b) NJ, 2017-2018

Fig. 5. State level ILI incidence forecasting performance on RMSE, MAPE, PCORR. A log y-scale is used in RMSE and MAPE.

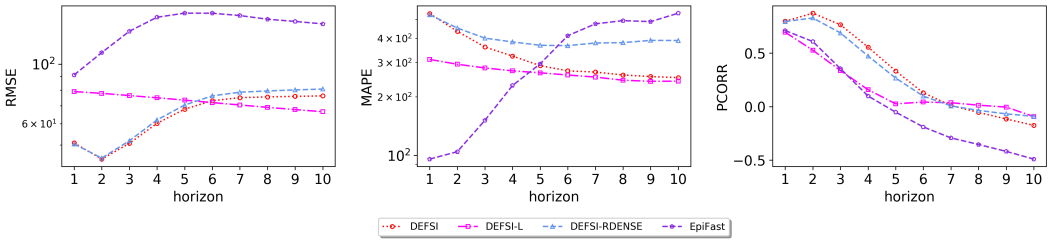


Fig. 6. County level ILI incidence forecasting performance on RMSE, MAPE, PCORR for NJ, 2017-2018. A log y-scale is used in RMSE and MAPE.

improvement compared with others, followed by DEFSI, DEFSI-RDENSE, and ARGO. LSTM is not robust to different states.

Overall, three major observations: (1) Our DEFSI and its variants achieve comparable/better performance than the other methods on state level ILI forecasting. (2) The methods of EpiFast and AdapLSTM perform not very well for both states on RMSE and MAPE. (3) LSTM is not robust to different states on all there metrics. We will briefly discuss the possible reasons for the above observations in Discussion.

5.6 Performance of High-resolution Forecasting

The performance of county-level forecasting is evaluated on NJ 2017-2018. The horizon is extended to 10 for better observations. In Fig. 6, we show county level ILI forecasting performance on three metrics. The metric value of each node in the figure is the average value across 21 counties in NJ. Our method consistently outperforms the comparison method EpiFast on RMSE (about 53%

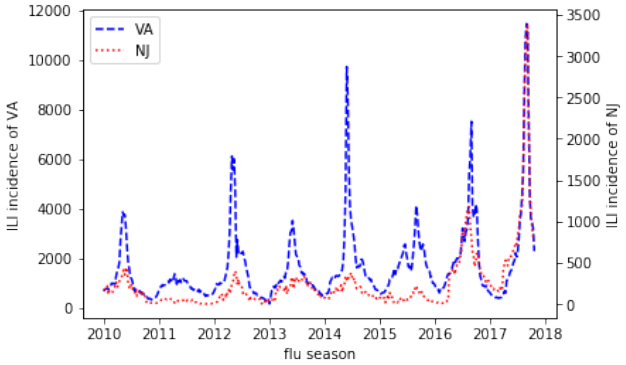


Fig. 7. CDC surveillance ILI incidence of VA (blue dash line) and NJ (red dot line). It is obvious to observe that, for testing season 2017-2018, a similar epi-curve (i.e. similar curve shape and the peak size) occurs at season 2014-2015 in VA, while no similar seasons could be found in NJ.

improvement) and PCORR (about 60% improvement). Among DEFSI models, DEFSI and DEFSI-RDENSE perform better than DEFSI-L, which indicates that the between-season observations are helpful for improving forecasting accuracy.

However, for MAPE, EpiFast performs better than our method with horizon less than 4, while the error increases dramatically as the horizon increases. This implies that EpiFast can make good prediction on short-term on county-level forecasting. According to the definition 5, MAPE is biased by small values of ILI incidence. By observing the ground truth of NJ county-level ILI incidence, we found that many counties have small infected numbers during the whole flu season. EpiFast may benefit from these small values. Overall, our method significantly outperforms the comparison method on county level forecasting.

5.7 Discussion

In general, AdapLSTM and EpiFast do not perform very well in our experiment compared with other methods.

For AdapLSTM, weather factors is considered for post adjustment of LSTM outputs. As stated in [33], the weather factors are estimated using time delays computed by apriori associations and selected by the largest confidence. However, in our experiment, they all show very low confidences (less than 0.3). This may probably cause arbitrary adjustment for predictions and consequently poor performance.

For EpiFast, one possible reason is that we did not find a good estimate of the underlying disease model for a specific region and season due to the noisy CDC observations. As we discussed in Introduction, if the observed data is too noisy then the learned underlying model tends to make predictions with large errors.

LSTM is not robust to different states. Specifically, LSTM performs relatively better on VA than on NJ. Similarly, ARIMA and ARGO also perform slightly better on VA than on NJ. One possible reason is that the trained models are overfitting due to the small size of the training dataset from CDC surveillance. This is especially true for LSTM which needs to learn large number of parameters. The performance of these models depend on whether a similar epicurve occurred in previous seasons. As shown in Fig. 7, the epicurve of VA 2017-2018 is similar to that of VA 2014-2015. However, the epicurve of NJ 2017-2018 seems to be much higher than all previous ones. On the contrary, DEFSI

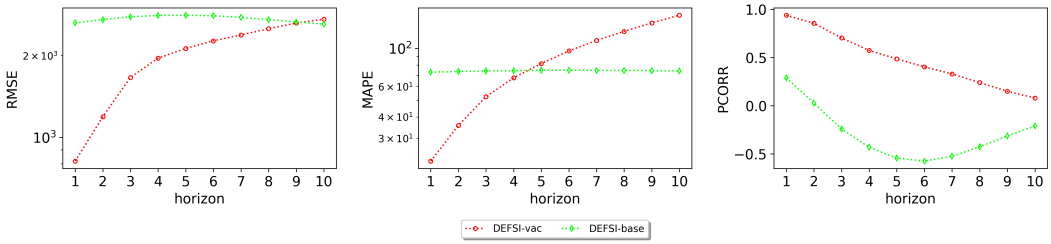


Fig. 8. Performance comparison between DEFSI models trained on base-case synthetic training dataset (DEFSI-base) and vaccine-case synthetic training dataset (DEFSI-vac). They test on VA, 2017-2018 with horizon up to 10 weeks ahead.

models have stable performance on both VA and NJ. They are kind of avoiding this overfitting through training on the large volume of synthetic training data.

Through the results, DEFSI enables high-resolution forecasting that outperforms EpiFast. Meanwhile, it achieves comparable/better performance than the comparison methods at state level forecasting. And in our framework, the large volume of realistic simulated data allows us to train a more complex DNN model and reduces the possibility of overfitting. Our experiments demonstrate that DEFSI integrates the strengths of ANN methods and causal methods to improve epidemic forecasting.

5.8 Consistency Constraint

We demonstrate the capability of DEFSI models in satisfying the consistency constraint in (3). In one case, we explicitly add an regularizer to the output of the model, whereas in the other case, no extra regularizer is used. Results show no significant difference between the outputs of two cases. DEFSI could always learn this linear constraint without an explicit regularizer. We omit the comparisons for brevity.

5.9 The Significance of I_V in $\mathcal{P}$

Our method described above constructs a parameter space $\mathcal{P}(p_E, p_I, \tau, N_I, I_V)$ that includes the vaccine intervention I_V. In this section, we investigate how I_V affects the performance of DEFSI by generating two synthetic training datasets: (1) ***vaccine-case***: simulations with I_V (the training dataset used in Experiments); (2) ***base-case***: simulations that share the common settings of p_E, p_I, τ, N_I with vaccine-case except $I_V = \emptyset$.

We train DEFSI on vaccine-case and base-case respectively with the same settings described in Experiment Setup, denoted as ***DEFSI-vac*** and ***DEFSI-base***. State level forecasting on VA 2017-2018 is evaluated. Fig. 8 shows the performance on RMSE, MAPE, and PCORR. DEFSI-vac significantly outperforms DEFSI-base on RMSE and PCORR, especially with small horizon less than 5. On MAPE, DEFSI-vac performs better than DEFSI-base with small horizon less than 5, while the situation inverses as the horizon increases. This indicates that realistic interventions can significantly improve the short-term forecasting.

Our experiments show the significance of realistic interventions in $\mathcal{P}$ for good forecasting performance of our method. Our proposed framework is extensible for further available realistic interventions, such as school closure and quarantine, to further improve forecasting performance.

Table 1. Marginal distributions of each parameters

Parameter	State	Name	Distribution	P-value
p_E	VA	-	(1:0.3, 2:0.5, 3:0.2) [24]	-
	NJ	-	(1:0.3, 2:0.5, 3:0.2) [24]	-
p_I	VA	-	(3:0.3, 4:0.4, 5:0.2, 6:0.1) [24]	-
	NJ	-	(3:0.3, 4:0.4, 5:0.2, 6:0.1) [24]	-
τ	VA	Normal	$\mathcal{N}(\mu = 4.88\text{e}{-5}, \delta = 9.33\text{e}{-7})$	0.74
	NJ	Normal	$\mathcal{N}(\mu = 4.63\text{e}{-5}, \delta = 1.05\text{e}{-6})$	0.85
N_I	VA	Uniform	$\mathcal{U}(7224, 14798)$	0.99
	NJ	Exponential	$exp(\mu = 708, \beta = 2013)$	0.93
I_V	VA	Discrete Uniform	6 vaccination schedules [9]	0.99
	NJ	Discrete Uniform	6 vaccination schedules [9]	0.93

6 CONCLUSION

In this paper we propose DEFSI, a novel epidemic forecasting framework combining deep neural network methods with casual models. In DEFSI, a two-branch neural network model and its variants are designed to combine within-season and between-season observations. The model is trained on region-specific synthetic dataset constructed at multiple spatially fine-grained scale. The trained model enables accurate high-resolution forecasting with flat-resolution observations as inputs. Extensive experiments on NJ and VA showed that DEFSI achieved comparable/better performance than the state-of-the-art methods on state level forecasting and consistently better performance than others on county level forecasting.

A APPENDIX

A.1 Fitting distributions for disease parameters

In $\mathcal{P}(p_E, p_I, \tau, N_I, I_V)$, (τ, N_I) are fitted distributions learned from collected samples through the following way:

Neighbors of VA: ['Kentucky', 'Maryland', 'North Carolina', 'Tennessee', 'West Virginia']. For all states and seasons, there are totally $7 * 6 = 42$ (ar, N_I) samples collected from 2010 to 2016.

Neighbors of NJ: ['Delaware', 'New York', 'Pennsylvania']. For all states and seasons, there are totally $4 * 6 = 42$ (ar, N_I) samples collected from 2010 to 2016.

τ is calibrated using EpiFast by Nelder-Mead [26] algorithm based on each pair of (ar, N_I).

We fit the samples to several reference distributions, including exponential, normal, uniform, gamma, and lognorm distributions. Then we run the KS-test (the null hypothesis being that the sample is drawn from the reference distribution) to choose a distribution with the highest significance. The best fitted distributions with the significance value for each parameter of each state are shown in Table 1.

A.2 Experiment Settings

We elaborate the details of settings for each method in the following.

(1) *DEFSI*: For each state, we generated $r = 5000$ simulated curves with the length $\ell = 52$ of weekly ILI incidence on both state level and county level. The simulations are initializing by samples from parameter space $\mathcal{P}$ (shown in Table 1). In DEFSI model, we set look back window size of within-season observations as 10, and between-season observations as 5.

(2) *Single layer LSTM model (LSTM)*: This is a baseline from artificial neural network methods. No surrogate indicator used in this model. We set the look back window size as 10.

(3) *AdapLSTM* [33]: This method makes predictions using a simple LSTM model, then adjusts the predictions by applying impacts of weather factors and spatio-temporal factors. LSTM model has the same setting with (2) in our experiment. In [33], the weather data features include maximum temperature, minimum temperature, humidity, and precipitation, while humidity is not included in our experiment. The confidences of symbolic pairs in our experiment are less than 0.3, which will lead to arbitrary adjustment for predictions. The neighbors of each state used for spatio-temporal adjustment factor are the same with neighbors described in **??**.

(4) *Simple ARIMA model (ARIMA)*: This is a baseline from statistical methods. No exogenous variable used in this model. The order for ARIMA is $(10, 1, 0)$.

(5) *AutoRegression with GOogle search data (ARGO)* [40]: The method proposes an autoregression model utilizing Google search data. We use the public available tool from [40]. In our experiment, We set look back window size as 52 and training window as 104. All of the top 100 Google correlate terms of VA are flu related, while only less than 1% of the top 100 Google correlated terms of NJ are flu related.

(6) *EpiFast* [4]: This model follows the definition in SEIR-based epidemic simulation, the parameters tuned in EpiFast are p_E, p_I, N_I, τ. They are optimized by minimizing the error of the predicted and the actual ILI incidence via Nelder-Mead method [26].

REFERENCES

[1] Ali Alessa and Miad Faezipour. 2018. A review of influenza detection and prediction through social networking sites. *Theoretical Biology & Medical Modelling* 15 (2018), 2. https://doi.org/10.1186/s12976-017-0074-5

[2] Norman T. J. Bailey. 1975. *The Mathematical Theory of Infectious Diseases and Its Applications* (2 ed.). Griffin.

[3] B. Bardak and M. Tan. 2015. Prediction of influenza outbreaks by integrating Wikipedia article access logs and Google flu trend data. In *IEEE 15th BIBE*. 1–6. https://doi.org/10.1109/BIBE.2015.7367640

[4] Richard J. Beckman, Keith R. Bisset, Jiangzhuo Chen, Bryan L. Lewis, Madhav V. Marathe, and Paula Elaine Stretz. 2014. ISIS: a networked-epidemiology based pervasive web app for infectious disease pandemic planning and response. In *KDD*.

[5] Michael A. Benjamin, Robert A. Rigby, and D. Mikis Stasinopoulos. 2003. Generalized Autoregressive Moving Average Models. *J. Amer. Statist. Assoc.* 98, 461 (2003), 214–223.

[6] Keith R. Bisset, Jiangzhuo Chen, Xizhou Feng, V.S. Anil Kumar, and Madhav V. Marathe. 2009. EpiFast: A Fast Algorithm for Large Scale Realistic Epidemic Simulations on Distributed Memory Systems. In *Proceedings of the 23rd ICS*. ACM, 430–439.

[7] CDC. 2018. Disease Burden of Influenza. https://www.cdc.gov/flu/about/disease/burden.htm. Accessed May 01, 2018.

[8] CDC. 2018. Fluview Interactive. https://www.cdc.gov/flu/weekly/fluviewinteractive.htm. Accessed April 20, 2018.

[9] CDC. 2018. Historical Seasonal Influenza Vaccine Schedule. https://www.cdc.gov/flu/professionals/vaccination/vaccinesupply.htm. Accessed November 1, 2017.

[10] CDO. 2018. Climate Data Online. https://www.ncdc.noaa.gov/cdo-web/datasets. Accessed August 28, 2018.

[11] Dennis L. Chao, M. Elizabeth Halloran, Valerie J. Obenchain, and Ira M. Longini, Jr. 2010. FluTE, a Publicly Available Stochastic Influenza Epidemic Simulation Model. *PLOS Computational Biology* 6, 1 (01 2010), 1–8. https://doi.org/10.1371/journal.pcbi.1000656

[12] Jean Paul Chretien, Dylan George, Jeffrey Shaman, Rohit A. Chitale, and F. Ellis McKenzie. 2014. Influenza Forecasting in Human Populations: A Scoping Review. *PLOS ONE* 9, 4 (04 2014), 1–8. https://doi.org/10.1371/journal.pone.0094130

[13] DOH. 2018. ILI Weekly Reports. http://www.nj.gov/health/cd/statistics/flu-stats/. Accessed April 20, 2018.

[14] Andrea Freyer Dugas, Mehdi Jalalpour, Yulia Gel, Scott Levin, Fred Torcaso, Takeru Igusa, and Richard E. Rothman. 2013. Influenza Forecasting with Google Flu Trends. *PLoS ONE* 8, 2 (2013), e56176.

[15] Antoine Flahault, Elisabeta Vergu, Laurent Coudeville, and Rebecca F. Grais. 2006. Strategies for containing a global influenza pandemic. *Vaccine* 24, 44 (2006), 6751–6755. https://doi.org/10.1016/j.vaccine.2006.05.079

[16] GHT. 2018. Google Health Trends. https://trends.google.com/trends. Accessed August 28, 2018.

[17] Google. 2018. Google Correlate Data. https://www.google.com/trends/correlate. Accessed August 28, 2018.

[18] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Computation* 9, 8 (1997), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735 arXiv:https://doi.org/10.1162/neco.1997.9.8.1735

[19] Ting Hua, Chandan K Reddy, Lei Zhang, Lijing Wang, Liang Zhao, Chang-Tien Lu, and Naren Ramakrishnan. 2018. Social Media based Simulation Models for Understanding Disease Dynamics. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*. International Joint Conferences on Artificial Intelligence Organization, 3797–3804.

[20] Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *CoRR* abs/1412.6980 (2014). arXiv:1412.6980 http://arxiv.org/abs/1412.6980

[21] Yu. A. Kuznetsov and C. Piccardi. 1994. Bifurcation analysis of periodic SEIR and SIR epidemic models. *Journal of Mathematical Biology* 32, 2 (01 01 1994), 109–121. https://doi.org/10.1007/BF00163027

[22] Jung Min Lee, Donghoon Choi, Giphil Cho, and Yongkuk Kim. 2012. The effect of public health interventions on the spread of influenza among cities. *Journal of Theoretical Biology* 293 (2012), 131–142. https://doi.org/10.1016/j.jtbi.2011.10.008

[23] Antonella Lunelli, Andrea Pugliese, and Caterina Rizzo. 2009. Epidemic patch models applied to pandemic influenza: Contact matrix, stochasticity, robustness of predictions. *Mathematical Biosciences* 220, 1 (2009), 24–33. https://doi.org/10.1016/j.mbs.2009.03.008

[24] Achla Marathe, Bryan Lewis, Jiangzhuo Chen, and Stephen Eubank. 2011. Sensitivity of Household Transmission to Household Contact Structure and Size. *PLoS ONE* 6 (08 2011).

[25] Noelle-Angelique M. Molinari, Ismael R. Ortega-Sanchez, Mark L. Messonnier, William W. Thompson, Pascale M. Wortley, Eric Weintraub, and Carolyn B. Bridges. 2007. The annual impact of seasonal influenza in the US: Measuring disease burden and costs. *Vaccine* 25, 27 (2007), 5086 – 5096.

[26] J. A. Nelder and R. Mead. 1965. A Simplex Method for Function Minimization. *Comput. J.* 7, 4 (01 1965), 308–313. http://dx.doi.org/10.1093/comjnl/7.4.308

[27] Elaine O. Nsoesie, Richard J. Beckman, Sara Shashaani, Kalyani S. Nagaraj, and Madhav V. Marathe. 2013. A Simulation Optimization Approach to Epidemic Forecasting. *PLOS ONE* 8, 6 (06 2013), 1–10. https://doi.org/10.1371/journal.pone.0067164

[28] Elaine O Nsoesie, John S Brownstein, Naren Ramakrishnan, and Madhav V Marathe. 2014. A systematic review of studies on forecasting the dynamics of influenza outbreaks. *Influenza and Other Respiratory Viruses* 8, 3 (05 2014), 309–316. https://doi.org/10.1111/irv.12226

[29] Jon Parker and Joshua M. Epstein. 2011. A Distributed Platform for Global-Scale Agent-Based Models of Disease Transmission. *ACM Trans Model Comput Simul* 22, 1, Article 2 (12 2011), 25 pages. https://doi.org/10.1145/2043635.2043637

[30] Jeffrey Shaman and Alicia Karspeck. 2012. Forecasting seasonal outbreaks of influenza. *Proceedings of the National Academy of Sciences* (2012). https://doi.org/10.1073/pnas.1208772109 arXiv:http://www.pnas.org/content/early/2012/11/21/1208772109.full.pdf

[31] Supplements. [n. d.]. Supplementary file. http://staff.vbi.vt.edu/chenj/pub/DEFSI-supplement.pdf.

[32] Ashleigh R. Tuite, Amy L. Greer, Michael Whelan, Anne-Luise Winter, Brenda Lee, Ping Yan, Jianhong Wu, Seyed Moghadas, David Buckeridge, Babak Pourbohloul, and David N. Fisman. 2010. Estimated epidemiologic parameters and morbidity associated with pandemic H1N1 influenza. *CMAJ* 182, 2 (2010), 131–136. https://doi.org/10.1503/cmaj.091807 arXiv:http://www.cmaj.ca/content/182/2/131.full.pdf

[33] Siva R. Venna, Amirhossein Tavanaei, Raju N. Gottumukkala, Vijay V. Raghavan, Anthony Maida, and Stephen Nichols. 2017. A novel data-driven model for real-time influenza forecasting. *bioRxiv* (2017).

[34] Svitlana Volkova, Ellyn Ayton, Katherine Porterfield, and Courtney D. Corley. 2017. Forecasting influenza-like illness dynamics for military populations using neural networks and social media. *PLOS ONE* 12, 12 (12 2017), 1–22. https://doi.org/10.1371/journal.pone.0188941

[35] Zheng Wang, Prithwish Chakraborty, Sumiko R. Mekaru, John S. Brownstein, Jieping Ye, and Naren Ramakrishnan. 2015. Dynamic poisson autoregression for influenza-like-illness case count prediction. In *KDD*.

[36] WHO. 2018. Seasonal Influenza. http://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal). Accessed May 01, 2018.

[37] Yuexin Wu, Yiming Yang, Hiroshi Nishiura, and Masaya Saitoh. 2018. Deep Learning for Epidemiological Predictions. In *SIGIR*.

[38] Qinneng Xu, Yulia R. Gel, L. Leticia Ramirez Ramirez, Kusha Nezafati, Qingpeng Zhang, and Kwok Leung Tsui. 2017. Forecasting influenza in Hong Kong with Google search queries and statistical model fusion. *PLOS ONE* 12, 5 (05 2017), 1–17. https://doi.org/10.1371/journal.pone.0176690

[39] Shihao Yang, Mauricio Santillana, John S. Brownstein, Josh Gray, Stewart Richardson, and S. C. Kou. 2017. Using electronic health records and Internet search information for accurate influenza forecasting. *BMC Infectious Diseases* 17, 1 (08 May 2017), 332. https://doi.org/10.1186/s12879-017-2424-7

[40] Shihao Yang, Mauricio Santillana, and S. C. Kou. 2015. Accurate estimation of influenza epidemics using Google search data via ARGO. *PNAS* 112, 47 (2015), 14473–14478. https://doi.org/10.1073/pnas.1515373112

arXiv:http://www.pnas.org/content/112/47/14473.full.pdf

[41] Wan Yang, Alicia Karspeck, and Jeffrey Shaman. 2014. Comparison of Filtering Methods for the Modeling and Retrospective Forecasting of Influenza Epidemics. *PLOS Computational Biology* 10, 4 (04 2014), 1–15. https://doi.org/10.1371/journal.pcbi.1003583

[42] Liang Zhao, Jiangzhuo Chen, Feng Chen, Wei Wang, Chang Tien Lu, and Naren Ramakrishnan. 2015. SimNest: Social Media Nested Epidemic Simulation via Online Semi-supervised Deep Learning. *Proceedings of IEEE ICDM* 2015 (11 2015), 639–648. https://doi.org/10.1109/ICDM.2015.39