CausalGNN: Causal-based Graph Neural Networks for Spatio-Temporal Epidemic Forecasting

Lijing Wang,[1,2] **Aniruddha Adiga,** [2] **Jiangzhuo Chen,** [2] **Adam Sadilek,** [3] **Srinivasan Venkatramanan,** [2] **Madhav Marathe** [1,2]

[1] Computer Science, University of Virginia
[2] Biocomplexity Institute and Initiative, University of Virginia
[3] Google Research
lw8bn,aa5dw,chenj,srini,marathe@virginia.edu sadilekadam@google.com

Abstract

Infectious disease forecasting has been a key focus in the recent past owing to the COVID-19 pandemic and has proved to be an important tool in controlling the pandemic. With the advent of reliable spatiotemporal data, graph neural network models have been able to successfully model the interrelation between the cross-region signals to produce quality forecasts, but like most deep-learning models they do not explicitly incorporate the underlying causal mechanisms. In this work, we employ a causal mechanistic model to guide the learning of the graph embeddings and propose a novel learning framework – **Causal**-based **G**raph **N**eural **N**etwork (CausalGNN) that learns spatiotemporal embedding in a latent space where graph input features and epidemiological context are combined via a mutually learning mechanism using graph-based non-linear transformations. We design an attention-based dynamic GNN module to capture spatial and temporal disease dynamics. A causal module is added to the framework to provide epidemiological context for node embedding via ordinary differential equations. Extensive experiments on forecasting daily new cases of COVID-19 at global, US state, and US county levels show that the proposed method outperforms a broad range of baselines. The learned model which incorporates epidemiological context organizes the embedding in an efficient way by keeping the parameter size small leading to robust and accurate forecasting performance across various datasets.

Introduction

Epidemic forecasting is crucial for helping inform policymakers on how to develop effective interventions and marshal limited healthcare resources. In general, modeling and forecasting the spatial and temporal evolution of infectious diseases has been an area of active research over the past couple of decades. Existing methodologies for epidemic forecasting can be broadly categorized into: 1) *Mechanistic causal methods*, including single patch/network-based compartmental models and agent-based models, employ a disease transmission model (e.g. Susceptible-Infectious-Recovered (SIR)) to incorporate the *causation* of disease spread in a population and to capture the underlying dynamics of disease transmission. Such models have been used extensively to study diseases in significant detail, including Ebola (Venkatramanan et al. 2018), influenza (Nsoesie, Mararthe, and Brownstein 2013), and more recently COVID-19 (Hoertel et al. 2020; Anastassopoulou et al. 2020; Giordano et al. 2020; Yamana, Pei, and Shaman 2020). Forecasting is made by calibrating disease parameters using simulation optimization to match observations and then by projecting forward using the most recent calibrated values for future predictions. 2) *Statistical time series methods* such as autoregressive models (e.g., AR, ARMA, and ARIMA) and Kalman filtering have been used for dengue and influenza forecasting in (Yang, Santillana, and Kou 2015; Wang et al. 2015; Kandula, Hsu, and Shaman 2017; Rangarajan, Mody, and Marathe 2019; Shaman and Karspeck 2012), and further used for COVID-19 forecasting in (Harvey and Kattuman 2020; Petropoulos and Makridakis 2020; Ribeiro et al. 2020). 3) *Deep learning methods* have gained increasing prominence in epidemic forecasting, such as Long-Short Term Memory (LSTM) for influenza forecasting (Volkova et al. 2017; Venna et al. 2019; Wu et al. 2018) and COVID-19 forecasting (Chimmula and Zhang 2020; Arora, Kumar, and Panigrahi 2020), and graph neural networks (GNNs) for spatiotemporal epidemic forecasting (Deng et al. 2020; Kapoor et al. 2020; Wang et al. 2020; Ramchandani, Fan, and Mostafavi 2020; Gao et al. 2021). These methods assume statistical properties about the data or employ complex spatiotemporal methodologies to learn patterns in historical data and leverage those patterns for forecasting.

In the context of new emerging epidemics, such as the COVID-19 pandemic, the forecasting problem has been particularly complicated as the surveillance data 1) is sparse due to the lack of historical data; 2) noisy due to reporting bias, testing prevalence etc.; 3) is a resultant of rapidly co-evolving dynamics of individual behavioral adaptations, government policies and disease spread. Given the data challenges, a plethora of models have been explored. One notable forecasting effort within the US is the COVID-19 Forecast Hub, a consortium of over 80 modeling teams initiated by the Centers for Disease Control and Prevention (CDC) in collaboration with academic partners.This effort has been aimed at real-time forecasting wherein multiple classes of statistical and mechanistic models have been employed by individual groups(Ray et al. 2020).Through these efforts and existing published works, we have observed several challenges to the forecasting problem:

- The network-based compartmental models (Balcan et al. 2009; Venkatramanan et al. 2017), compared to single-patched models, explicitly account for the connectivity among patches. Thus, it is more promising in capturing the relation between the model parameters and spatiotemporal data. However, calibrating such models, especially at the high geographical resolution, is challenging given the need to capture time-varying inter- and intra-regional effects. For example, for the United States with 3000+ counties and W weeks of data, there are technically over $3000 \times W$ entries in the spatiotemporal transmissibility matrix to be calibrated, making traditional Bayesian techniques computationally intensive and susceptible to overfitting due to the limited training data size.
- Deep learning models especially GNN-based models usually require a sufficiently large quality dataset to train the large number of model parameters to avoid overfitting. Existing spatiotemporal forecasting models such as for traffic forecasting (Li et al. 2017; Lai et al. 2018; Yu, Yin, and Zhu 2017; Wu et al. 2019, 2020; Bai et al. 2020), tend to overfit in epidemic data because epidemic data is sparser and noisier than traffic data. Existing spatiotemporal epidemic forecasting models (Wu et al. 2018; Deng et al. 2020) whose parameter size increases with graph node size failed to forecast over a large number of regions. Reducing the complexity of such models is crucial for accurate forecasting.
- Prior works in physics and biology (Karpatne et al. 2017) have shown the evidence that incorporating domain knowledge into data-driven models can help improve spatiotemporal forecasting algorithms. However, for epidemic forecasting, apart from a few models (Wu et al. 2018; Deng et al. 2020), existing deep learning models rarely consider explicit incorporation of epidemiological context[1]. Such models are prone to be overfitting leading to failures in long-term forecasting, especially when the data is noisy and sparse such as COVID-19 surveillance data at the US county level. Recent work (Gao et al. 2021) leverages causal features generated by the ordinary differential equations (ODE) to regularize model predictions for GNN learning but has not considered these features in the graph embedding process.

To address the above challenges, we propose **Causal**-based **G**raph **N**eural **N**etwork (CausalGNN). The CausalGNN attempts to capture the spatial and temporal dynamics via a well designed GNN module and uses a causal module to mutually provide and embed causal features to get epidemiological context. Causal constraints are also added to further improve model forecasting performance. The major contributions are summarized below:

- We propose a novel spatiotemporal learning framework that learns a latent space to combine the spatiotemporal and causal embeddings using graph-based non-linear transformations. We present a jointly learning process for incorporating epidemiological context in GNN learning.
- We design an attention-based dynamic GNN module to embed spatial and temporal signals from disease dynamics. The parameter size in our design is agnostic to the number of regions thus leading to a robust forecasting performance on datasets of varying region numbers.
- We incorporate a causal module of single-patched compartmental models into the framework to provide epidemiological context. Compared to traditional network-based compartmental models, in our framework, the patches are connected via a learned GNN. The calibration is done through GNN training, which is computationally efficient. The causal outputs are embedded as graph node features and used to regularize GNN forecasts for causal-based forecasting, leading to better forecasting performance.
- In order to allow for interaction between the causal and GNN modules, we design a causal encoder to encode causal features as node embedding to propagate over the graph and a causal decoder to infer mechanistic model parameters from latent space at each time step. To the best of our knowledge, we are the first to propose this iterative feedback mechanism that benefits from the learning in both modules.
- We evaluate the proposed framework for forecasting daily new confirmed cases of COVID-19 at global, US state, and US county levels. Our model outperforms a broad range of baselines with up to 7% improvement of performance. Through an ablation study, we demonstrate the effectiveness of causal module in improving model performance.

Methods

Problem Formulation

We assume N regions in total and define a dynamic graph on the N regions as $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathcal{T})$, where $\mathcal{V}$ is the set of N nodes, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges, and $\mathcal{T}$ is the set of T time points. At each time step t, the graph $\mathcal{G}$ is associated with a feature matrix $\mathbf{C}_t \in \mathbb{R}^{N \times C}$ where C is the feature number and the graph nodes are connected via an adjacency matrix $\mathbf{A}_t \in \mathbb{R}^{N \times N}$. Given a graph $\mathcal{G}$ and its historical $K \leq T$ steps of feature and adjacency matrices, the objective is to predict an epidemiological target at future time $T + h$ for N regions where h denotes the horizon time. The important notations are described in Appendix Table 1.

Framework

The proposed framework (shown in Figure 1) consists of two major modules: 1) an *attention-based dynamic GNN (ADGNN) module* to capture the spatial and temporal disease dynamics via graph-based neural networks; 2) a *causal module* to provide epidemiological context for GNN learning via ordinary differential equations. The working of the model is as follows (the left part of Figure 1): at each time

[1] In our paper, epidemiological context refers to all the parameters and features corresponding to the disease models, such as S,I,R counts and disease specific parameters, which can provide context specific information for model learning.

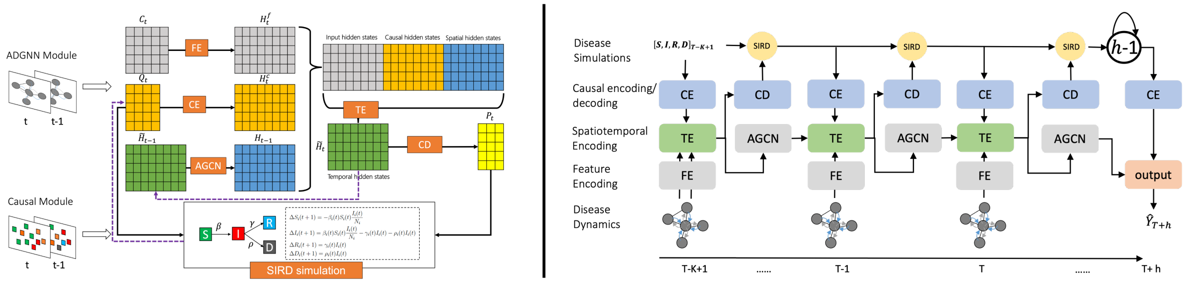


Figure 1: Framework of CausalGNN which consists of a causal module and an attention-based dynamic GNN module.

step t, the graph $\mathcal{G}$ is embedded via nonlinear transformations including feature encoding (FE), causal encoding (CE), dynamic graph encoding (AGCN), temporal encoding (TE), causal decoding (CD), and susceptible(S)-infected(I)-recovered(R)-deceased(D) (SIRD) computation. In causal module, we run a single-patched SIRD model for each region. In ADGNN module, the causal feature matrix $\mathbf{Q}_t$ is encoded as a causal hidden matrix $\mathbf{H}_t^c$ via the CE layer. The input feature matrix $\mathbf{C}_t$ is encoded as a hidden matrix $\mathbf{H}_t^f$ via the FE layer. The current TE layer combines $\mathbf{H}_t^f, \mathbf{H}_t^c, \mathbf{H}_{t-1}$ to generate the temporal hidden matrix $\tilde{\mathbf{H}}_t$ which is then fed into the CD layer to infer causal parameters $\mathbf{P}_t$ and fed into the AGCN layer to generate the spatial hidden matrix $\mathbf{H}_t$ ($\mathbf{H}_{T-K} = \mathbf{H}_{T-K+1}^f$) for the next round of computation. $\mathbf{Q}_{t+1}$ ($\mathbf{Q}_{T-K+1}$ is given as the input) is updated via Equation 1 from $\mathbf{Q}_t$ and $\mathbf{P}_t$. We save $\mathbf{Q}_{t+1}$ for the next computation This iteration is repeated by K times. At time $T+1$, we run SIRD computation for $h-1$ steps further with the most recent disease model parameters $\mathbf{P}_T$ to get causal features $\mathbf{Q}_{T+h}$ which is then fed into a CE layer. The hidden matrices from the last AGCN layer and the last CE layer are combined and fed into an output layer to get the final prediction. The pseudocode of the model training process is described in Appendix. We will elaborate details of each step in the following sections.

Causal Modeling We import epidemiological context by incorporating causal-based differential equations into a deep learning framework. In this paper, we focus on COVID-19 forecasting. Based on the availability of surveillance data (i.e., daily confirmed, death, and recovered counts), we choose a single-patched compartmental SIRD model (Loli Piccolomini and Zama 2020) to simulate the COVID-19 spread in each region. Other models such as SIR can also work. We discuss this in Appendix via sensitivity analysis. Consider a population of N_i individuals in patch i (i.e., region i), each of whom can be in one of the following states: S, I, R, D. Compartmental models operate under a homogeneous mixing assumption, i.e., every individual can directly infect any other individual. We assume that individuals who become recovered do not get infected again. The dynamics of epidemic spread in patch i at time t are described by the following equations:

$$\begin{aligned}
\Delta S_i(t+1) &= -\beta_i(t)S_i(t)\frac{I_i(t)}{N_i} \\
\Delta I_i(t+1) &= \beta_i(t)S_i(t)\frac{I_i(t)}{N_i} - \gamma_i(t)I_i(t) - \rho_i(t)I_i(t) \\
\Delta R_i(t+1) &= \gamma_i(t)I_i(t) \\
\Delta D_i(t+1) &= \rho_i(t)I_i(t)
\end{aligned} \tag{1}$$

where $\mathbf{q}_{i,t} : S_i(t), I_i(t), R_i(t), D_i(t)$ denote the cumulative number of individuals in each of the states and $S_i(t) + I_i(t) + R_i(t) + D_i(t) = N_i$; $\mathbf{p}_{i,t} : \beta_i(t), \gamma_i(t), \rho_i(t)$ denote the transmissibility, the recovery rate, and mortality rate, respectively; Δ means the newly added number of individuals in each state. In our framework, let $\mathbf{Q}_t = (\mathbf{q}_{i,t}) \in \mathbb{R}^{N\times 4}$ denote the causal feature matrix of N regions at time t, then $\mathbf{Q}_{t+1}$ is updated as $\mathbf{Q}_t + \Delta\mathbf{Q}_{t+1}$ where the initial values of $\mathbf{Q}_{T-K+1}$ are given as the input; $\mathbf{P}_t = (\mathbf{p}_{i,t}) \in \mathbb{R}^{N\times 3}$ denotes the causal parameter matrix of N regions at time t which is inferred by a neural network.

Causal Encoding A causal encoder (CE) is designed to encode causal features as node embedding. At time t, it works as:

$$\mathbf{H}_t^c = \tanh\left(\mathbf{Q}_t\mathbf{W}_e^{(t)} + \mathbf{b}_e^{(t)}\right) \in \mathbb{R}^{N\times F_c}, \tag{2}$$

where $\mathbf{W}_e^{(t)} \in \mathbb{R}^{4\times F_c}$ and $\mathbf{b}_e^{(t)} \in \mathbb{R}^{F_c}$ are model parameters.

Feature Encoding Let $\mathbf{H}_t^f \in \mathbb{R}^{N\times F_f}$ represent the matrix of hidden states of node features $\mathbf{C}_t$ for N nodes by a feature encoder (FE):

$$\mathbf{H}_t^f = \sigma\left(\mathbf{C}_t\mathbf{W}_f^{(t)} + \mathbf{b}_f^{(t)}\right) \in \mathbb{R}^{N\times F_f}, \tag{3}$$

where $\mathbf{W}_f^{(t)} \in \mathbb{R}^{C\times F_f}$, $\mathbf{b}_f^{(t)} \in \mathbb{R}^{F_f}$ are model parameters and σ is sigmoid activation function.

Dynamic Graph Encoding We leverage GCN (Kipf and Welling 2017) to generate node embedding based on local network neighborhoods through message passing. The neighborhoods are defined using an attention matrix. A traditional GCN model consists of multiple layers for a single graph convolution. In our problem, the node features and graph structure vary across time, hence we implemented a dynamic attention-based GCN (AGCN) that an AGCN layer corresponds to a time step to learn spatial features. The number of AGCN layers is the number of time points in the input sequence K and they share a common parameter set. This

AGCN architecture allows our model to recurrently propagate forward the spatial and temporal features with a small parameter size.

Let $\mathbf{H}_t \in \mathbb{R}^{N\times F}$ denotes the matrix of hidden states from the AGCN layer at time t, which maps from $\tilde{\mathbf{H}}_t$ as:

$$\mathbf{H}_t = g\big(\mathbf{A}_t\tilde{\mathbf{H}}_t\mathbf{W}^{(t)} + \mathbf{b}^{(t)}\big) \in \mathbb{R}^{N\times F}, \tag{4}$$

where $\mathbf{W}^{(t)} \in \mathbb{R}^{F\times F}$, $\mathbf{b}^{(t)} \in \mathbb{R}^{F}$ are model parameters. $\mathbf{A}_t$ is an attention matrix which is defined below. $\tilde{\mathbf{H}}_t \in \mathbb{R}^{N\times F}$ is the output from the temporal encoder (TE) layer (will be described in the following section). g is rectified linear units (ReLU) (Nair and Hinton 2010). The spatial embedding $\mathbf{H}_{T-K} = \mathbf{H}^f_{T-K+1}$.

In real world scenarios, the disease dynamics change and co-evolve at each time step thus the traditional geographical adjacency matrix failed to reveal the true connectivity. Recent works (Kapoor et al. 2020; Wang et al. 2020) use aggregate mobility data to understand COVID-19 dynamics. However, they usually require adequate data sources to achieve decent performance in epidemic forecasting which are usually not available for public usages. Furthermore, as mentioned in (Wang et al. 2020), these data may not be representative of the population as whole, and their representativeness may vary by region. We want the model to learn an adaptive relationship between two nodes. Thus, we define an asymmetric attention matrix to reflect the dynamic connectivity among regions at each time step, denoted as $\mathbf{A}_t = (a_{ij,t}) \in \mathbb{R}^{N\times N}$ where $a_{ij,t}$ represents the impact of node j on node i. It is computed from $\tilde{\mathbf{H}}_t$ as:

$$a_{ij,t} = \mathbf{v}^T g(\mathbf{W}_{sc}\mathbf{h}_{i,t} + \mathbf{W}_{tg}\mathbf{h}_{j,t} + \mathbf{b}_s) + b_s \in \mathbb{R}, \tag{5}$$

where $\mathbf{h}_{i,t}, \mathbf{h}_{j,t} \in \mathbb{R}^F$ are the transpose of the ith and jth rows in $\tilde{\mathbf{H}}_t$; $\mathbf{W}_{sc}, \mathbf{W}_{tg} \in \mathbb{R}^{F_s\times F}$, $\mathbf{b}_s \in \mathbb{R}^{F_s}$, and $b_s \in \mathbb{R}$ are model parameters; g is ReLU that is applied elementwise. We use softmax function to normalize each row in $\mathbf{A}_t$.

Temporal Encoding Inspired by recurrent neural networks, to consider temporal features in the graph, at each time step, we employ a temporal encoder (TE) layer to reencode the hidden representatives $\mathbf{H}^f_t, \mathbf{H}^c_t, \mathbf{H}_{t-1}$ from FE and CE at the current time t, and AGCN at the previous time $t-1$. $\tilde{\mathbf{H}}_t$ in Equation 4 is computed as:

$$\begin{aligned}
\dot{\mathbf{H}}^f_t &= \mathbf{H}^f_t\mathbf{W}^{(t)}_a + \mathbf{b}^{(t)}_a,\\
\dot{\mathbf{H}}^c_t &= \mathbf{H}^c_t\mathbf{W}^{(t)}_b + \mathbf{b}^{(t)}_b,\\
\dot{\mathbf{H}}_{t-1} &= \mathbf{H}_{t-1}\mathbf{W}^{(t)}_c + \mathbf{b}^{(t)}_c,\\
\tilde{\mathbf{H}}_t &= \tanh\big(\big[\dot{\mathbf{H}}^f_t \| \dot{\mathbf{H}}^c_t \| \dot{\mathbf{H}}_{t-1}\big]\big) \in \mathbb{R}^{N\times F},
\end{aligned} \tag{6}$$

where $\mathbf{W}^{(t)}_a \in \mathbb{R}^{F_f\times a}$, $\mathbf{W}^{(t)}_b \in \mathbb{R}^{F_c\times b}$, $\mathbf{W}^{(t)}_c \in \mathbb{R}^{F\times c}$, $\mathbf{b}^{(t)}_a \in \mathbb{R}^a$, $\mathbf{b}^{(t)}_b \in \mathbb{R}^b$, and $\mathbf{b}^{(t)}_c \in \mathbb{R}^c$ are model parameters, and $a+b+c = F$. $\|$ represents concatenate operation. The TE module can be replaced by existing RNN modules such as RNN, GRU, or LSTM. This will be discussed in sensitivity analysis in Appendix.

Causal Decoding The disease model parameter matrix $\mathbf{P}_t$ is inferred dynamically from $\tilde{\mathbf{H}}_t$ via a causal decoder (DE):

$$\mathbf{P}_t = \sigma\big(\tilde{\mathbf{H}}_t\mathbf{W}^{(t)}_d + \mathbf{b}^{(t)}_d\big) \in \mathbb{R}^{N\times 3}, \tag{7}$$

where $\mathbf{W}^{(t)}_d \in \mathbb{R}^{F\times 3}$ and $\mathbf{b}^{(t)}_d \in \mathbb{R}^3$ are model parameters and σ is the sigmoid activation function.

Output Layer As described in the framework overview, the causal parameters $\mathbf{P}_T$ are used to run the SIRD model $h-1$ steps further to generate causal predictions $\mathbf{Q}_{T+h}$ and will then be fed into a CE layer to generate $\mathbf{H}^c_{T+h}$. We concatenate $\mathbf{H}^c_{T+h}$ and the output of the last AGCN layer $\mathbf{H}_T$ and feed them to an output layer for final prediction:

$$\hat{\mathbf{Y}} = \phi\Big(\Big[\mathbf{H}_T \parallel \mathbf{H}^c_{T+h}\Big]\mathbf{W}_o + b_o\Big) \in \mathbb{R}^N, \tag{8}$$

where $\mathbf{W}_o \in \mathbb{R}^{(F_c+F)}$, $b_o \in \mathbb{R}$ are model parameters, ϕ is an identity function, and $\hat{\mathbf{Y}}$ denotes the predicted target for N regions at time $T+h$.

Optimization

We consider causal loss together with ADGNN prediction loss in the loss function and then optimize a ℓ_1-norm loss via gradient descent:

$$\mathcal{L}(\Theta) = \|\mathbf{Y} - \hat{\mathbf{Y}}\| + \sum_{t=T-K+2}^{T+h} \|\mathbf{Y}^c_t - \hat{\mathbf{Y}}^c_t\|, \tag{9}$$

where $\hat{\mathbf{Y}}^c_t$ denotes the causal prediction of the target from SIRD simulations for N regions at time t; $\mathbf{Y}$ and $\mathbf{Y}^c$ represent the corresponding ground truth values. We do not distinguish regions in calculating loss. In our framework, given an epidemiological target, we have two predictions from causal module and ADGNN module respectively. We use the prediction from the ADGNN module as our final prediction as it embeds hidden information from both modules via the output layer.

Model Complexity

The number of parameters of the proposed model is $O(F\times(F{+}F_s{+}c){+}F_c{\times}b{+}F_f{\times}(C{+}a))$. It is agnostic to the number of regions in the dataset. In our setting, C, F, F_c, F_f, F_s are limited to small numbers. Thus, our model can capture spatiotemporal patterns of disease transmissions in an efficient way. We will provide more detailed analysis in the experiment section.

Experiments

Data We use three kinds of data for our experiments. Their data sources are elaborated in Appendix. **COVID-19 datasets**: It contains daily cumulative confirmed, death, and recovered counts at global, US state and county levels, as well as regions' latitude and longitude, from May 3, 2020 to April 23, 2021. We select countries with population size of more than 8.7 million and US counties with more than 3000 confirmed cases by March 20, 2021 to ensure the data source accuracy. Finally, we include 93 countries, 52 states, and 1351 counties. Their statistics are shown in Table 1. **Geographical adjacency datasets**: It contains country adjacency, US state adjacency, and US county adjacency information. **Population datasets**: It contains country population (2020), US state and county population (2019) information. It is used to calculate $S(T-K+1)$.

Table 1: Dataset statistics: min, max, mean, and standard deviation (std) of patient counts; dataset size means number of locations multiplied by # of days.

Dataset	Size	Min	Max	Mean	std
Globe	93×355	0	823225	3988	15381
US-State	52×355	0	62168	1670	3192
US-County	1351×355	0	34497	59	238

Metrics The metrics used to evaluate the forecasting performance are: *mean absolute error (MAE)* which is a measure of absolute difference between two variables, and *mean absolute percentage error (MAPE)* which measures the size of the error between two variables in percentage terms. Both MAE and MAPE range in $[0, +\infty]$ and smaller values are better. The detailed definition and calculation equations are shown in Appendix Equation (1) and (2).

Baselines To serve as baselines, we implemented a broad range of classic and state-of-the-art forecasting models. A detailed description is shown in Appendix.

- *Mechanistic causal models*: **SIR** is a single patch SIR compartmental model. **PatchSEIR** (Venkatramanan et al. 2017) is a network-based SEIR compartmental model for infectious disease forecasting.
- *Statistical models*: **Autoregressive (AR)** and **Autoregressive Moving Average (ARMA)** (Contreras et al. 2003).
- *Classic deep learning models*: **Recurrent Neural Network (RNN)** (Werbos 1990), **Gated Recurrent Unit (GRU)** (Cho et al. 2014), and **Long-Short Term Memory (LSTM)** (Hochreiter and Schmidhuber 1997).
- *Spatio-temporal deep learning models*: **DCRNN** (Li et al. 2017), **CNNRNN-Res** (Wu et al. 2018), and **LSTNet** (Lai et al. 2018). They combine convolutional neural networks (CNNs) and RNNs to extract spatial and temporal patterns for learning time series trends.
- *Graph-based models*: **STGCN** (Yu, Yin, and Zhu 2017), **Cola-GNN** (Deng et al. 2020), and **STAN** (Gao et al. 2021). These models use GNNs to combine graph structures and time series features to capture a dynamic propagation process.

Settings and Implementation Details In the graph $\mathcal{G}$, each node's features include dynamic features (i.e., daily new-confirmed cases, recovered cases, and deaths) and static features (i.e., population density, latitude, and longitude). In our model, the hidden dimensions F_c,F_f,F are set as 32, F_s is set as 16 ($\frac{F}{2}$), and $a = 12, b = 10, c = 10$ in Equation 6. All the parameters are initialized with Glorot initialization. We set batch size as 32, epoch number as 1000. We use Adam optimizer with default settings, and early stopping with patience of 100 epochs for all model training. For all models, the historical window $K = 28$. Unless otherwise specified, all baselines have parameters set in accordance with the original paper. The collected COVID-19 datasets are split into training-validation datasets (from May 3, 2020, to March 20, 2021) and testing datasets (from March 21, 2021, to April 23, 2021). For each targeted data point in a testing dataset, we make 7, 14, 21, and 28 days ahead forecasting of the data point. All results are an average of 5 randomized trials. The random seeds for reproducing the results are 42, 52, 62, 72, and 82. We show experiment results with their means and 95% confidence intervals. All programs are implemented using Python 3.7.4 and PyTorch 1.4.0 with CUDA 10.1 in a Simple Linux Utility for Resource Management (SLURM) system with K80, P100, V100, and RTX2080 NVIDIA GPU devices that serve in random.

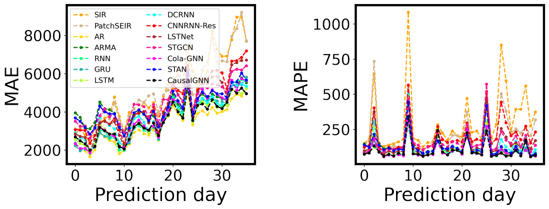
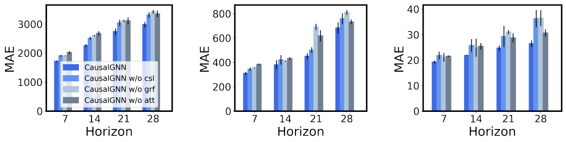


Figure 2: Performance of MAE and MAPE computed across all regions at the Global level across various forecast days.



(a) Globe (b) US-State (c) US-County

Figure 3: Ablation analysis on major components of the proposed model.

Forecasting Performance We evaluate our method and all baselines on forecasting COVID-19 daily new confirmed case count at global, US state, and US county levels. Table 2 shows the model performance in terms of MAE and MAPE.

We observe that CausalGNN performs consistently better than the baselines across multiple scales and with increasing horizons. Compared among spatiotemporal forecasting models, epidemic forecasting models (e.g., Cola-GNN, STAN, and CausalGNN) outperform models proposed for traffic forecasting (e.g., DCRNN, LSTNet, and STGCN). A possible reason is that data sampling for epidemic data is different with traffic data. For instance, traffic sensors transmit data at 5-minute intervals while COVID-19 data collection shows a larger granularity (i.e., days) with a delay. Traffic forecasting models tend to overfit in epidemic data.

CausalGNN performs better than STAN in most cases because it not only adds a causal-based regularizer in the loss function (like STAN did) but also mutually encodes causal features into graph learning, which provides epidemiolog-

Table 2: MAE and MAPE performance of different methods on the three datasets with horizon= 7, 14, 21, 28. Mean and 95% confidence interval of 5 runs are shown. Bold face indicates the best result of each column and underlined is the second-best. Improvement values are the improved ratio made by CausalGNN when compared with the second-best method.

	Globe				**US-State**				**US-County**			
MAE(↓)	7	14	21	28	7	14	21	28	7	14	21	28
SIR	4777±819	4880±615	5090±1125	5182±282	677±31	738±58	831±51	854±60	38.8±3.3	44.2±2.7	51.3±2.8	58.5±2.0
PatchSEIR	4419±500	4562±601	4737±349	5167±298	633±78	687±78	757±37	876±77	73.5±5.4	84.4±6.7	100.8±14.6	110.6±6.4
AR	2298±10	3024±7	3619±17	4258±246	377±1	580±3	683±11	758±28	24.8±0.1	33.6±0.3	34.2±0.3	35.6±0.4
ARMA	2254±13	2987±14	3596±23	4239±60	379±3	583±3	686±9	750±17	23.8±0.1	27.0±0.1	33.0±3.8	35.9±0.7
RNN	2395±44	2871±28	3328±27	3596±178	369±13	525±38	660±191	745±181	21.5±1.0	25.0±2.4	35.9±5.0	36.1±3.7
GRU	2189±48	2916±30	3379±37	3620±130	385±27	504±80	660±100	833±174	31.6±4.1	30.1±11.5	35.1±8.8	38.1±11.1
LSTM	1911±16	2585±11	3050±21	3598±140	344±9	421±24	552±161	748±134	23.4±1.0	**23.7±0.3**	33.0±4.7	37.6±1.3
DCRNN	2287±189	2892±137	3369±71	3804±177	393±20	470±26	657±172	702±165	22.4±1.3	25.3±1.0	31.2±6.7	36.3±6.4
CNNRNN-Res	4143±649	4526±572	4467±437	4479±390	642±31	658±41	732±94	856±148	29.8±1.3	31.0±1.1	33.4±1.1	36.0±2.8
LSTNet	2693±91	3535±125	3909±209	4285±155	443±19	597±34	744±73	815±53	24.5±0.7	28.0±1.7	31.5±1.0	33.2±1.6
STGCN	4750±796	4325±357	4669±202	4494±162	580±19	630±19	699±95	793±60	23.7±1.1	26.5±2.7	30.9±2.7	32.7±5.4
Cola-GNN	2314±231	3012±682	3225±263	3755±175	384±30	497±19	613±124	810±343	22.5±1.4	37.7±19.1	34.5±7.5	37.5±9.3
STAN	**1851±172**	2628±144	3163±138	3574±142	350±16	428±27	512±80	**622±122**	22.2±0.7	25.3±1.9	28.5±1.7	31.2±4.6
CausalGNN	1911±192	**2502±144**	**3041±211**	**3310±58**	**339±13**	**416±16**	**498±68**	640±51	**21.3±0.4**	24.1±0.1	**26.8±0.7**	**29.0±0.7**
MAPE(↓)	7	14	21	28	7	14	21	28	7	14	21	28
SIR	577±46	335±61	285±10	298±19	141±27	147±34	139±26	146±31	233.7±9.3	260.9±11.5	217.8±9.3	234.0±2.3
PatchSEIR	342±26	268±17	225±13	228±21	152±26	143±24	153±31	180±23	472.9±15.0	479.4±18.2	546.3±10.0	642.1±31.5
AR	**108±0.3**	109±0.7	110±0.7	130±13.1	93±1.0	114±1.7	150±8.4	178±18.6	79.7±0.1	79.4±0.9	81.8±0.9	84.7±0.9
ARMA	109±0.3	110±2.8	110±1.3	127±9.7	91±2.4	111±2.0	146±12.0	175±25.5	75.6±0.1	89.4±0.6	92.2±13.4	86.8±0.7
RNN	131±13	98±10	108±13	112±3	86±9	130±34	158±61	192±85	62.9±13.4	87.0±6.3	98.2±9.8	137.2±31.9
GRU	124±10	115±13	99±13	113±11	95±14	98±40	118±28	210±97	64.3±4.6	79.9±12.0	118.6±20.0	134.3±42.0
LSTM	126±4	104±1	95±4	119±16	84±3	89±7	122±48	182±54	62.5±6.0	**62.1±3.7**	108.2±27.2	134.1±33.3
DCRNN	135±14	120±10	122±7	132±10	95±3	107±3	132±37	137±38	63.5±6.1	71.0±4.6	85.4±6.8	96.1±25.8
CNNRNN-Res	230±41	218±20	206±48	204±20	108±11	136±26	150±34	167±11	90.8±12.2	91.9±8.4	97.3±13.8	103.7±19.2
LSTNet	131±4	114±17	129±14	147±19	86±3	110±9	137±11	171±11	72.7±3.0	81.3±8.3	86.5±4.8	107.9±9.7
STGCN	210±13	173±9	168±24	163±16	129±11	146±21	155±37	180±31	62.9±4.4	71.4±6.3	83.8±10.1	85.3±12.1
Cola-GNN	125±31	119±53	96±10	100±11	95±16	119±27	122±14	218±144	**56.5±5.7**	101.1±10.4	110.2±12.7	123.4±22.5
STAN	126±5	**96±7**	92±10	109±11	86±1	96±1	108±1	**109±3**	75.4±1.9	82.8±0.6	95.4±1.0	104.9±2.4
CausalGNN	122±4	100±6	**90±9**	**97±17**	**81±4**	**87±4**	**105±11**	140±10	62.0±3.7	64.7±1.0	**77.3±4.4**	**79.0±2.8**

ical context recurrently for future forecasting. Compared with GNN-based models, Cola-GNN performs the worst on the US-County dataset. A possible reason is that its model size increases linearly with the squared number of regions (N^2) leading to overfitting to the dataset of 1351 regions.

SIR and PatchSEIR perform worse than data-driven methods, especially for long-term forecasting. PatchSEIR performs worse than SIR at county level. As we mentioned in the introduction section, SIR does not consider the spatial connectivity thus fail to capture spatial disease transmission dynamics. PatchSEIR leverages a gravity model-generated network but may not represent real world mobility activities. Further, calibrating it is prone to overfitting on the US-County dataset due to the large number of counties. In our framework, the patches are connected via a learned GNN that allows the spatial and temporal disease dynamics to exchange information in a latent space. The results demonstrate the practical value of our design.

Compared with GNN-based models including STGCN, Cola-GNN, STAN, and CausalGNN, the vanilla RNN, GRU, LSTM models perform well in horizon=7,14. However, as the horizon increases their advantages have diminished. This indicates the importance of capturing spatial disease transmission patterns in the input data for long term forecasting. In most cases, the classic statistical methods (AR, ARMA) show a poorer performance than the classic RNNs (RNN, GRU, LSTM). This implies the importance of modeling non-linear patterns for achieving good forecasting performance.

Figure 2 shows the model performance of MAE and MAPE computed across all regions at the Global level at various forecast days. We observe that the model performance varies across the days but our model performs the best in most of the days. We also observe that the MAE values increase by days. The trend in MAE values coincides with the trend in the number of global daily new confirmed cases, which increases day by day from March 21, 2021, to April 23, 2021. Similar observations are obtained at US-State and US-County levels. However, the MAPE results show a flat trend with interval spikes across days. The spikes of MAPE are caused by the noise in the testing datasets (variability in reporting across day of a week). These observations indicate that all models are implemented in a fair manner and perform stably across days. We obtain similar observations from performance at the US-state and US-county levels, which are presented at the Appendix.

Ablation Study To explore the effect of the causal module and graph structure in our model, we conduct an ablation analysis on the three datasets.

- **CausalGNN w/o csl**: Remove the SIRD causal encoder and decoder layers from the proposed model, and remove the second term from the loss function in Equation 9. We call the removed components as CSL.
- **CausalGNN w/o grf**: Remove the AGCN layers from the

model architecture. This means remove the graph structure called GRF.

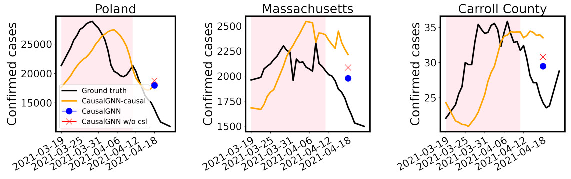
- **CausalGNN w/o att**: Replace the attention matrix with geographical adjacency matrix, which means remove the attention mechanism called ATT.

We present the comparisons of forecasting performance in terms of MAE for the above described model configurations in Figure 3. Each comparison group (of the same metric, dataset and horizon) involves four models: CausalGNN, CausalGNN w/o csl, CausalGNN w/o grf, and CausalGNN w/o att. Within a group, a model with a larger MAE value than CausalGNN indicates a more important role of the missing component in that model. The forecasting performance in terms of MAPE shows similar observations, thus is shown in Appendix due to the page limit.

Major observations and discussion: CausalGNN always performs the best among the four models on different datasets and horizons. This implies that all three components play important roles in improving our model performance. Specifically, in short-term forecasting (horizon=7,14), CausalGNN w/o att performs the worst. It indicates that an adaptive adjacency matrix is crucial in capturing near future dynamics. In long-term forecasting (horizon=21,28), CausalGNN w/o grf performs the worst on three datasets. This indicates that GRF plays the most important role in improving long-term forecasting performance. It complies with the fact that incorporating cross-spatial signals is crucial for a good epidemic forecasting model. Also, GRF's importance increases with increasing spatial resolution which is intuitive as the spatial interdependence is higher at state and county level. The results also show that adding the CSL to the framework can lead to a performance improvement. This demonstrates the effectiveness of the CSL in improving epidemic forecasting performance.

Epidemiological Context In Figure 4, we present examples of the causal module impact by comparing 7 days ahead forecasts of 2021-04-18 by CausalGNN (blue dots) and CausalGNN w/o csl (red crosses) in Poland, Massachusetts, and Carroll County. Both solid lines and dots are smoothed and the shaded area is the input window. We can observe that 1) CausalGNN makes better forecasts than CausalGNN w/o csl (i.e., the blue dots are closer to the black curves than the red crosses on the forecasting day). This means that the causal module proposed in our model can help in improving the model performance; 2) the causal module in CausalGNN can generate meaningful curves (orange curves) compared with the ground truth curves (black curves). This indicates that CausalGNN can reveal mechanistic causal process by producing meaningful causal parameters which can provide meaningful epidemiological context for GNN learning. Limitations are discussed in Appendix.

Model Complexity The number of parameters of our model is agnostic to the number of regions N, as well as RNN, GRU, and LSTM models. The parameter sizes of AR, ARMA, and LSTNet increase linearly with N while those of CNNRNN-Res and Cola-GNN increase linearly with N^2. The parameter sizes of SIR and PatchSEIR are linearly increasing with $N \times T$. We compare the model parameter size of all methods in Table 3. The results show that compared with the other GNN-based models, CausalGNN keeps a relatively small parameter size even when the number of regions increases. This demonstrates that our method can achieve robust performance across different datasets.



(a) Poland (b) Massachusetts (c) Carroll County

Figure 4: Examples of causal module impact.

Table 3: Model parameter size comparison on the US-State, Globe, and US-county datasets. κ denotes the real parameter size on US-State level. We show real parameter size for US-State level and relative values for US-County and Global level.

Methods	US-State	Globe (κ)	US-County (κ)
SIR	16.6K	1.79	2.60
PatchSEIR	16.6K	1.79	2.60
AR	1.5K	1.79	25.98
ARMA	2.9K	1.79	25.98
RNN	0.5K	1.00	1.00
GRU	1.4K	1.00	1.00
LSTM	1.9K	1.00	1.00
DCRNN	21	1.00	1.00
CNNRNNRes	9.7K	2.04	201.98
LSTNet	13.3K	1.61	20.48
STGCN	14.6K	1.01	1.35
ColaGNN	5.7K	2.05	323.51
STAN	8K	0.96	0.96
CausalGNN	1.5K	0.97	0.97

Conclusion

This paper introduces CausalGNN which is a GNN-based model combining with causal computations for spatiotemporal epidemic forecasting. CausalGNN is well-designed by keeping a small number of parameters and considering epidemiological context via a mutually learning mechanism, leading to better spatiotemporal forecasting performance compared to baselines. Future work may include: 1) multi-task learning, such as confirmed and death counts; 2) exploring counterfactual forecasting via the causal module; 3) conducting a deeper analysis on the learned model for explainability.

Acknowledgement

The authors would like to thank members of the Biocomplexity COVID-19 Response Team and the Network Systems Science and Advanced Computing (NSSAC) Division for their thoughtful comments and suggestions related to epidemic modeling and response support. We thank members of the Biocomplexity Institute and Initiative, University of Virginia, for useful discussion and suggestions. This work was partially supported by National Institutes of Health (NIH) Grant 1R01GM109718, NSF BIG DATA Grant IIS-1633028, NSF Grant No.: OAC-1916805, NSF Expeditions in Computing Grant CCF-1918656, CCF-1917819, NSF RAPID CNS-2028004, NSF RAPID OAC-2027541, US Centers for Disease Control and Prevention 75D30119C05935, a grant from Google, University of Virginia Strategic Investment Fund award number SIF160, Defense Threat Reduction Agency (DTRA) under Contract No. HDTRA1-19-D-0007, and Virginia Dept of Health Grant VDH-21-501-0141. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding agencies.

References

Anastassopoulou, C.; Russo, L.; Tsakris, A.; and Siettos, C. 2020. Data-based analysis, modelling and forecasting of the COVID-19 outbreak. *PloS one*, 15(3): e0230405.

Arora, P.; Kumar, H.; and Panigrahi, B. K. 2020. Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India. *Chaos, Solitons & Fractals*, 110017.

Bai, L.; Yao, L.; Li, C.; Wang, X.; and Wang, C. 2020. Adaptive graph convolutional recurrent network for traffic forecasting. *arXiv preprint arXiv:2007.02842*.

Balcan, D.; Colizza, V.; Gonccalves, B.; Hu, H.; Ramasco, J. J.; and Vespignani, A. 2009. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences*, 106(51): 21484–21489.

Chimmula, V. K. R.; and Zhang, L. 2020. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos, Solitons & Fractals*, 109864.

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*.

Contreras, J.; Espinola, R.; Nogales, F. J.; and Conejo, A. J. 2003. ARIMA models to predict next-day electricity prices. *IEEE transactions on power systems*, 18(3): 1014–1020.

Deng, S.; Wang, S.; Rangwala, H.; Wang, L.; and Ning, Y. 2020. Cola-GNN: Cross-location Attention based Graph Neural Networks for Long-term ILI Prediction. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 245–254.

Gao, J.; Sharma, R.; Qian, C.; Glass, L. M.; Spaeder, J.; Romberg, J.; Sun, J.; and Xiao, C. 2021. STAN: spatio-temporal attention network for pandemic prediction using real-world evidence. *Journal of the American Medical Informatics Association*, 28(4): 733–743.

Giordano, G.; Blanchini, F.; Bruno, R.; Colaneri, P.; Di Filippo, A.; Di Matteo, A.; and Colaneri, M. 2020. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine*, 1–6.

Harvey, A.; and Kattuman, P. 2020. Time series models based on growth curves with applications to forecasting coronavirus. *Covid Economics, Vetted and Real-Time Papers*, (24).

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

Hoertel, N.; Blachier, M.; Blanco, C.; Olfson, M.; Massetti, M.; Rico, M. S.; Limosin, F.; and Leleu, H. 2020. A stochastic agent-based model of the SARS-CoV-2 epidemic in France. *Nature medicine*, 26(9): 1417–1421.

Kandula, S.; Hsu, D.; and Shaman, J. 2017. Subregional nowcasts of seasonal influenza using search trends. *Journal of medical Internet research*, 19(11): e370.

Kapoor, A.; Ben, X.; Liu, L.; Perozzi, B.; Barnes, M.; Blais, M.; and O'Banion, S. 2020. Examining COVID-19 Forecasting using Spatio-Temporal Graph Neural Networks. *arXiv preprint arXiv:2007.03113*.

Karpatne, A.; Atluri, G.; Faghmous, J. H.; Steinbach, M.; Banerjee, A.; Ganguly, A.; Shekhar, S.; Samatova, N.; and Kumar, V. 2017. Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on Knowledge and Data Engineering*, 29(10): 2318–2331.

Kipf, T. N.; and Welling, M. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *ICLR*.

Lai, G.; Chang, W.-C.; Yang, Y.; and Liu, H. 2018. Modeling long-and short-term temporal patterns with deep neural networks. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, 95–104.

Li, Y.; Yu, R.; Shahabi, C.; and Liu, Y. 2017. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv preprint arXiv:1707.01926*.

Loli Piccolomini, E.; and Zama, F. 2020. Monitoring Italian COVID-19 spread by a forced SEIRD model. *PloS one*, 15(8): e0237417.

Nair, V.; and Hinton, G. E. 2010. Rectified linear units improve restricted boltzmann machines. In *ICML*.

Nsoesie, E.; Mararthe, M.; and Brownstein, J. 2013. Forecasting peaks of seasonal influenza epidemics. *PLoS currents*, 5.

Petropoulos, F.; and Makridakis, S. 2020. Forecasting the novel coronavirus COVID-19. *PloS one*, 15(3): e0231236.

Ramchandani, A.; Fan, C.; and Mostafavi, A. 2020. DeepCOVIDNet: An Interpretable Deep Learning Model for Predictive Surveillance of COVID-19 Using Heterogeneous Features and Their Interactions. *arXiv preprint arXiv:2008.00115*.

Rangarajan, P.; Mody, S. K.; and Marathe, M. 2019. Forecasting dengue and influenza incidences using a sparse representation of Google trends, electronic health records, and time series data. *PLoS computational biology*, 15(11): e1007518.

Ray, E. L.; Wattanachit, N.; Niemi, J.; Kanji, A. H.; House, K.; Cramer, E. Y.; Bracher, J.; Zheng, A.; Yamana, T. K.;

Xiong, X.; et al. 2020. Ensemble Forecasts of Coronavirus Disease 2019 (COVID-19) in the us. *MedRXiv*.

Ribeiro, M. H. D. M.; da Silva, R. G.; Mariani, V. C.; and dos Santos Coelho, L. 2020. Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil. *Chaos, Solitons & Fractals*, 109853.

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

Venkatramanan, S.; Chen, J.; Gupta, S.; Lewis, B.; Marathe, M.; Mortveit, H.; and Vullikanti, A. 2017. Spatio-temporal optimization of seasonal vaccination using a metapopulation model of influenza. In *2017 IEEE International Conference on Healthcare Informatics (ICHI)*, 134–143. IEEE.

Venkatramanan, S.; Lewis, B.; Chen, J.; Higdon, D.; Vullikanti, A.; and Marathe, M. 2018. Using data-driven agent-based models for forecasting emerging infectious diseases. *Epidemics*, 22: 43–49.

Venna, S. R.; Tavanaei, A.; Gottumukkala, R. N.; Raghavan, V. V.; Maida, A. S.; and Nichols, S. 2019. 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.; Ben, X.; Adiga, A.; Sadilek, A.; Tendulkar, A.; Venkatramanan, S.; Vullikanti, A.; Aggarwal, G.; Talekar, A.; Chen, J.; et al. 2020. Using Mobility Data to Understand and Forecast COVID19 Dynamics. *medRxiv*.

Wang, Z.; Chakraborty, P.; Mekaru, S. R.; Brownstein, J. S.; Ye, J.; and Ramakrishnan, N. 2015. Dynamic poisson autoregression for influenza-like-illness case count prediction. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1285–1294. ACM.

Werbos, P. J. 1990. Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, 78(10): 1550–1560.

Wu, Y.; Yang, Y.; Nishiura, H.; and Saitoh, M. 2018. Deep learning for epidemiological predictions. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, 1085–1088. ACM.

Wu, Z.; Pan, S.; Long, G.; Jiang, J.; Chang, X.; and Zhang, C. 2020. Connecting the dots: Multivariate time series forecasting with graph neural networks. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 753–763.

Wu, Z.; Pan, S.; Long, G.; Jiang, J.; and Zhang, C. 2019. Graph wavenet for deep spatial-temporal graph modeling. *arXiv preprint arXiv:1906.00121*.

Yamana, T.; Pei, S.; and Shaman, J. 2020. Projection of COVID-19 Cases and Deaths in the US as Individual States Re-open May 4, 2020. *medRxiv*.

Yang, S.; Santillana, M.; and Kou, S. C. 2015. Accurate estimation of influenza epidemics using Google search data via ARGO. *Proceedings of the National Academy of Sciences*, 112(47): 14473–14478.

Yu, B.; Yin, H.; and Zhu, Z. 2017. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv preprint arXiv:1709.04875*.