

Epidemiology-aware Deep Learning for Infectious Disease Dynamics Prediction

Mutong Liu
csmtliu@comp.hkbu.edu.hk
Department of Computer Science,
Hong Kong Baptist University
Kowloon, Hong Kong, China

Yang Liu
csygliu@comp.hkbu.edu.hk
Department of Computer Science,
Hong Kong Baptist University
Kowloon, Hong Kong, China

Jiming Liu
jiming@comp.hkbu.edu.hk
Department of Computer Science,
Hong Kong Baptist University
Kowloon, Hong Kong, China

ABSTRACT

Infectious disease risk prediction plays a vital role in disease control and prevention. Recent studies in machine learning have attempted to incorporate epidemiological knowledge into the learning process to enhance the accuracy and informativeness of prediction results for decision-making. However, these methods commonly involve single-patch mechanistic models, overlooking the disease spread across multiple locations caused by human mobility. Additionally, these methods often require extra information beyond the infection data, which is typically unavailable in reality. To address these issues, this paper proposes a novel epidemiology-aware deep learning framework that integrates a fundamental epidemic component, the next-generation matrix (NGM), into the deep architecture and objective function. This integration enables the inclusion of both mechanistic models and human mobility in the learning process to characterize within- and cross-location disease transmission. From this framework, two novel methods, Epi-CNNRNN-Res and Epi-Cola-GNN, are further developed to predict epidemics, with experimental results validating their effectiveness.

CCS CONCEPTS

• **Computing methodologies** → **Neural networks**; **Regularization**; • **Applied computing** → **Forecasting**.

KEYWORDS

Epidemiological constraints; Deep learning; Infectious disease dynamics prediction

ACM Reference Format:
Mutong Liu, Yang Liu, and Jiming Liu. 2023. Epidemiology-aware Deep Learning for Infectious Disease Dynamics Prediction. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM '23), October 21–25, 2023, Birmingham, United Kingdom.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3583780.3615139

1 INTRODUCTION

Accurate prediction of disease transmission dynamics over space and time is essential for controlling the spread of infectious diseases, as it provides crucial information for quantifying the expected impact of potential outbreaks, and thus informs appropriate public-health responses such as intervention deployment and resource allocation to high-risk locations or populations [7].

Mechanistic modeling is the most fundamental way to quantitatively describe infectious disease dynamics [11]. Classical mechanistic models of infectious diseases include the susceptible-infected-recovered (SIR) model [15] and its variants [12]. Although such mechanistic modeling approaches have improved the characterization of disease transmission processes, they generally assumed that the transmission is governed ideally by predefined differential equations (DEs). This assumption, however, may not hold in reality, as such dynamics are generally determined by many transmission-related factors in a complex manner. Therefore, machine learning has become an indispensable tool for uncovering the underlying transmission patterns of diseases in a data-driven manner.

Various statistical and machine learning methods have been developed to model and predict disease transmission dynamics. Representative statistical methods include the autoregressive (AR) model and its variants [8, 18]. Recently, modern machine learning models, especially deep learning methods, have been widely utilized to model the spatiotemporal patterns of disease transmission. Typical examples include graph neural network (GNN) or convolutional neural network (CNN) based methods for spatial modeling [13, 16, 21], recurrent neural network (RNN) based methods for temporal modeling [1, 24], and hybrid-structured methods for spatiotemporal modeling [5, 6, 20, 22, 23, 25]. However, most of the above learning methods pursue the best fitting with training data while ignoring epidemiological mechanisms behind the disease transmission process (i.e., the prior knowledge about the disease propagation investigated by empirical studies). As a result, the model parameters and disease transmission risks inferred using such a purely data-driven manner may lack clear epidemiological meaning, thus leading to unsatisfactory generality and being less informative for public-health decision-making.

Recently, some efforts have been made in integrating mechanistic models of infectious diseases into machine learning methods as regularizers or constraints to help the prediction of disease risks. Arik et al. integrated a generalized additive model into the susceptible-exposed-infected-recovered (SEIR) model for COVID-19 progression prediction [2]. Kargas et al. utilized the SIR model to regularize the temporal dimension of a three-way tensor of reported case numbers, and developed a regularized tensor factorization method for future case number prediction [14]. Gao et al. proposed a spatio-temporal attention network (STAN), which combined the GNN, the attention mechanism, and the SIR constraints

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CIKM '23, October 21–25, 2023, Birmingham, United Kingdom

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0124-5/23/10…$15.00
https://doi.org/10.1145/3583780.3615139

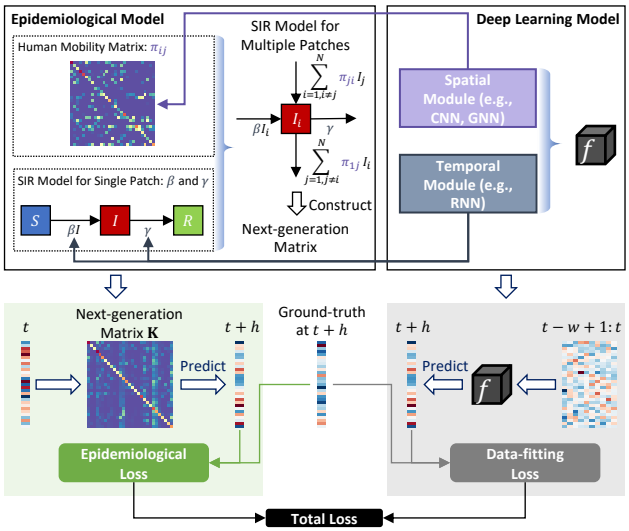


Figure 1: The schematic illustration of our framework.

for COVID-19 pandemic prediction [10]. Wang et al. developed the causal-based graph neural network (CausalGNN), with an attention-based dynamic GNN for spatiotemporal modeling and a susceptible-infected-recovered-deceased (SIRD) model for causal-based regularization [19]. However, the above approaches face two drawbacks. Firstly, they incorporate the single-patch mechanistic models of infectious diseases into the learning process. Such single-patch epidemiological constraints or regularizers only describe the disease dynamics within individual geographical locations; they overlook the disease spread across multiple locations caused by human mobility, which is of great importance in characterizing the spatiotemporal transmission patterns of infectious diseases. Secondly, most of these epidemic-regularized/constrained methods, in addition to the infection data, require extra information such as the susceptible and recovered data to calibrate the epidemiological parameters or optimize the loss function. This extra information, however, is often unavailable in reality, thus making the aforementioned methods less useful when only the infection data is available.

To address these two challenges, in this paper, we propose a novel epidemiology-aware deep learning framework that (1) integrates a fundamental epidemic component—the next-generation matrix (NGM) [9] into the deep learning architecture with spatial and temporal modules, and (2) only requires the infection data as the model input. Figure 1 illustrates the idea of the proposed framework. First, we integrate human mobility into single-patch mechanistic models to formulate a multi-patch epidemic model for cross-location transmission dynamics description (upper-left). Based on this multi-patch model, we construct its NGM, which is a function of the human mobility matrix and epidemiological parameters, to represent the disease dynamics. To make epidemiological knowledge compatible with data-driven models, we utilize spatial and temporal modules in the deep learning architecture (upper-right) to represent the human mobility matrix and to predict epidemiological parameters, respectively (the top of Figure 1). Finally, we jointly minimize the data-fitting loss from the deep learning model and the epidemiological loss from the NGM module to optimize the model parameters, so as to provide an epidemiology-aware characterization of disease dynamics (the bottom of Figure 1). Under the proposed framework, we further develop two novel deep learning methods based on the representative CNNRNN-Res [20] and Cola-GNN [6], respectively, for epidemic prediction.

2 EPIDEMIOLOGY-AWARE DEEP LEARNING

2.1 Problem Statement

Given the spatiotemporal observations of disease dynamics: $\mathbf{D} = \{\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_T\}^{\mathrm{T}} \in \mathbb{R}^{T\times N}$, where N denotes the number of locations of interest, T denotes the number of time steps, and $\mathbf{d}_t = \{d_{t,1}, d_{t,2}, \cdots, d_{t,N}\}^{\mathrm{T}} \in \mathbb{R}^{N\times 1}$ $(t = 1, \cdots, T)$ is an N-dimensional vector, with $d_{t,i}$ $(i = 1, \cdots, N)$ being the disease risk (generally quantified by the infection number or disease activity level) of location i at time step t. The task of disease dynamics prediction can be formulated as follows:

$$\mathbf{D}_{Future} = f(\mathbf{D}_{Historical}). \tag{1}$$

Here $\mathbf{D}_{Future}$ can be $\mathbf{d}_{T+1}$ (i.e., one-step ahead prediction) or $\mathbf{d}_{T+h}$ (i.e. multiple-step ahead prediction); h, called prediction horizon, is the number of time steps to be predicted ahead; and $\mathbf{D}_{Historical}$ can be $\mathbf{d}_T$ or $\{\mathbf{d}_1, \cdots, \mathbf{d}_T\}$. In this work, we consider multiple-step ahead prediction, which includes one-step ahead prediction as a special case. The objective of the epidemic prediction task is to learn the prediction function f from given observations, so that it can make accurate predictions of $\boldsymbol{D}_{Future}$ based on $\boldsymbol{D}_{Historical}$. In this paper, we expect to utilize the epidemiological dynamics of infectious diseases to regularize and constrain the learning process, thus avoiding potential overfitting on training data and further enhancing the model's generality. In the following subsections, we elaborate more details on the design of the proposed framework, especially on its epidemiological component.

2.2 Mechanistic Modeling of Disease Dynamics

In this paper, we use the classical SIR model [15] as an example to demonstrate how epidemiological dynamics can be incorporated into our framework using the NGM representation for infectious disease prediction. Other epidemic models such as the SEIR model and the SIRD model [12] can also be flexibly integrated into our framework using a similar way, depending on the need for modeling specific infectious diseases or considering specific scenarios.

Mechanistic modeling aims to characterize the disease dynamics process using a set of DEs [11]. In our case, the DEs of the SIR model for location i can be represented as follows:

$$\begin{cases} dS_i/dt = -\beta_i I_i S_i/N_i \\ dI_i/dt = \beta I_i S_i/N_i - \gamma_i I_i, \\ dR_i/dt = \gamma_i I_i \end{cases} \tag{2}$$

where S_i, I_i, and R_i denote the numbers of susceptible, infected, and recovered human individuals in location i, respectively; γ_i and β_i are the recovery rate of infected individuals and the infected rate of susceptible individuals in location i, respectively; and N_i is the population size in location i. For simplicity, we follow a widely used assumption that the disease is transmitted in a totally susceptible environment [17], i.e., $\frac{S_i}{N_i} = 1$.

2.3 Mobility-integrated Representation for Spatiotemporal Transmission Modeling

The above single-patch mechanistic model describes the disease dynamics within individual geographical locations. To further consider the spread of infectious diseases over space, which is generally caused by human mobility across different locations, we define a mobility matrix $\mathbf{\Pi}$ as follows:

$$\mathbf{\Pi} = \begin{pmatrix} \pi_{1,1} & \cdots & \pi_{1,N} \\ \vdots & \ddots & \vdots \\ \pi_{N,1} & \cdots & \pi_{N,N} \end{pmatrix}, \tag{3}$$

where $\pi_{i,j}$ denotes the intensity of human mobility from location i to location j. In many real-world scenarios, however, human mobility information is not directly accessible due to various reasons such as data privacy issues. In our framework, therefore, we estimate the above human mobility matrix via a data-driven manner, which will be introduced in the Subsection 2.5. With the human mobility integrated, we generalize the DE of the compartment I in the SIR model for location i ($i = 1, \cdots, N$) as follows:

$$\frac{dI_i}{dt} = \beta_i I_i - \gamma_i I_i - \sum_{j=1,\ j\neq i}^{N} \pi_{i,j} I_i + \sum_{j=1,\ j\neq i}^{N} \pi_{j,i} I_j. \tag{4}$$

Note that here we only show the equation for I_i — the variable that we aim to predict. The generalized equations for S_i and R_i can also be formulated accordingly if needed.

2.4 Construction of NGM

To represent the DE-characterized spatiotemporal transmission dynamics in a form that is easily compatible with data-driven models, we construct the NGM [9] of infectious diseases based on Eq. (4). Specifically, the NGM for all N locations can be formulated as

$$\mathbf{K} = \boldsymbol{\beta} \cdot (\boldsymbol{\gamma} + \mathbf{A})^{-1}, \tag{5}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are $N \times N$ diagonal matrices with the epidemiological parameters β_i and γ_i ($i = 1, \cdots, N$) from the SIR model being the diagonal elements, respectively. Here we have $\mathbf{A} = \mathbf{W} - \mathbf{\Pi}^T$, where $\mathbf{W}$ is the degree matrix of $\mathbf{\Pi}$. Given the formulation of $\mathbf{K}$, the DE-characterized dynamics of the compartment I for all N locations can be represented as follows:

$$\begin{aligned} \mathbf{I}^{(t+1)} &= \mathbf{K} \cdot \mathbf{I}^{(t)} \\ &= \boldsymbol{\beta} \cdot (\boldsymbol{\gamma} + (\mathbf{W} - \mathbf{\Pi}^T))^{-1} \cdot \mathbf{I}^{(t)}, \end{aligned} \tag{6}$$

where $\mathbf{I}^{(t)} = [I_1^{(t)}, \cdots, I_i^{(t)}, \cdots, I_N^{(t)}]^T$. Based on the definition of NGM, Eq. (6) can be used to directly predict the disease risks at time step $t+1$ (quantified by $\mathbf{d}_{t+1}$) based on the observations at time step t (i.e., $\mathbf{d}_t$), and has clear epidemiological meaning. Therefore, we will embed this representation in our framework to enable the desired epidemiology-aware learning.

2.5 Epidemiology-aware Deep Learning

To model the spatiotemporal transmission dynamics of infectious diseases, we incorporate the epidemiology-aware component into deep architectures by (1) utilizing the spatial and temporal modules to represent and infer epidemiology-related parameters and (2) designing the epidemiology-informed loss function. In this subsection, we explain our idea of integrating the aforementioned NGM component for epidemiology-aware learning based on two representative deep architectures, the CNNRNN-Res [20] and Cola-GNN [6], as examples. In fact, the epidemiological component can also be flexibly incorporated into other deep learning structures designed for spatiotemporal prediction.

2.5.1 Epi-CNNRNN-Res. The CNNRNN-Res structure is composed of three modules [20]: (1) a CNN module to model spatial interactions between different locations, (2) an RNN module to capture the temporal dependency of data, and (3) a residual module to avoid overfitting. In their CNN module, an adjacency parameter matrix $\mathbf{\Phi}_G$ is utilized to replace the grid filter in conventional CNNs for better representation of spatial relations; for the RNN and residual modules, the Gated Recurrent Unit (GRU) [4] is used in the RNN module to capture the temporal dependency, while the residual link method is adopted by the residual module to prevent overfitting via densely connecting the last layer to almost all previous layers.

In our design, we expect the $\mathbf{\Phi}_G$ to explicitly characterize the infectious disease transmission over locations. As such transmission is often triggered by cross-location human mobility, we use $\mathbf{\Phi}_G$ to represent the defined human mobility matrix. We apply the softmax function to each row of $\mathbf{\Phi}_G$ to ensure that the value of each element in $\mathbf{\Phi}_G$ fall into the interval $[0, 1]$, which can be understood as the probability of the human mobility from one location to another.

For β and γ, they generally vary with locations and change over time. Thus, we design an additional temporal module to capture and predict the changing patterns of these two epidemiological parameters from the historical data. Specifically, for each of them, we adopt a standard RNN to generate the temporal embedding based on the time series of historical disease risks and feed the embedding into a multi-layer perceptron (MLP) to predict future values. Finally, we use the learned matrix and epidemiological parameters to calculate $\mathbf{K}$ using Eq. (5).

2.5.2 Epi-Cola-GNN. The Cola-GNN consists of three modules [6]: (1) a dilated convolution module to capture the temporal dependency; (2) the RNN and attention module to capture the spatial dependency; and (3) the GNN module to pass the message of learned temporal features on the learned spatial relation graph. The final prediction of disease risks will be generated by an MLP, which takes the learned node embedding of GNN as the input.

In the spatial module of Cola-GNN, a spatial dependency matrix (denoted as $\mathbf{\Phi}'_G$) is obtained by applying the attention mechanism. Similar to our design in Section 2.5.1, we apply the softmax function to re-weight each row of $\mathbf{\Phi}'_G$ and use the re-weighted $\mathbf{\Phi}'_G$ to represent the human mobility matrix. For β and γ, we use the same RNN and MLP structures introduced in Section 2.5.1 to predict their spatiotemporally varying values.

2.5.3 Objective Function. The objective function of the proposed framework is given as follows:

$$\min \sum_{t=w}^{T_{train}+w} \left(\|\mathbf{d}_{t+h} - f(\mathbf{d}_{t-w+1:t})\|_2^2 + \lambda \|\mathbf{d}_{t+h} - \mathbf{K}\mathbf{d}_t\|_2^2 \right), \tag{7}$$

where w is the window length of the input historical data, T_{train} denotes the number of samples used for training, h is the prediction

Table 1: Performance comparison of baselines and proposed methods on the US-HHS-Flu dataset. The best and the second-best results in each setting are highlighted using boldface and underlining, respectively. H denotes the prediction horizon.

H	Metrics	AR	GAR	VAR	CNNRNN-Res	Cola-GNN	Epi-CNNRNN-Res	Epi-Cola-GNN
1	RMSE	0.2616	0.2593	0.3031	0.2514	0.2444	**0.2402**	0.2430
	MAE	0.1832	0.1790	0.2163	0.1751	**0.1601**	0.1615	0.1602
	CORR	0.9430	0.9422	0.9301	0.9508	0.9510	0.9499	**0.9511**
2	RMSE	0.3639	0.3797	0.4142	0.3467	0.3415	**0.3091**	0.3284
	MAE	0.2580	0.2623	0.2818	0.2429	0.2256	0.2093	**0.2135**
	CORR	0.8938	0.8819	0.8677	0.9121	0.9154	**0.9291**	0.9161
4	RMSE	0.4718	0.5230	0.5042	0.4369	0.4989	**0.3952**	0.4094
	MAE	0.3449	0.3862	0.3402	0.3088	0.2954	0.2670	**0.2626**
	CORR	0.8227	0.7710	0.8070	0.8430	0.8262	**0.8812**	0.8773

horizon (refer to Section 2.1). The temporal resolution of the horizon is consistent with that of the training data. In Eq. (7), the first term is the prediction error of the deep learning model. We aim to minimize it to achieve good prediction accuracy. The second term could be explained as an epidemiology-aware regularizer to enforce the prediction results as well as the learned parameters (e.g., the parameter matrices $\mathbf{\Phi}_G$, $\mathbf{\Phi}'_G$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$) to be consistent with an epidemiologically meaningful dynamics defined by the NGM, thus improving the generality of the model. The parameter λ is used to balance these two terms. To avoid overfitting, we enforce the sparsity of the learned matrices $\mathbf{\Phi}_G$ and $\mathbf{\Phi}'_G$ according to the geographical adjacency of different locations.

3 EXPERIMENTS

In this section, we validate the effectiveness of the proposed methods[1], Epi-CNNRNN-Res and Epi-Cola-GNN, by comparing them with five representative disease dynamics prediction algorithms on US-HHS-Flu[2], a real-world public dataset of influenza-like illness (ILI). This dataset records weekly influenza activity levels (measured by the weighted ILI metric) in 10 districts (divided by HHS) of the mainland United States between the first week of 2010 and the 52nd week of 2016 ($7 \times 52 = 364$ weeks in total). Therefore, the spatiotemporal size of this dataset is 10×364. The five methods selected for performance comparison are Autoregression (AR) [8], Global AR (GAR), Vector AR (VAR) [18], CNNRNN-Res [20], and Cola-GNN [6]. The first three are typical time series prediction algorithms while the last two are state-of-the-art deep learning methods designed for epidemic prediction.

We use two standard criteria, the Root-Mean-Square Error (RMSE) [3, 6] and the Mean Absolute Error (MAE) [5], to evaluate the prediction accuracy of proposed methods. To further test the overall consistency between the predicted dynamics and the real epidemic trend, we use Pearson's Correlation Coefficient (CORR) [6, 20] to quantify the correlation between the prediction and ground truth. For each method in our comparison, we calculate its average CORR over all N locations.

[1]Code available at https://github.com/gigg1/CIKM2023EpiDL

[2]https://gis.cdc.gov/grasp/fluview/fluportaldashboard.html, 25 May 2023, date last accessed

3.1 Experimental Settings

For all methods, we partition the dataset into 60%, 20%, and 20% for training, validation, and testing, respectively. For AR, GAR, VAR, and CNNRNN-Res, we use the settings reported in [20] and re-run the experiments. For Cola-GNN, we follow the description in [6]: for hidden dimensions of the RNN module and the initial learning rate, we search the optimal values from $\{10, 20, 30\}$ and $\{0.001, 0.005, 0.01\}$, respectively; for other parameters, we use the default values. For our Epi-CNNRNN-Res, we set the candidates of window size as $\{32, 64\}$, the hidden dimension for GRU as $\{5, 10\}$, the dropout rate for GRU layers as 0.2, the number of residual links as $\{8, 16\}$, the weight of epidemiological loss λ as $\{0.5, 1\}$, and the ratio of the output of CNNRNN and residual links as $\{0.05, 0.1\}$. For our Epi-Cola-GNN, we set hidden dimensions of the RNN module to $\{10, 20, 30\}$, the initial learning rate to $\{0.001, 0.01, 0.1\}$, and the weight of epidemiological loss λ to $\{0.1, 0.5, 1\}$. For both of our methods, we conduct a grid search over the above parameter combinations and report the best performance.

3.2 Results and Analysis

The performance of our methods and other baselines on the US-HHS-Flu dataset is reported in Table 1. For each setting, we highlight the best performance using boldface and the second-best results using underlining. From the table, we can observe that the deep learning based methods, i.e., CNNRNN-Res, Cola-GNN, Epi-CNNRNN-Res, and Epi-Cola-GNN, consistently perform better than the classical AR, GAR, and VAR. This shows the power of deep architectures in extracting discriminative representations for prediction. By integrating the epidemiological component into the learning framework, the proposed Epi-CNNRNN-Res and Epi-Cola-GNN further outperform original algorithms, i.e., CNNRNN-Res and Cola-GNN, and achieve the best or the second-best performance among all methods in most scenarios.

4 CONCLUSIVE REMARKS

In this paper, we developed a novel framework to integrate the epidemiological model into deep learning methods for disease dynamics prediction. By integrating the NGM into the learning procedure and the objective function of deep models, the proposed methods can characterize both within-location disease dynamics and cross-location disease transmission for integrative and epidemiologically meaningful spatiotemporal modeling. Experimental results on a public infectious disease dataset validate the effectiveness of our methods in terms of prediction accuracy. In our future work, we plan to generalize our design to include more complex mechanistic models and incorporate the proposed epidemiology-aware design into different deep structures.

ACKNOWLEDGMENTS

This work was supported in part by the Ministry of Science and Technology of China under Grants 2021ZD0112501 and 2021ZD0112502, the General Research Fund of Hong Kong SAR under Grants RGC/HKBU 12201619, 12202220, and 12203122, and the Research Fund from the Guangdong Basic and Applied Basic Research Foundation under Project 2022A1515010124.

REFERENCES

[1] Nurul Absar, Nazim Uddin, Mayeen Uddin Khandaker, and Habib Ullah. 2022. The efficacy of deep learning based LSTM model in forecasting the outbreak of contagious diseases. *Infectious Disease Modelling* 7, 1 (2022), 170–183.

[2] Sercan Arik, Chun-Liang Li, Jinsung Yoon, Rajarishi Sinha, Arkady Epshteyn, Long Le, Vikas Menon, Shashank Singh, Leyou Zhang, Martin Nikoltchev, et al. 2020. Interpretable sequence learning for COVID-19 forecasting. *Advances in Neural Information Processing Systems* 33 (2020), 18807–18818.

[3] Defu Cao, Yujing Wang, Juanyong Duan, Ce Zhang, Xia Zhu, Congrui Huang, Yunhai Tong, Bixiong Xu, Jing Bai, Jie Tong, et al. 2020. Spectral temporal graph neural network for multivariate time-series forecasting. *Advances in neural information processing systems* 33 (2020), 17766–17778.

[4] Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 1724–1734.

[5] Yue Cui, Chen Zhu, Guanyu Ye, Ziwei Wang, and Kai Zheng. 2021. Into the unobservables: A multi-range encoder-decoder framework for COVID-19 prediction. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 292–301.

[6] Songgaojun Deng, Shusen Wang, Huzefa Rangwala, Lijing Wang, and Yue Ning. 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.

[7] R. Detels, M. Gulliford, K. Q. Abdool, and C. C. Tan. 2017. *Oxford Textbook of Global Public Health*. Oxford University Press.

[8] Marcel Dettling. 2013. Applied time series analysis. *ETH Scripts* (2013).

[9] Odo Diekmann, JAP Heesterbeek, and Michael G Roberts. 2010. The construction of next-generation matrices for compartmental epidemic models. *Journal of the royal society interface* 7, 47 (2010), 873–885.

[10] Junyi Gao, Rakshith Sharma, Cheng Qian, Lucas M Glass, Jeffrey Spaeder, Justin Romberg, Jimeng Sun, and Cao Xiao. 2021. STAN: spatio-temporal attention network for pandemic prediction using real-world evidence. *Journal of the American Medical Informatics Association* 28, 4 (2021), 733–743.

[11] N. C. Grassly and C. Fraser. 2008. Mathematical models of infectious disease transmission. *Nature Reviews Microbiology* 6 (2008), 477–487.

[12] H. W. Hethcote. 2000. The Mathematics of Infectious Diseases. *SIAM Rev.* 42, 4 (2000), 599–653.

[13] Amol Kapoor, Xue Ben, Luyang Liu, Bryan Perozzi, Matt Barnes, Martin Blais, and Shawn O'Banion. 2020. Examining covid-19 forecasting using spatio-temporal graph neural networks. *arXiv preprint arXiv:2007.03113* (2020).

[14] Nikos Kargas, Cheng Qian, Nicholas D Sidiropoulos, Cao Xiao, Lucas M Glass, and Jimeng Sun. 2021. Stelar: Spatio-temporal tensor factorization with latent epidemiological regularization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 4830–4837.

[15] William Ogilvy Kermack and Anderson G McKendrick. 1927. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character* 115, 772 (1927), 700–721.

[16] Kookjin Lee, Jaideep Ray, and Cosmin Safta. 2021. The predictive skill of convolutional neural networks models for disease forecasting. *PLOS ONE* 16, 7 (2021), 1–26.

[17] Jiming Liu and Shang Xia. 2020. *Computational Epidemiology*. Springer.

[18] Helmut Lütkepohl and Markus Krätzig. 2004. *Applied time series econometrics*. Cambridge university press.

[19] Lijing Wang, Aniruddha Adiga, Jiangzhuo Chen, Adam Sadilek, Srinivasan Venkatramanan, and Madhav Marathe. 2022. CausalGNN: Causal-based Graph Neural Networks for Spatio-Temporal Epidemic Forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 12191–12199.

[20] Yuexin Wu, Yiming Yang, Hiroshi Nishiura, and Masaya Saitoh. 2018. Deep learning for epidemiological predictions. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. 1085–1088.

[21] Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, Xiaojun Chang, and Chengqi Zhang. 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.

[22] Feng Xie, Zhong Zhang, Liang Li, Bin Zhou, and Yusong Tan. 2023. EpiGNN: Exploring spatial transmission with graph neural network for regional epidemic forecasting. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2022, Grenoble, France, September 19–23, 2022, Proceedings, Part VI*. Springer, 469–485.

[23] Feng Xie, Zhong Zhang, Xuechen Zhao, Bin Zhou, and Yusong Tan. 2022. Inter- and Intra-Series Embeddings Fusion Network for Epidemiological Forecasting. *arXiv preprint arXiv:2208.11515* (2022).

[24] Zifeng Yang, Zhiqi Zeng, Ke Wang, Sook-San Wong, Wenhua Liang, Mark Zanin, Peng Liu, Xudong Cao, Zhongqiang Gao, Zhitong Mai, et al. 2020. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of thoracic disease* 12, 3 (2020), 165.

[25] Shun Zheng, Zhifeng Gao, Wei Cao, Jiang Bian, and Tie-Yan Liu. 2021. HierST: A Unified Hierarchical Spatial-temporal Framework for COVID-19 Trend Forecasting. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 4383–4392.