TLife-LSTM: Forecasting Future COVID-19 Progression with Topological Signatures of Atmospheric Conditions

Ignacio Segovia-Dominguez[1,2(✉)], Zhiwei Zhen[1], Rishabh Wagh[1], Huikyo Lee[2], and Yulia R. Gel[1]

[1] The University of Texas at Dallas, Richardson, TX 75080, USA
ignacio.segoviadominguez@utdallas.edu
[2] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA

Abstract. Understanding the impact of atmospheric conditions on SARS-CoV2 is critical to model COVID-19 dynamics and sheds a light on the future spread around the world. Furthermore, geographic distributions of expected clinical severity of COVID-19 may be closely linked to prior history of respiratory diseases and changes in humidity, temperature, and air quality. In this context, we postulate that by tracking topological features of atmospheric conditions over time, we can provide a quantifiable structural distribution of atmospheric changes that are likely to be related to COVID-19 dynamics. As such, we apply the machinery of persistence homology on time series of graphs to extract topological signatures and to follow geographical changes in relative humidity and temperature. We develop an integrative machine learning framework named Topological Lifespan LSTM (TLife-LSTM) and test its predictive capabilities on forecasting the dynamics of SARS-CoV2 cases. We validate our framework using the number of confirmed cases and hospitalization rates recorded in the states of Washington and California in the USA. Our results demonstrate the predictive potential of TLife-LSTM in forecasting the dynamics of COVID-19 and modeling its complex spatio-temporal spread dynamics.

Keywords: Dynamic networks · COVID-19 · Topological Data Analysis · Long Short Term Memory · Environmental factors · Clinical severity

1 Introduction

Nowadays, there is an ever-increasing spike of interest in enhancing our understanding of hidden mechanisms behind transmission of SARS-CoV2 (i.e., the virus that causes COVID-19) and its potential response to atmospheric conditions including temperature and relative humidity [4,16,23]. Understanding the impact of atmospheric conditions on COVID-19 trajectory and associated mortality is urgent and critical, not only in terms of efficiently responding to the

K. Karlapalem et al. (Eds.): PAKDD 2021, LNAI 12712, pp. 201–212, 2021.
https://doi.org/10.1007/978-3-030-75762-5_17

current pandemic (e.g., preparing an adequate health care response in areas with expected higher clinical coronavirus severity), but also in terms of forecasting impending hotspots and potential next-wave occurrences.

However, as shown in many recent biosurveillance studies [2,20], the non-trivial relationship between the spatio-temporal dynamics of atmospheric data and disease transmission may not be captured well by conventional metrics based on Euclidean distances. This phenomenon can be partially explained by a sophisticated spatio-temporal dependence structure of the atmospheric conditions. A number of recent results on (re)emerging infectious have introduced the concepts of topological data analysis (TDA) into modeling the spread of climate-sensitive viruses. In particular, TDA and, specifically, persistent homology have been employed by [11] for analysis of influenza-like illness during the 2008–2013 flu seasons in Portugal and Italy. Most recently, [19,25] show the explanatory and predictive power of TDA for analysis of Zika spread. The obtained results show that topological descriptors of spatio-temporal dynamics of atmospheric data tend to improve forecasting of infectious diseases. These findings are largely due to the fact that topological descriptors allow for capturing higher-order dependencies among atmospheric variables that otherwise might be unassessable via conventional spatio-temporal modeling approaches based on geographical proximity assessed via Euclidean distance.

In this paper, we develop a novel predictive deep learning (DL) platform for COVID-19 spatio-temporal spread, coupled with topological information on atmospheric conditions. The key idea of the new approach is based on the integration of the most essential (or *persistent*) topological descriptors of temperature and humidity, into the Long Short Term Memory (LSTM) model. The new Topological Lifespan Long Short Term Memory (TLife-LSTM) approach allows us to track and forecast COVID-19 spread, while accounting for important local and global variability of epidemiological factors and its complex dynamic interplay with environmental and socio-demographic variables.

The significance of our paper can be summarized as follows:

- To the best of our knowledge, this is the first paper, systematically addressing complex nonlinear relationships between atmospheric conditions and COVID-19 dynamics.
- The proposed TLife-LSTM model, coupled with TDA, allows for efficient and mathematically rigorous integration of hidden factors behind COVID-19 progression which are otherwise inaccessible with conventional predictive approaches, based on Euclidean distances.
- Our case studies indicate that the new TLife-LSTM approach delivers a highly competitive performance for COVID-19 spatio-temporal forecasting in the states of California and Washington on a county-level basis. These findings suggest that TLife-LSTM and, more generally, biosurveillance tools based on topological DL might be the most promising predictive tools for COVID-19 dynamics under the scenarios of limited and noisy epidemiological data records.

2 Related Work

Throughout the last year numerous studies have found that DL and, particularly, LSTM exhibit predictive utility for COVID-19 spread at the country level and varying forecasting horizons e.g. 2, 4, 6, 8, 10, 12 and 14 days [1,3,5,27]. Similar biosurveillance performance in terms of Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), have been also obtained for various Recurrent Neural Networks (RNN) architectures [24,27]. Hence, despite requiring higher historical epidemiological records, DL tools nowadays are viewed as one of the most promising modeling approaches to address short- and medium-term prediction of COVID-19 dynamics [22].

Topological Data Analysis (TDA) provides a rigorous theoretical background to explore topological and geometric features of complex geospatial data. TDA has been proven useful in a broad range of data science applications [8,15,21], including biosurveillance [11,19,25], COVID-19 spread visualization [10,12,18] and COVID-19 therapy analysis [9]. However, to the best of our knowledge, there yet exists no *predictive models for COVID-19, harnessing the power of TDA*. Even more, there are, still, few research studies incorporating atmospheric conditions as inputs of RNNs to forecasting COVID-19 spread. Our research makes contributions in both areas and creates connections between TDA and DL through adding topological signatures, i.e. covariates, into a LSTM architecture.

3 Problem Statement

Our primary goal is to construct a RNN model which harnesses the power of TDA to forecast COVID-19 dynamics at various spatial resolutions. We exploit the coordinate-free PH method to extract topological features from environmental variables and to summarize the most essential topological signatures of atmospheric conditions.

In this project we use two distinctive data types: 1) the number of COVID-19 confirmed cases/hospitalizations from US states at county-level, and 2) daily records of environmental variables from weather stations. Our goal is to model the COVID-19 dynamics and predict its spread and mortality at county-level in US states, while accounting for complex relationships between atmospheric conditions and current pandemic trends.

4 Background

We start from providing a background on the key concepts employed in this project, namely, TDA and LSTM neural networks.

4.1 Preliminaries on Topological Data Analysis

The fundamental idea of TDA is that the observed data $\mathbb{X}$ represent a discrete sample from some metric space and, due to sampling, the underlying structure of this space has been lost. The main goal is then to systematically retrieve the lost underlying structural properties, by quantifying dynamics of topological properties exhibited by the data, assessed through multiple user-defined (dis)similarity scales [6,7,14]. Such (dis)similarity measure can be, for instance, similarity of temperature values or homogeneity of COVID-19 counts in various counties. Hence, given this idea, the derived TDA summaries are expected to be inherently robust to missing data, to matching multiple spatio-temporal data resolutions, and other types of uncertainties in the observed epidemiological information.

The TDA approach is implemented in the three main steps (see the toy example in Fig. 1):

1. We associate $\mathbb{X}$ with some filtration of $\mathbb{X}$: $\mathbb{X}_1 \subseteq \mathbb{X}_2 \subseteq \ldots \subseteq \mathbb{X}_k = \mathbb{X}$.
2. We then monitor the evolution of various pattern occurrences (e.g., cycles, cavities, and more generally k-dimensional holes) in nested sequence of subsets. To make this process efficient and systematic, we equip $\mathbb{X}$ with certain combinatorial objects, e.g., simplicial complexes. Formally, a simplicial complex is defined as a collection $\mathcal{C}$ of finite subsets of $\mathcal{G}$ such that if $\sigma \in \mathcal{C}$ then $\tau \in \mathcal{C}$ for all $\tau \subseteq \sigma$. The basic unit of simplicial complexes is called the *simplex*, and if $|\sigma| = m + 1$ then σ is called an m-simplex. Here we employ a Vietoris–Rips (VR) complex which is one of the most widely used complexes due its computational cost benefits [8,14]. Filtration of $\mathbb{X}_1 \subseteq \mathbb{X}_2 \subseteq \ldots$ is then associated with filtration of VR complexes $VR_1 \subseteq VR_2 \subseteq \ldots$.
3. We track the index of VR complex t_b when each topological feature is first recorded (i.e., born) and the index of VR complex t_d when this feature is last seen (i.e., died). Then lifespan of this feature is $t_d - t_b$. The extracted topological characteristics with a longer lifespan are called *persistent* and are likelier connected to some fundamental mechanisms behind the underlying system organization and dynamics, e.g., higher-order interactions between disease transmissibility and changes in atmospheric variables. In turn, topological features with a shorter lifespan are referred to as *topological noise*.

We then use distributions of the recorded lifespans of topological features as inherent signatures, characterizing the observed data $\mathbb{X}$ and integrate these topological descriptors into our TLife-LSTM model.

4.2 Long Short Term Memory

RNNs and, particularly, Long Short Term Memory (LSTM) approaches have been successfully used to model time series and other time-dependent data [26]. The LSTM architecture addresses the problem of the gradient instability of predecessors and adds extra flexibility due to the memory storage and forget gates. Each LSTM unit contains three transition functions: input gate i_t, output gate o_t and forget gate f_t;

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{1}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{2}$$

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{3}$$

where σ represents the sigmoid function. The gates' information and the output vector h_{t-1} of the hidden layer obtained at step $t-1$ serve as input to the memory cell c_t, Eq. (5). It allows to recursively use vector output h_t to extract patterns from complex time series.

$$g_t = \tanh(W_g \cdot [h_{t-1}, x_t] + b_g) \tag{4}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \tag{5}$$

$$h_t = o_t \odot \tanh c_t \tag{6}$$

During the training stage, LSTM neural networks learn the weight matrices W_i, W_o, W_f and W_g, while the performance is affected by bias vectors b_i, b_o, b_f and b_g. Most of variants of LSTM architecture performs similarly well in large scale studies, see [17].

5 The Proposed Topological Lifespan Long Short Term Memory (TLife-LSTM) Approach

We now introduce a new DL framework to capture topological patterns in the observed spatio-temporal data and to model complex relationships in time series. Details are divided into two sections. First, we explain the construction of dynamic networks and posterior extraction of n-dimensional features. Then, we expand on the general methodology and provide specifics on the neural network architecture.

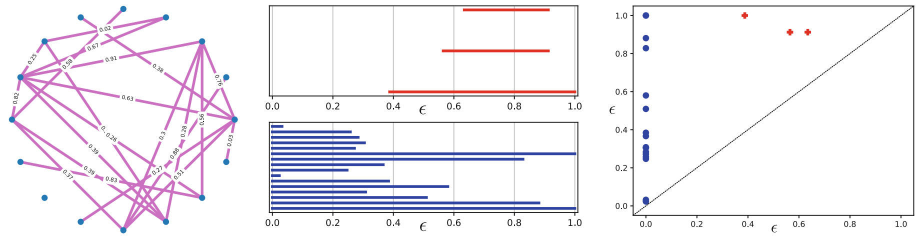
5.1 Topological Features of Environmental Dynamic Networks

Topological features, as defined in Sect. 4.1, describe the shape structure of underlying data which are invariant under continuous transformations such as twisting, bending, and stretching. In this study, we primarily focus on lifespans of such topological features computed on environmental dynamic networks.

Definition 1. *Let $\mathcal{G} = (V, E, \omega)$ be a weighted graph, where V is a set of vertices (e.g., weather stations), $E = \{e_1, e_2, \ldots\} \subseteq V \times V$ is a set of edges, and $\omega = \omega(e) : E \to \mathbb{Z}^+$ for all $e \in E$ are edge weights. (Here $\omega(e)$ may represent a level of similarity exhibited by atmospheric conditions at two weather stations).*

In particular, our biosurveillance methodology is inspired by the recent advancements of applying PH on dynamic networks. Let $\mathbb{G} = \{\mathcal{G}_t\}_{t=1}^{T} = \{\mathcal{G}_1, \ldots, \mathcal{G}_T\}$ be a sequence of time-evolving weighted networks observed over index-time t, with $1 \le t \le T < \infty$. Then, the objective is to compute and summarize persistent topological features at each $\mathcal{G}_t$; particularly lifespans as defined in Sect. 4.1. Hence, PH emerges as a topological technique to track changes in dynamic networks.

Meteorologists monitor atmospheric conditions using ground-based weather stations and satellites. Ground-based stations collect observations at different time resolutions. Let $O^{(i)}$ be historical-meteorological data, i.e., time series observations, collected from station $S^{(i)} = \{\text{latitude}, \text{longitude}\}$ such that $O^{(i)}$ is a time-ordered sample where each element contains λ observed measures of



(a) (b) (c)

Fig. 1. Example of applying persistent homology (PH) in a weighted graph. (a) Weighted graph. (b) Barcode, birth and death of topological summaries. Blue: 0-dimensional features. Red: 1-dimensional features. (c) Persistence diagram. (Color figure online)

weather conditions, i.e. $O_t^{(i)} = \{O_{t1}^{(i)}, O_{t2}^{(i)}, \dots, O_{t\lambda}^{(i)}\}$, and indexed by time t. Given a set of stations $\{S^{(i)}\}_{i=1}^{M}$ sited on the area of study (California and Washington), we construct a dynamic network $\mathbb{G} = \{\mathcal{G}_t\}_{t=1}^{T} = \{\mathcal{G}_1, \dots, \mathcal{G}_T\}$ using each station as vertex $v_i \in V$. Each undirected graph $\mathcal{G}_t$ contains edge weights $\omega_t(e)$ which vary over time according with weather measurements.

Our goal is to build time series of graphs that reflect the underlying connections of atmospheric conditions across local regions. That is, attaching weighted edges as a function of temporal-varying temperature or relative humidity provides a natural way to track weather changes between regions and link these with the trend in COVID-19 dynamics.

Algorithm 1. Topological Feature Extraction from Environmental Dynamic Networks

INPUT: Weather Conditions $\{O^{(i)}\}_{i=1}^{M}$; Station Locations $\{S^{(i)}\}_{i=1}^{M}$
OUTPUT: Topological summaries

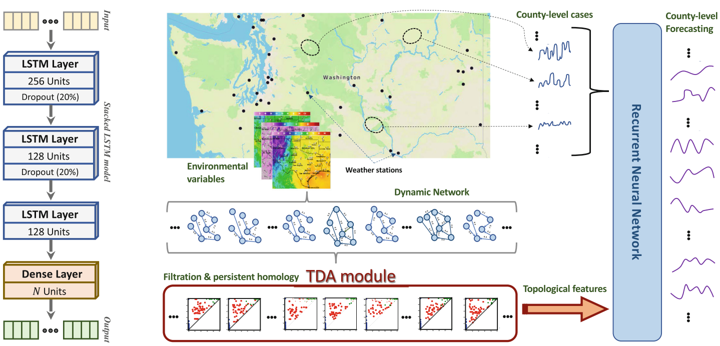
for $i \leftarrow 1 : M-1$ **do**
  for $j \leftarrow i+1 : M$ **do**
    Compute $L_{ij} = Norm(S^{(i)} - S^{(j)})$
for $t \leftarrow 1 : T$ **do**
  for $h \leftarrow 1 : \lambda$ **do**
    for $i \leftarrow 1 : M-1$ **do**
      for $j \leftarrow i+1 : M$ **do**
        Compute $D_{th}^{(ij)} = O_{th}^{(i)} - O_{th}^{(j)}$
    Standardize matrix D_{th}
  Compute $D_t = \frac{1}{2\lambda}\sum_{h=1}^{\lambda} D_{th} + \frac{1}{2}L$
  Compute $\omega_t(e) = D_t/\text{max element in } D_t$
  Generate $\mathcal{G}_t$ based on $\omega_t(e)$
Apply PH on dynamic networks $\mathbb{G} = \{\mathcal{G}_t\}_{t=1}^{T}$ for 0, 1, 2 and 3 dimensions
Calculate mean, total sum, variance of features' lifespan, and the number of topological features obtained through PH in each dimension

Algorithm 1 describes our methodology to create each weighted graph $\mathcal{G}_t$, based on atmospheric conditions, $O_t^{(i)}$, and its corresponding feature extraction. We propose to use the TDA tools, particularly, persistent homology (PH) to track qualitative features that persist across multiple scales. First, we generate a lower triangular matrix L to represent the distance between stations. Then for each weather measurement, we use D_{th} to record the difference in O_t between stations. For D_{th}, we take all low-triangular elements as a sample to do the standardization and generate a new matrix D_t based on the atmospheric connections D_{th} and the distance L. To ensure that weights in the dynamic networks are between 0 and 1, we divide D_t by its maximum element. Finally, we generate our dynamic networks and use the VR complex to get one persistence diagram on each $\mathcal{G}_t$. Based on features' lifespans, we compute the mean, total sum, and variance of the lifespan of topological features, and use these topological signatures as input for our RNN architecture.

5.2 Topological Machine Learning Methodology: TLife-LSTM

Let $\mathbf{Y} = \{Y_i\}_{i=1}^N$ be a multivariate time series, where $Y_i \in \mathbf{Y}$ and each Y_{it} is a sequence of outcomes from random variables Z_{it}, $t = 1, 2, \ldots, T$ indexed by time t. Hence, historical data of the $i - esim$ element Y_i is a time-ordered sample of observations $Y_{i1}, Y_{i2}, \ldots, Y_{iT}$. Given the partition of $\mathbf{Y} = \{\mathbf{A}, \mathbf{B}\}$ such that $\mathbf{A} = \{Y_1, Y_2, \ldots, Y_\tau\}$ and $\mathbf{B} = \{Y_{\tau+1}, Y_{\tau+2}, \ldots, Y_T\}$, vectors $Y_j \in \mathbf{A}$ are formed by historical data of COVID-19 progression, whilst vectors $Y_k \in \mathbf{B}$ are topological signatures from an environmental dynamic network $\mathbb{G}$.

Given multivariate time-series dataset $\mathbf{Y} = \{Y_i\}_{i=1}^N$, we train a LSTM model to capture complex relationships between variables. Notice that input comes from two different sources: 1) historical data of COVID-19 dynamics, and 2) persistent signatures of atmospheric conditions. We then create a suitable RNN architecture to handle these data. Figure 2a shows a graphical representation of the main modules in our RNN architecture. First, observations are received



(a) (b)

Fig. 2. Topological LSTM (a) RNN architecture. (b) Topological Lifespan LSTM.

as real vectors and passed through a series of three-stacked LSTM layers with 256, 128, and 128 units, respectively. To avoid overfitting, we follow the dropout regularization technique, i.e., randomly dropping units in the training step. Next, a densely connected layer accounts for additional relationship and prepares the data-flow to produce N outputs. In the current study, each output corresponds to each local region, e.g., counties in the two US states.

Figure 2b depicts our proposed Topological Lifespan LSTM (TLife-LSTM). The key idea is to assemble the computed topological summaries along with the time series of confirmed COVID-19 cases. Hence, we complement conventional biosurveillance information with the topological summaries, in order to enrich the input data in our deep learning phase. Although there have been various approaches using neural networks to predict COVID-19 dynamics, our methodology is unique as it integrates atmospheric conditions in its learning step. Furthermore, incorporation of topological signatures aims to ensure that the learning phase only relates distinctive and persistent features of atmospheric conditions along with intrinsic dynamics of COVID-19 progression. Hence, with TLife-LSTM, we extract higher-order dependence properties among temperature, relative humidity, and COVID-19 spread which are otherwise inaccessible with more conventional methods based on the analysis of geographic proximity.

6 Experiments

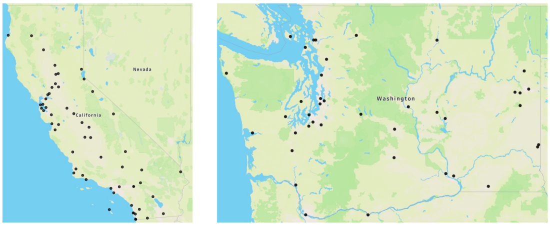
To assess the predictive capabilities of the proposed approach, we present experimental forecasts of COVID-19 dynamics: 1) the number of confirmed cases (**Cases**), and 2) the number of hospitalizations (**Hospi**). To evaluate the performance of our forecasts, we compare results using common evaluation metrics and present insightful visualizations at county-level. All experiments are run using Keras and Tensorflow, source codes[1] are published online to encourage reproducibility and replicability.

6.1 Data, Experimental Setup and Comparative Performance

Our experiments have been carried out using collected data in California and Washington. Particularly, our methodology produces daily COVID-19 progression and hospitalization forecasts at county-level resolution.

To build environmental dynamic networks and extract topological features, we select atmospheric measurements from hourly-updated reports of the National Center for Environmental Information, NCEI[2]. In this study, we focus our attention on temperature, humidity, and visibility. Since the observations are made hourly, we use the first record of each day as our representative daily measurement from 66 meteorological stations in each state, see Fig. 3. Daily records on COVID-19 cases and hospitalizations is from the data repository by Johns

[1] Available at Source codes (repository).
[2] Available at https://www.ncei.noaa.gov.



(a) (b)

Fig. 3. Land-based selected meteorological stations. (a) California state. (b) Washington state.

Hopkins University[3], see [13], which includes aggregated data sources from the World Health Organization, European Center for Disease Prevention and Control, and US Center for Disease Control and Prevention. For information about the COVID-19 disease progression, and additional modeling resources, we use the MIDAS online portal for COVID-19 modeling research[4].

We split the datasets into a training set, from April 15 to August 31, and a test set, the whole month of September. We train our RNN architecture to produce daily predictions for the next 30 days, i.e., one month ahead of forecasting, and use RMSE as a metric to assess the predictive capability of the derived forecasts.

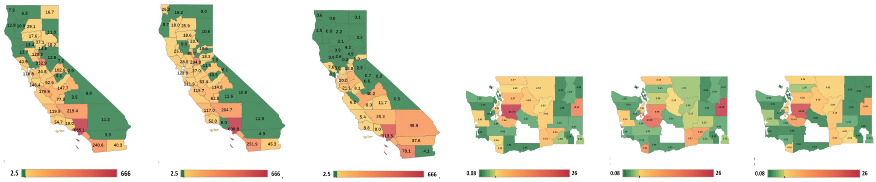
To verify the value added by the topological descriptors, we perform predictions on three different input-data-cases: a) using only the number of confirmed cases (LSTM), b) using the number of confirmed cases plus environmental variables (LSTM+EV), and c) using the number of confirmed cases plus topological summaries (TLife-LSTM). Notice that case (a) corresponds to the traditional application of LSTM on COVID-19 data as in literature, and case (b) is equal to use LSTM with environmental variables. As Fig. 4 show, forecasting performance of new daily COVID-19 cases at a county level tend to benefit from integrating topological summaries of atmospheric conditions. Similarly, the topological features extracted from environmental dynamic networks also tend to enhance the

Table 1. RMSE Results in California (CA) and Washington (WA). Performance comparison.

Statistics	LSTM **Cases**	LSTM+EV **Cases**	TLife-LSTM **Cases**	LSTM **Hospi**	LSTM+EV **Hospi**	TLife-LSTM **Hospi**
CA (means)	64.5662	58.7211	60.5828	13.8351	19.3277	12.1730
CA (freq)	16	22	21	16	9	31
WA (means)	17.2531	11.4251	17.10648	3.1329	2.4663	3.0746
WA (freq)	4	32	4	11	15	11

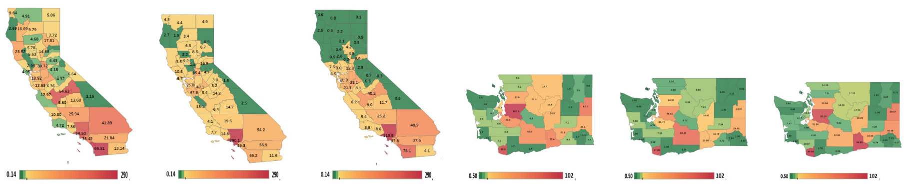
3 Available at https://github.com/CSSEGISandData/COVID-19.
4 Available at https://midasnetwork.us/covid-19/.

forecasts of the hospitalization, as shown in Fig. 5. Table 1 presents the summary of the prediction results in CA and WA at county level. The 1st and 3rd row show the means of RMSE for all counties, while the 2nd and 4th rows show the numbers of counties for which each method delivers the lowest RMSE. The columns represent the input-data-cases. The first 3 columns are focusing on the cases prediction and the other are results from hospitalization prediction. As Table 1 suggests, integrating environmental information tends on average to improve the forecasting performance of the number of cases and hospitalizations in both states. While we tend to believe that the impact of environmental variables on COVID-19 transmission is likely to be only indirect, the impact of environmental information on future hospitalization appears to be much more profound. This phenomenon is likely to be linked to the connection of atmospheric variables and multiple pre-existing health conditions, e.g., respiratory and cardiovascular diseases, that result in elevating risks of COVID-19 severity. As such, the important future extensions of this analysis include integration of various sociodemographic variables (e.g., population, health conditions, and median age) into TLife-LSTM.



(a) (b) (c) (d) (e) (f)

Fig. 4. RMSE of new cases prediction for each county in California(CA) and Washington(WA). (a) LSTM: Cases CA. (b) LSTM+EV: Cases CA. (c) TLife-LSTM: Cases CA. (d) LSTM: Cases WA. (e) LSTM+EV: Cases WA. (f)TLife-LSTM: Cases WA.



(a) (b) (c) (d) (e) (f)

Fig. 5. RMSE of hospitalization prediction for each county in California (CA) and Washington (WA). (a) LSTM: Cases CA. (b) LSTM+EV: Hospi CA. (c) TLife-LSTM: Hospi CA. (d) LSTM: Hospi WA. (e) LSTM+EV: Hospi WA. (f) TLife-LSTM: Hospi WA.

7 Conclusions and Future Scope

We have developed a new topological machine learning model TLife-LSTM to forecast COVID-19 dynamics, with a focus on using environmental and topological features along with time series of the historical COVID-19 records. Our experiments have indicated that TLife-LSTM yields more accurate forecasts of future number of cases and hospitalizations rates in the states of California and Washington on a county-level basis. As a result, TLife-LSTM indicates the impact of atmospheric conditions on COVID-19 progression and associated mortality; critical to efficiently respond to the current pandemic and forecast impeding hotspots and potential waves. In the future, we plan to introduce socio-demographic variables to our model and encode multi-parameter persistent characteristics of epidemiological and atmospheric conditions.

Acknowledgments. This work has been supported in part by grants NSF DMS 2027793 and NASA 20-RRNES20-0021. Huikyo Lee's research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004).

References

1. Alazab, M., Awajan, A., Mesleh, A., Abraham, A., Jatana, V., Alhyari, S.: COVID-19 prediction and detection using deep learning. Int. J. Comput. Inf. Syst. Ind. Manage. Appl. **12**, 168–181 (2020)
2. de Ángel Solá, D.E., Wang, L., Vázquez, M., Méndez Lázaro, P.A.: Weathering the pandemic: how the Caribbean Basin can use viral and environmental patterns to predict, prepare and respond to COVID-19. J. Med. Virol. **92**(9), pp. 1460–1468 (2020)
3. Arora, P., Kumar, H., Panigrahi, B.: Prediction and analysis of COVID-19 positive cases using deep learning models: a descriptive case study of India. Chaos Solitons Fractals **139**, 110017 (2020). https://doi.org/10.1016/j.chaos.2020.110017
4. Berumen, J., et al.: Trends of SARS-Cov-2 infection in 67 countries: role of climate zone, temperature, humidity and curve behavior of cumulative frequency on duplication time. medRxiv (2020)
5. Bouhamed, H.: COVID-19 cases and recovery previsions with Deep Learning nested sequence prediction models with Long Short-Term Memory (LSTM) architecture. Int. J. Sci. Res. Comput. Sci. Eng. **8**, 10–15 (2020)
6. Carlsson, G.: Topology and data. BAMS **46**(2), 255–308 (2009)
7. Carlsson, G.: Persistent homology and applied homotopy theory. In: Handbook of Homotopy Theory. CRC Press, Boca Raton (2019)
8. Chazal, F., Michel, B.: An introduction to topological data analysis: fundamental and practical aspects for data scientists. arxiv:1710.04019 (2017)
9. Chen, J., Gao, K., Wang, R., Nguyen, D.D., Wei, G.W.: Review of COVID-19 antibody therapies. Annu. Rev. Biophys. **50** (2020)
10. Chen, Y., Volic, I.: Topological data analysis model for the spread of the coronavirus. arXiv:2008.05989 (2020)
11. Costa, J.P., Škraba, P.: A topological data analysis approach to epidemiology. In: European Conference of Complexity Science (2014)

12. Dlotko, P., Rudkin, S.: Visualising the evolution of English COVID-19 cases with topological data analysis ball mapper. arXiv:2004.03282 (2020)
13. Dong, E., Du, H., Gardner, L.: An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect. Dis. **20**(5), 533–534 (2020). https://doi.org/10.1016/S1473-3099(20)30120-1
14. Edelsbrunner, H., Harer, J.: Persistent homology - a survey. Contemp. Math. **453**, 257–282 (2008)
15. Falk, M., et al.: Topological data analysis made easy with the topology toolkit, what is new? (2020)
16. Franch-Pardo, I., Napoletano, B.M., Rosete-Verges, F., Billa, L.: Spatial analysis and GIS in the study of COVID-19. A review. Sci. Total Environ. **739**, 140033 (2020)
17. Greff, K., Srivastava, R.K., Koutník, J., Steunebrink, B.R., Schmidhuber, J.: LSTM: a search space odyssey. IEEE Trans. Neural Netw. Learn. Syst. **28**(10), 2222–2232 (2017). https://doi.org/10.1109/TNNLS.2016.2582924
18. Johnson, L., Schieberl, L.: Topological visualization of COVID-19 spread in California, Florida, and New York (2020)
19. Lo, D., Park, B.: Modeling the spread of the Zika virus using topological data analysis. PLoS One **13**(2), e0192120 (2018)
20. Metcalf, C.J.E., et al.: Identifying climate drivers of infectious disease dynamics: recent advances and challenges ahead. Proc. R. Soc. B Biol. Sci. **284**(1860), 20170901 (2017)
21. Otter, N., Porter, M.A., Tillmann, U., Grindrod, P., Harrington, H.A.: A roadmap for the computation of persistent homology. EPJ Data Sci. **6**(1), 1–38 (2017). https://doi.org/10.1140/epjds/s13688-017-0109-5
22. Ramchandani, A., Fan, C., Mostafavi, A.: DeepCOVIDNet: an interpretable deep learning model for predictive surveillance of COVID-19 using heterogeneous features and their interactions. IEEE Access **8**, 159915–159930 (2020). https://doi.org/10.1109/ACCESS.2020.3019989
23. Rouen, A., Adda, J., Roy, O., Rogers, E., Lévy, P.: COVID-19: relationship between atmospheric temperature and daily new cases growth rate. Epidemiol. Infect. **148** (2020)
24. Shahid, F., Zameer, A.: Predictions for COVID-19 with deep learning models of LSTM, GRU, and Bi-LSTM. Chaos, Solitons Fractals **140**, 110212 (2020)
25. Soliman, M., Lyubchich, V., Gel, Y.: Ensemble forecasting of the Zika space-time spread with topological data analysis. Environmetrics **31**(7), e2629 (2020). https://doi.org/10.1002/env.2629
26. Yu, Y., Si, X., Hu, C., Zhang, J.: A review of recurrent neural networks: LSTM cells and network architectures. Neural Comput. **31**(7), 1235–1270 (2019)
27. Zeroual, A., Harrou, F., Abdelkader, D., Sun, Y.: Deep learning methods for forecasting COVID-19 time-series data: a comparative study. Chaos Solitons Fractals **140**, 110121 (2020). https://doi.org/10.1016/j.chaos.2020.110121