

– Supplementary Material –

NASA Satellite Data: Uncovering Hidden Patterns in COVID-19 Clinical Severity and Beyond

Anonymous submission

Further Details of the collected **NASAdat** Dataset

Data Preprocessing and Format AOD is a measure of the amount of light that atmospheric aerosols scatter and absorb and a monotonic function of air quality related to particulate matter near the ground. We generated daily climatology of AOD using the 19-year observations between January 1st, 2001 and December 31st, 2019 (Figure 1 (a)) and used the climatological AOD in the team's previous studies (Segovia-Dominguez et al. 2021a,b). To calculate a climatological mean for each day of the year, we average 19 observations between January 1st, 2001 and December 31st, 2019. For example, the climatological AOD on January 1st is an average of the 19 New Year's days from 2003 through 2019.

We also provide data on daily climatology of surface air temperature and RH from the Atmospheric InfraRed Sounder (Aumann et al. 2003) as shown in Figures 1 (b) and (c). To fully take advantage of its high spatial resolution, we use surface air temperature and relative humidity from AIRS and CrIs in years 2020 and 2021. For example, GDL models can use topological summaries of the Community Long-term Infrared Microwave Coupled Atmospheric Product System (CLIMCAPS) products as input. The underlying hypothesis to be tested over the next three years is that surface air temperature and RH may affect COVID-19 hospitalization and death indirectly.

The collected dataset include a unique identifier for each county and is saved in the netCDF format. *NASAdat* can be accessed via:

Temperature

DOI: 10.48577/jpl.z31y-2r10

<https://doi.org/10.48577/jpl.z31y-2r10>

Metadata (url)

<https://commons.datacite.org/doi.org/10.48577/jpl.z31y-2r10>

Relative Humidity

DOI: 10.48577/jpl.ws86-1q81

<https://doi.org/10.48577/jpl.ws86-1q81>

Metadata (url)

<https://commons.datacite.org/doi.org/10.48577/jpl.ws86-1q81>

AOD

DOI: 10.48577/jpl.k37v-y751
<https://doi.org/10.48577/jpl.k37v-y751>
Metadata (url)
<https://commons.datacite.org/doi.org/10.48577/jpl.k37v-y751>

By including the Federal Information Processing Standard code (FIPS) of each county, now NASA's atmospheric data in *NASAdat* is easily matched with county level datasets from other public and private entities.

Uniqueness The collected *NASAdat* dataset is unique in multiple aspects. First, long-term AOD observations from a single instrument over the entire CONUS, such as our *NASAdat*, is only available from satellites. While AOD observations are also available from NASA's remote sensing Aerosol Robotic Network (AERONET) stations, AERONET coverage is noticeably sparser. In turn, many previous studies which compare AOD observations from MODIS with those from AERONET report reasonable agreement between the two, which also can serve as an additional measure of data quality control. Second, while NOAA through NCEI provides data on such weather variables as temperature, precipitation, drew point, visibility, etc. Almost all of NOAA's records rely on ground-based stations. As a result, in contrast to *NASAdat*, the NOAA data are limited to the resolution on covered areas across U.S., and many counties are far away from land-based stations which further increases uncertainty in applications requiring better resolution, such as biosurveillance. Third, in comparison to all other existing data, our daily climatologies of temperature and relative humidity provide annual cycles in these variables for each county with the Federal Information Processing Standard Publication 6-4 (FIPS 6-4) code, thereby making it easier to match *NASAdat* with various key biosurveillance, socio-economic and socio-demographic information of the best available granularity (i.e., at a county level) such as COVID-19 hospitalizations, cancer rates, and number of houses with solar panels. Fourth, temperature and relative humidity data for the entire globe including those over ocean are another benefit of using satellite observations when running ML models for different spatial domains other than the US. Fifth, the climatology datasets such as *NASAdat* can be used to study the impacts of the nation's climate change on various

sectors, from digital agriculture to resilience of critical infrastructures to adverse climate events. Moreover, given multiple types of ground truth instances associated with these data, e.g., dust storms and teleconnection patterns, the presented benchmark NASAdat can serve as a test bed for a very broad range of ML tasks such as spatio-temporal forecasting with graph neural networks, transfer learning of climatic scenarios, dynamic clustering, anomaly detection, and multi-resolution pattern matching.

Quality of the dataset NASAdat undergoes standard data quality control checks under NASA guidelines. The original datasets were generated by averaging quality-controlled observations. As a part of retrieval algorithms, a quality flag is automatically assigned to each retrieved value of temperature, relative humidity, and AOD. The algorithms assign a quality flag of each pixel by comparing the observed values with predefined ranges of valid observations. A quality flag is a kind of automated annotation by a machine that is already considered in the original datasets. As such, we were confident about the quality of our newly generated datasets. Due to low-quality retrievals, there exists a small fraction of missing values in the original datasets. As per the standard statistical practice, these missing values are stripped when calculating a spatial and temporal average for each county.

Both MODIS and AIRS missions provide more detailed information on the quality flag.

MODIS

https://atmosphere-imager.gsfc.nasa.gov/sites/default/files/ModAtmo/documents/QA_Plan_C61_Master_2021_09_22.pdf

AIRS

https://docserver.gesdisc.eosdis.nasa.gov/public/project/AIRS/V7_L2_Quality_Control_and_Error_Estimation.pdf

Both AIRS and MODIS datasets cover the entire globe. The total sizes of 6209 AIRS and and 6939 MODIS files are about 2.5 and 4.2 gigabytes respectively. In our processed data, each file for temperature, relative humidity, or AOD has a size of 95 MB.

Maintenance Plan Our previous work (Lee et al. 2018) indicates that even 19 years (2001-2019) may not be long enough to define statistically stable AOD climatology. Also, we recognize that continuous updates are the key for these data utilities, especially for biosurveillance and other time sensitive applications. JPL NASA/Caltech will update our datasets 2 times per year and also whenever new versions of the NASA products are released through NASA's Distributed Active Archive Centers (DAACs).

In our maintenance plan we are taking advantage from the fact that these benchmark data are one of the first projects within the most recent broader NASA's JPL initiative on hosting datasets, such as these and assigning DOIs so there is persistence for papers, and also capturing the raw and any derived results. As such, JPL will continue updating and maintaining these benchmark data under this broader NASA's initiative, with external access to a hub under the

subdomain of jpl.nasa.gov. Our team will keep producing daily temperature, relative humidity, and AOD datasets from AIRS/CrIS and MODIS/VIIRS in a NetCDF format which can serve as input for multiple projects across the ML and atmospheric sciences communities. To take full advantage of the highest spatial resolution, we plan to expand and use level 2 surface air temperature and relative humidity from AIRS and CrIs of next years. With the combination of using NASA front-end servers, NVIDIA DGX clusters at the NASA Center for Climate Simulation, and parallel processing capabilities and elastic scalability of the Advanced Data Analytics Platform (ADAPT) science cloud, we expect to have no issue maintaining our data for years to come as these services will provide us all the resources necessary with no cost to NASAdat end-users.

Further Experimental Results

Benchmarking neural network models We benchmark two broad classes of neural networks (i) Recurrent Neural Networks (RNNs): Long Short-Term Memory (LSTM) (Hochreiter and Schmidhuber 1997) can forecast univariate time series with LSTM hidden units; (ii) Spatio-Temporal Graph Convolutional Networks: spatio-temporal model with the framework of graph convolutional network (GCN) exploit GCN and temporal convolution to capture dynamic spatial and temporal patterns and correlations; we report performances for eight types of state-of-the-arts methods on our benchmark datasets including (1) Diffusion Convolutional Recurrent Neural Network (DCRNN) (Li et al. 2018): diffusion convolution recurrent neural network that captures both spatial and temporal dependencies through random walks on graph and encoder-decoder architecture for multi-step forecasting, (2) Long Short-Term Memory R-GCN (LRGCN) (Li et al. 2019): time-evolving neural network which integrates relational GCN (R-GCN) into the LSTM to fully investigate both intra-time and inter-time relations, (3) Attention Temporal Graph Convolutional Network (A3T-GCN) (Bai et al. 2021): an attention temporal GCN that combines GCNs and GRUs with attention mechanism which can capture both spatio-temporal dependencies and global variation trends; (4) Message Passing Neural Networks with LSTM (MPNN+LSTM) (Panagopoulos, Nikolentzos, and Vazirgiannis 2021): a time-series version of message passing neural networks consists of a series of neighborhood aggregation layers to model in detail the dynamics of the spreading process; (5) Evolving Graph Convolutional Networks (EvolveGCNO and EvolveGCNH) (Pareja et al. 2019): evolving graph convolutional network that utilizes the recurrent model to update the trainable parameters of GCN for understanding and forecasting graph structure dynamics; (6) Graph Convolutional Recurrent Network (GconvLSTM) (Seo et al. 2018): graph convolutional recurrent network model which replaces convolution by graph convolution to extract the spatial-temporal information; (7) Gated Graph Neural Networks for Dynamic Graphs (DyGrEncoder) (Taheri, Gimpel, and Berger-Wolf 2019): gated graph neural networks for dynamic graphs which uses a gated graph neural network equipped with standard LSTM

for dynamic graph classification.

Under/Over Prediction analysis Table 1 shows fraction of days the DL and GDL models deliver COVID-19 hospitalization forecasts which are higher than the true records (i.e., over-predict). We find that under all considered scenarios the DL and GDL models, with or without atmospheric variables, largely tend to under-forecast COVID-19 related hospitalizations. Such phenomena may be due to inherent model bias and can be addressed by developing ensemble approaches allowing us to more systematically quantify predictive uncertainty. Under-forecasting is particularly prominent for PA and TX. While under- and over-forecasting have its own limitations, it is especially important to account for such under-forecasting phenomena in tasks related to allocation of healthcare resources such ICU beds.

Time complexity Table 2 shows the average running time and standard error of all models on CA, PA, and TX.

Percentage of improvements To further illustrate the benefits of including the atmospheric factors into GCN-based models for COVID-19 hospitalizations analysis, we also present percentages of counties in each state, where the DL predictive performance has been improved upon adding (one) atmospheric variable as predictor (see Table 3). As Table 3 suggests, compared to Temp and RH, most GCN-based models with AOD tend to deliver noticeably higher numbers of counties where the DL model performance improvement is recorded. These findings also echo the results in Table 1 and reconfirm the earlier hypothesis that AOD is an important factor for understanding COVID-19 clinical severity.

Why RMSE? In our experiments we use the RMSE metric rather than R^2 since RMSE is the standard metric for validation of predictive models in space-time forecasting (Brockwell et al. 1991). Despite statistical criticism, R^2 is still used in epidemiology. As such, we present a summary of results for R^2 . While we find that R^2 for actual observations and hospitalization forecasts with/without AOD are generally similar in CA, in TX and PA R^2 for GCNs with AOD tends to be from 0.05 to 0.25 higher than R^2 for the same GCN but *without* AOD, with ranges from 0.6 to 0.88 in PA and from 0.71 to 0.93 in TX. These findings echo our conclusions on contributions of AOD to COVID-19 clinical severity, based on predictive RMSE.

Why Not Regression Models? Furthermore, we do not consider simpler models, such as regression, ARIMA and other Box-Jenkins class of models, because such tools focus only on linear relationships between variables and, as a result, cannot capture nonlinear nonseparable spatio-temporal dependencies of COVID-19 dynamics (and many other infectious diseases with high virulence). In turn, our analysis includes a broad range of DL architectures that allow us to address such nonlinear dependencies. Furthermore, the model consensus analysis presented in our paper enables us to address such pressing question as whether a relative risk to be affected by COVID-19 is higher for some areas due to

their higher exposure to poor air quality.

Addition of Social Variables

IGN: Working on it... Adding socioeconomic variables. We include... network structure based on.... boundary.... include additional variable... i.e. AOD + SOCIAL.... MAPS in ...

Further comments on experimental settings We do not explicitly incorporate social variables *by purpose*. First, it is questionable which variables and how impact COVID-19 clinical severity (Hu et al. 2021). Second, even if someone smokes (potentially a negative factor), does it mean that smokers in polluted and unpolluted areas have the same COVID-19 prognosis? To address such questions, we focus on assessing conditional predictive utility of atmospheric variables, given that socio-economic, socio-demographic, social-mobility etc factors are fixed. Third, as noted by (Sesé et al. 2021; Persico 2020), more polluted areas tend to be populated by economically disadvantaged groups, thereby further increasing unfairness in healthcare outcomes, and it is of critical importance to account for such predisposition in a systematic manner such that the contribution of a single environmental risk factor is analyzed, conditionally on all other factors being fixed.

IGN: Since now we have added social variable in appendix, this remark is outdated. **Remark 3.** Note that we intentionally do not incorporate any social variables into analysis of COVID-19 clinical severity. First, it is questionable which variables and how impact COVID-19 clinical severity [3]. Second, our primary focus is to assess conditional predictive utility of atmospheric variables, given that socio-economic, socio-demographic, social-mobility etc factors are fixed. Third, as noted by [1, 2], more polluted areas tend to be populated by economically disadvantaged groups, thereby further increasing unfairness in healthcare outcomes, and it is of critical importance to account for such predisposition in a systematic manner such that the contribution of a single environmental risk factor is analyzed, conditionally on all other factors being fixed.

Model Voting with Temperature and Relative Humidity

Global Impact and Applications Beyond Bio-surveillance

NASA's satellite observations such as NASAdat have played an important role in the international assessment reports on climate change that provide scientific resources for understanding human-induced climate change and assessing impacts crucial for informing policy. As such, the climatology datasets such as NASAdat can be used to study the impacts of the nation's climate change on various sectors, from digital agriculture to resilience of critical infrastructures to adverse climate events. Moreover, given multiple types of ground truth instances associated with these data, e.g., dust storms and teleconnection patterns, the presented benchmark NASAdat can serve as a test bed for

Table 1: Fraction of days (in %) the DL models deliver forecasts higher than the true COVID-19 hospitalization numbers (i.e., over-predict) in three U.S. states: CA, PA, and TX. Forecasting horizon is 15-day ahead. Results (RMSE \pm s.d.) are averaged over 10 runs with different seeds.

Model	CA				PA				TX			
	Baseline	AOD	Temp	RH	Baseline	AOD	Temp	RH	Baseline	AOD	Temp	RH
LSTM (Hochreiter and Schmidhuber 1997)	23.90 \pm 3.70	23.20 \pm 2.82	19.72 \pm 1.40	18.70 \pm 1.37	5.50 \pm 0.63	9.14 \pm 1.63	9.71 \pm 1.24	7.43 \pm 2.84	30.86 \pm 1.23	28.46 \pm 6.86	27.51 \pm 2.78	17.12 \pm 3.07
DCRNN (Li et al. 2018)	19.33 \pm 0.69	28.38 \pm 2.85	20.70 \pm 0.70	15.17 \pm 0.42	4.41 \pm 0.33	5.34 \pm 0.74	5.23 \pm 0.55	3.64 \pm 0.70	24.45 \pm 0.49	6.86 \pm 1.54	28.02 \pm 3.95	10.27 \pm 0.44
LRGCN (Li et al. 2019)	24.36 \pm 4.95	19.37 \pm 2.62	20.49 \pm 1.51	19.25 \pm 1.22	3.70 \pm 0.25	1.27 \pm 0.92	2.76 \pm 0.47	3.06 \pm 0.99	30.20 \pm 1.60	1.27 \pm 1.04	22.73 \pm 5.28	16.48 \pm 3.00
AT3-GCN (Bai et al. 2021)	28.20 \pm 0.63	18.97 \pm 0.56	20.23 \pm 0.88	19.81 \pm 0.53	11.05 \pm 0.95	11.69 \pm 0.59	10.51 \pm 1.22	10.28 \pm 0.63	24.65 \pm 0.68	22.92 \pm 2.04	7.82 \pm 1.41	26.07 \pm 1.10
MPNN+LSTM (Panagopoulos, Nikolentzos, and Vaziriannis 2021)	27.19 \pm 1.27	25.36 \pm 0.88	26.75 \pm 2.45	13.59 \pm 1.84	5.77 \pm 0.12	6.91 \pm 0.42	5.05 \pm 1.37	8.47 \pm 0.73	13.82 \pm 0.69	8.48 \pm 3.13	7.81 \pm 2.64	11.46 \pm 1.37
EvoLveGCNO (Pareja et al. 2019)	23.05 \pm 0.69	16.63 \pm 0.55	20.13 \pm 0.83	20.89 \pm 0.47	14.39 \pm 4.90	5.55 \pm 0.32	8.88 \pm 1.32	21.57 \pm 4.38	9.31 \pm 0.76	2.60 \pm 0.14	4.03 \pm 0.61	12.57 \pm 1.17
EvoLveGCNH (Pareja et al. 2019)	22.22 \pm 1.20	20.19 \pm 1.89	21.61 \pm 1.77	21.90 \pm 0.72	11.06 \pm 5.09	9.43 \pm 6.24	12.06 \pm 5.38	30.65 \pm 15.14	10.57 \pm 3.60	4.42 \pm 4.75	9.38 \pm 13.63	14.54 \pm 5.54
GconvLSTM (Seo et al. 2018)	24.64 \pm 3.29	23.47 \pm 1.28	18.84 \pm 2.65	19.77 \pm 1.30	8.68 \pm 1.31	10.92 \pm 0.93	11.66 \pm 0.91	9.96 \pm 0.77	20.69 \pm 3.26	33.39 \pm 2.77	21.45 \pm 4.53	17.13 \pm 6.22
DyGrEncoder (Taheri, Gimpel, and Berger-Wolf 2019)	24.14 \pm 8.06	16.45 \pm 2.63	19.74 \pm 1.44	15.60 \pm 1.32	6.05 \pm 0.70	4.49 \pm 2.56	3.93 \pm 1.88	4.18 \pm 1.63	29.14 \pm 1.45	5.25 \pm 5.25	10.65 \pm 5.85	24.32 \pm 3.05

Table 2: Time complexity in seconds, mean and std, for each DL and GDL models in U.S. states: CA, PA, TX.

Model	CA			PA			TX		
	AOD	Temp	RH	AOD	Temp	RH	AOD	Temp	RH
LSTM (Hochreiter and Schmidhuber 1997)	192.40 \pm 0.44	197.40 \pm 0.94	196.93 \pm 0.57	201.66 \pm 0.55	206.47 \pm 1.02	207.50 \pm 1.13	576.75 \pm 2.96	585.27 \pm 3.35	580.16 \pm 2.70
DCRNN (Li et al. 2018)	412.44 \pm 3.80	287.70 \pm 1.55	286.80 \pm 3.19	427.46 \pm 1.16	295.72 \pm 1.82	296.23 \pm 2.98	1250.86 \pm 10.65	613.05 \pm 1.16	847.26 \pm 7.30
LRGCN (Li et al. 2019)	707.08 \pm 16.32	667.43 \pm 16.46	521.15 \pm 11.53	668.91 \pm 7.99	677.56 \pm 11.91	577.46 \pm 10.47	1286.92 \pm 15.31	1295.01 \pm 8.39	861.55 \pm 40.12
AT3-GCN (Bai et al. 2021)	704.70 \pm 16.50	738.78 \pm 59.67	728.41 \pm 2.77	753.81 \pm 23.40	752.51 \pm 7.51	809.73 \pm 129.26	3343.85 \pm 31.39	3154.69 \pm 141.87	2949.90 \pm 31.44
MPNN+LSTM (Panagopoulos, Nikolentzos, and Vaziriannis 2021)	713.54 \pm 4.11	959.21 \pm 10.67	960.71 \pm 11.14	949.18 \pm 3.82	994.15 \pm 9.35	931.53 \pm 24.72	3040.30 \pm 35.26	3124.62 \pm 74.75	3894.68 \pm 51.19
EvoLveGCNO (Pareja et al. 2019)	127.37 \pm 0.41	128.40 \pm 2.73	131.35 \pm 0.77	128.39 \pm 0.44	128.10 \pm 0.39	131.14 \pm 1.61	213.53 \pm 0.33	217.54 \pm 3.38	220.19 \pm 1.31
EvoLveGCNH (Pareja et al. 2019)	227.10 \pm 2.03	224.09 \pm 2.10	223.81 \pm 2.14	228.85 \pm 2.28	228.38 \pm 1.61	230.25 \pm 3.49	348.20 \pm 4.40	344.20 \pm 2.34	342.21 \pm 2.50
GconvLSTM (Seo et al. 2018)	627.27 \pm 2.06	625.42 \pm 0.62	618.33 \pm 0.92	640.47 \pm 1.20	637.65 \pm 1.06	630.42 \pm 1.02	1188.53 \pm 1.74	1189.27 \pm 5.56	1183.04 \pm 1.08
DyGrEncoder (Taheri, Gimpel, and Berger-Wolf 2019)	367.77 \pm 25.47	353.97 \pm 5.66	352.04 \pm 5.29	103.16 \pm 5.83	95.27 \pm 3.77	102.46 \pm 7.73	73.03 \pm 10.52	65.36 \pm 7.97	70.69 \pm 4.10

a very broad range of ML tasks such as spatio-temporal forecasting with graph neural networks, transfer learning of climatic scenarios, dynamic clustering, anomaly detection, and multi-resolution pattern matching. For example, scientists can explore transferability of the developed topological and geometric machine learning approaches to monitoring and forecasting spread and clinical severity of (re)emerging SARS- and MERS-associated viruses and arboviral diseases such as Zika. Alternatively, we can study how transferable is the response of the Dallas power grid network to Texas’ winter storms to other adverse climate events and/or across power grid networks in other US cities.

Although the current NASAdat benchmark is limited to the CONUS, it is important to take advantage of a global coverage in the observations from polar-orbiting satellites. Prior to application of NASAdat worldwide, it is also important to test the data over the CONUS where ground-based observations of temperature and relative humidity observations are available from NOAA. Not surprisingly, the comparison between the AIRS observations and those from NOAA’s network would show some difference. However, based on the previous studies conducted at JPL, we are confident that there is almost negligible difference in annual cycles of these two variables.

The biggest advantage of using satellite observations is their wide coverage over the entire globe. We already prepared the same datasets of temperature, relative humidity, and AOD averaged for each country, but have not applied to them to model COVID-19 clinical severity outside the United States yet. Such COVID-19 biosurveillance analysis using GDL and other DL applied to the worldwide dataset will be our future work. Furthermore, we have experimented with multi-resolution pattern matching using topological ML in application to the worldwide dataset (Ofori-Boateng et al. 2021), and we think that such multi-resolution pattern matching will be of interest not only in environmental sciences but in broader problems of image processing and

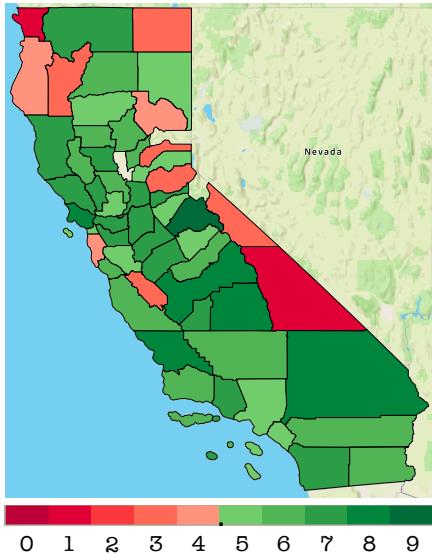
computer vision.

Acknowledgements

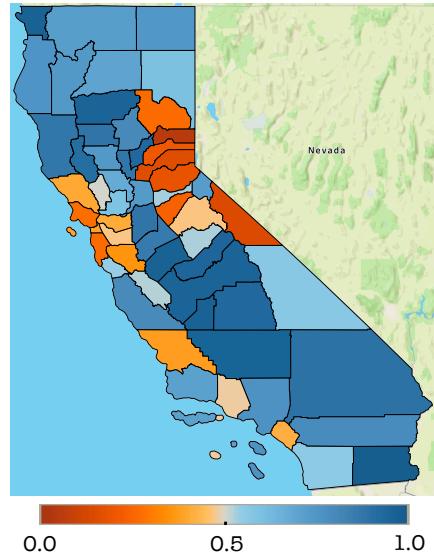
The project has beted by NASA grant 20-RRNES20-0021 under the Rapid Response and Novel Research in Earth Science, the UTSystem-CONACYT ConTex program, and NSF ATD grant DMS 1925346.

Table 3: Fraction of counties (in %) in three US states: CA, PA, and TX where the forecasting performance for COVID-19 hospitalizations has been improved upon adding atmospheric variables from the proposed NASAdat dataset into DL models. Highest fractions for each state are in bold.

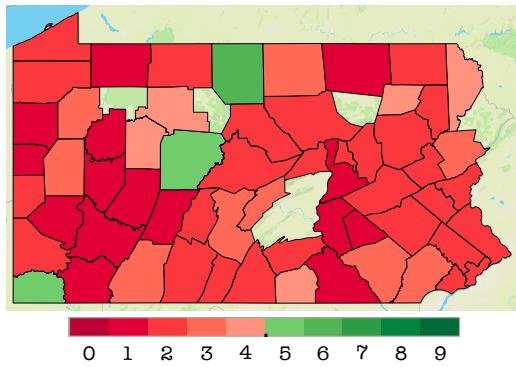
Model	CA			PA			TX		
	AOD	Temp	RH	AOD	Temp	RH	AOD	Temp	RH
LSTM (Hochreiter and Schmidhuber 1997)	5.45	5.45	5.45	88.33	90.00	78.33	82.47	33.86	47.41
DCRNN (Li et al. 2018)	92.73	1.82	3.64	23.33	33.33	26.67	96.41	10.36	4.38
LRGCN (Li et al. 2019)	9.09	5.45	5.45	6.67	58.33	65.00	65.74	31.08	75.70
AT3-GCN (Bai et al. 2021)	7.27	5.45	5.45	1.67	1.67	1.67	71.71	73.31	59.36
MPNN+LSTM (Panagopoulos, Nikolentzos, and Vaziriannis 2021)	92.73	10.91	5.45	98.33	75.00	90.00	41.43	42.23	10.36
EvolveGCNO (Pareja et al. 2019)	7.27	7.27	5.45	6.67	11.67	95.00	3.19	3.19	86.06
EvolveGCNH (Pareja et al. 2019)	5.45	3.64	30.91	0.00	48.33	61.67	3.19	1.20	47.41
GconvLSTM (Seo et al. 2018)	9.09	5.45	5.45	53.33	73.33	35.00	64.54	15.54	64.94
DyGrEncoder (Taheri, Gimpel, and Berger-Wolf 2019)	10.91	12.73	10.91	91.67	95.00	80.00	45.02	49.40	26.29



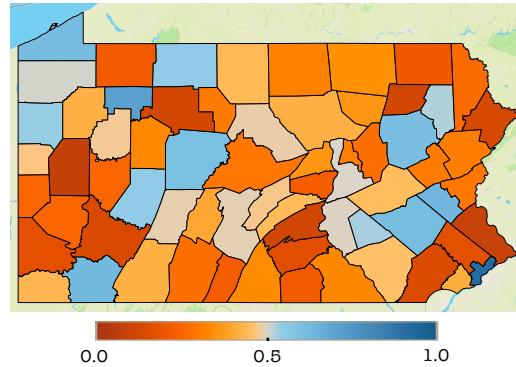
(a) Model Voting with Social Vulnerability Index for CA



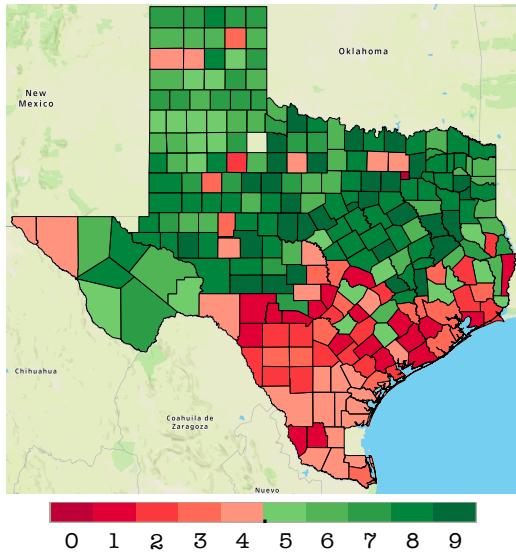
(b) Social Vulnerability Index Map for CA



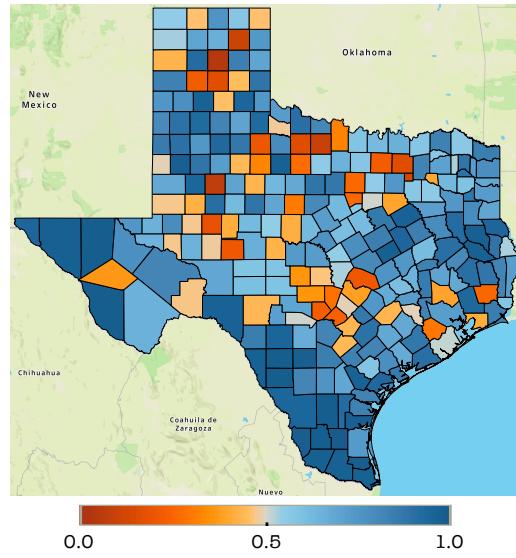
(c) Model Voting with Social Vulnerability Index for PA



(d) Social Vulnerability Index Map for PA



(e) Model Voting with Social Vulnerability Index for TX



(f) Social Vulnerability Index Map for TX

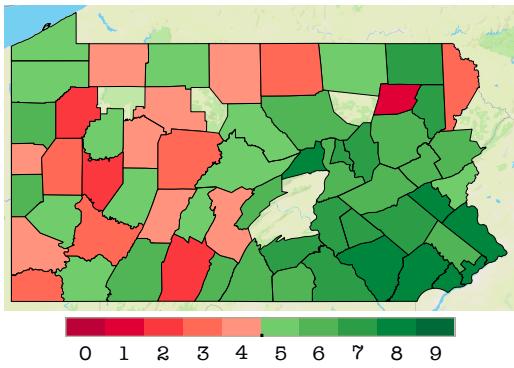
Figure 1: Spatial distribution of the consensus among DL models which find that Social Vulnerability Index exhibits predictive utility for COVID-19 clinical severity (a, c, e) and Social Vulnerability Index maps (b, d, f) at the county level (i.e., number of model votes agreeing that Social Vulnerability Index is a useful predictor for COVID-19 hospitalizations, see Table 3 in Appendix B).



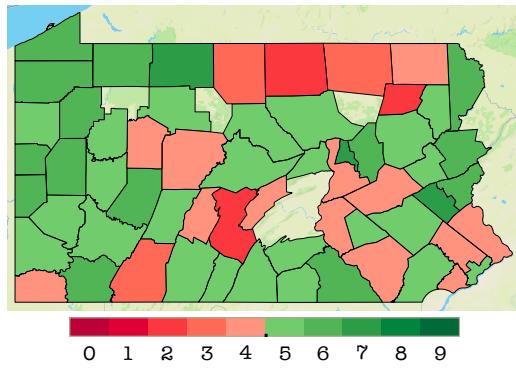
(a) Model Voting with RH for CA



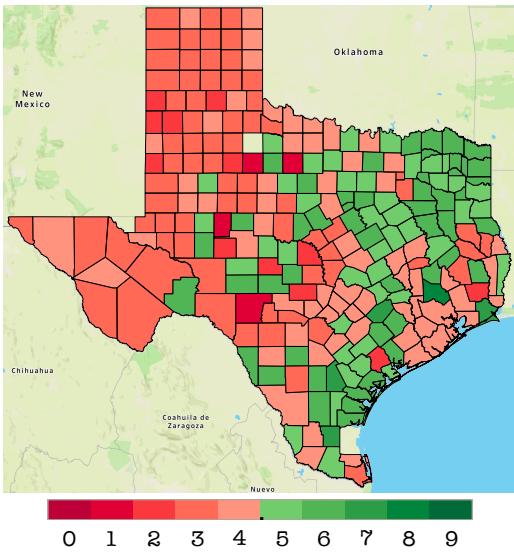
(b) Model Voting with TEMP for CA



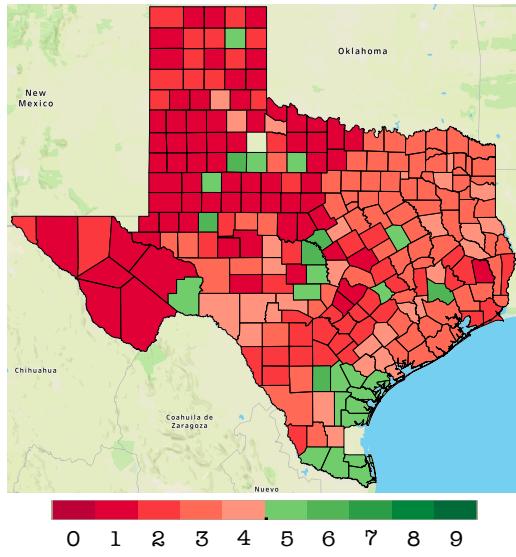
(c) Model Voting with RH for PA



(d) Model Voting with TEMP for PA



(e) Model Voting with RH for TX



(f) Model Voting with TEMP for TX

Figure 2: Spatial distribution of the consensus among DL models which find that RH and TEMP exhibits predictive utility for COVID-19 clinical severity (RH: a, c, e; TEMP: b, d, f) at the county level (i.e., number of model votes agreeing when RH and TEMP might be a useful predictor for COVID-19 hospitalizations, see Table 3 in Appendix B).