COVID-19 Forecasting using Web-Based and Participatory Survey Data

Kavya Kopparapu
Harvard University
Cambridge, MA, USA
kavyakopparapu@college.harvard.edu

Bryan Wilder
Harvard University
Cambridge, MA, USA
bwilder@g.harvard.edu

Milind Tambe
Harvard University
Cambridge, MA, USA
milind_tambe@harvard.edu

Maia Majumder
Boston Children's Hospital and Harvard Medical School
Boston, MA, USA
maimuna.majumder@childrens.harvard.edu

ABSTRACT

The COVID-19 pandemic has had a significant and unprecedented impact on the world since 2020. Concurrently, the rise of internet technology has led to the development of several significant data sources for forecasting the development of the pandemic, including but not limited to mobility data, crowdsourced symptoms, and search trends. While there has been significant algorithmic interest in developing forecasting *models*, previous work has not provided a systematic investigation of the relative utility of different *data sources* for COVID-19 forecasting. We present the first work comparing internet data sources for use in deep learning models to predict case incidence of COVID-19 across states the US. Our work affirms the relative utility of incidence and mobility data, which closely models the epidemiological interactions, compared to search trends and crowdsourced survey data, which encourage overfitting of deep learning models.

KEYWORDS

COVID19, Time Series, Deep Learning, LSTMs, Prediction

ACM Reference Format:
Kavya Kopparapu, Bryan Wilder, Milind Tambe, and Maia Majumder. 2021. COVID-19 Forecasting using Web-Based and Participatory Survey Data. In *Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), London, UK, May 3–7, 2021*, IFAAMAS, 7 pages.

1 INTRODUCTION

The World Health Organization declared the novel coronavirus disease (COVID-19) as a global pandemic on March 11th, 2020 [24]. As of February 25th, 2021, over 110 million cases have been reported globally, with over 28 million cases and 500,000 deaths in the United States (US). The economic impact of the pandemic has also disproportionately affected women and the socioeconomically disadvantaged [19]. Furthermore, the pandemic has exacerbated critical healthcare infrastructure in the US, especially evidenced by the shortage of personal protective equipment and available hospital ICU beds [25].

Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), U. Endriss, A. Nowé, F. Dignum, A. Lomuscio (eds.), May 3–7, 2021, London, UK.

Given these impacts, predicting the spread of epidemics is essential to mounting an appropriate public health response. A predictive model is only as good as its underlying data sources and so understanding the role played by different data sources is an important digital epidemiology question. More nuanced comparisons can help guide the development of predictive models, as well as inform priorities for the development of new data sources. In the past, forecasting has targeted seasonal diseases (such as influenza) or else more localized outbreaks (e.g., dengue, Ebola, or Zika). In response to these outbreaks, traditional forecasting utilizes epidemiological data – such as patient-level data or aggregated data regarding cases, hospitalizations, and deaths, from local authorities – to understand the spread of disease. In recent years, Internet-based digital epidemiological data in the form of social media posts, search engine queries, and online surveys has proven valuable for predicting the spread of epidemics [16, 27]. In particular, the historical and large volume of data available from rich data sources has spurred a rise in deep learning-based approaches to epidemiological forecasting [43].

However, predicting the spread of COVID-19 is fundamentally different than predicting other smaller-scale seasonal and local outbreaks due to its unprecedented impact on the operation of our society. While previous work has built a range of systems for predicting COVID-19 cases, to our knowledge, a systematic investigation of the promises and pitfalls of using different data sources in this forecasting task does not currently exist. In particular, COVID-19 has prompted both the usage of both existing digital data sources (e.g., search trends data [17]) as well as the introduction of new data sources such as mobility data collected from smartphones. While a large number of predictive models have been developed to forecast COVID-19 outbreaks, there has been little systematic effort to understand the extent to which these new and emerging data sources contribute to prediction specifically in the COVID-19 context.

A range of characteristics differentiate COVID-19 from the outbreak settings studied in earlier predictive work, potentially altering the role that different data sources may play:

- **Lack of historical data**: Deep learning models benefit from a significant volume of data, for example training on several historical influenza seasons. Due to the emerging nature of COVID-19, there is no historical data to supplement data

from the current pandemic. While, this limitation can be addressed by combining insights from several geographic regions [14, 17], it is nevertheless possible that the comparatively smaller amount of data available for COVID-19 could reduce the ability to learn complex models.
- **Prominent coverage and interest**: Unlike typical seasonal outbreaks or established diseases, developments about COVID-19 are well-reported by the media and followed by the public. Specific changes in public policy also significantly affect the spread of the disease, including mask mandates, social distancing, and indoor dining restrictions [46]. These public policy shifts are likely to register in local and national search trends as the public seeks more information about them. However, prominent media coverage also has significant downsides. For example, in the "worried well" phenomenon, excessive media coverage or misinformation about particular topics could result in healthy individuals over-reporting COVID-19 symptoms [36]. It is unclear whether digital data sources will become more or less reliable in this environment.
- **Stronger signals from participatory surveys**: The COVID-19 pandemic is of international importance, which may help overcome limitations in participation and selection bias previously documented with crowdsourcing symptoms of influenza-like illness (ILI) [34].
- **Novel Data Sources**: Newer data sources have emerged due the scale of the COVID-19 pandemic, including those pertaining to mobility, contact tracing, and social distancing metrics. Part of the objective of this study is to systematically evaluate the utility of these new signals.

In light of these opportunities and challenges, several data sources have emerged as candidates to inform the prediction of disease incidence on a localized level: Google Trends, participatory health surveys, mobility data, hospitalization information, and more [5, 13, 32]. Indeed, the COVID-19 Forecast Hub, which aggregates the predictions of over 20 leading models, shows a wide range of data sources in current use [7]. However, no previous work has provided a direct comparison of the utility of these data sources in COVID-19 forecasting. This study makes the following contributions:

- We systematically compare the success of a deep learning model at predicting COVID-19 outbreaks using a range of data sources representative of those in current use: incidence data supplemented by any of mobility data, search trends data, and crowd-sourced symptom reports. This allows us to assess the additional value provided by digital data sources above and beyond basic incidence reports.
- We find that the best-performing mode *uses only the incidence data.* Performance decreases with the inclusion of any of the digital data sources.
- We study potential explanations for this phenomenon. Examination of the train vs test performance of the models, alongside an investigation of feature importances, suggests that COVID-19 models are particularly susceptible to overfitting, potentially due to the limited amount of available training data. The inclusion of additional data sources can thus pose an unfavorable bias-variance tradeoff by increasing the number of parameters required in the model.

Overall, our results highlight the need for careful curation of the data sources used in COVID-19 forecasting, and for outbreak prediction more broadly. More data is not necessarily better, if additional data sources contain only marginal or noisy signal which the model can easily overfit to. Particularly when model simplicity is at a premium, it is necessary to develop a detailed understanding of the potential biases in input data sources (e.g., debiasing search trend data to account for the effects of media coverage [17]) since models can easily overfit to extraneous patterns in an input signal.

2 RELATED WORK

There has been significant work exploring the use of mechanistic, statistical, and deep learning models for predicting the spread of past epidemics. For example, Volkova et al. utilized long short-term memory (LSTMs) networks to predict ILI in military populations using social media communication and topics [41]. Venna et al. used similar methods but focused on geographic proximity and environmental variables as the data source [40]. Zou et al. utilized Google search data to predict weekly incidence of ILI [47]. Several authors use mechanistic models, especially susceptible-exposed-infected-recovered (SEIR) models, for their ability to represent the underlying transmission dynamics and interactions [23, 28].

Forecasting for COVID-19 is an incredibly active area of research, with a wide range of models in development. The COVID-19 Forecasting Hub [7] provides an overview of many of the models in current use. These models vary in the data sources used as input. However, one category of data source, introduced uniquely in response to COVID-19, is mobility data. Several studies have utilized the rise in access to this data to predict COVID-19 incidence, study dynamics of the pandemic, and understand the impact of public policy decisions [4, 12, 15, 42, 44].

More broadly, the common approach of incorporating several distinct data streams into a forecast for COVID-19 combines methodology explored in the time series prediction community. Long short-term memory networks have historically shown incredible promise in time-series prediction across a variety of tasks [10, 20, 21]. Several LSTM-based approaches have been implemented to forecast COVID-19 across different geographies, ground truths, and data streams [3, 30, 45]. Jin et al. leverages similar methods of learnable detrending to predict cases, deaths, and hospitalizations at the US state level from demographics, mobility indices, and health interventions [14].

Overall, previous work focuses on advancing model design and implementation when applied to epidemic forecasting, largely overlooking the question of the utility of data sources compared to each other under a fixed learning framework. One counterexample is work by Samaras et al., who explored the utility of search trends versus social media for predicting influenza in Greece [29]. However, explicit comparisons of the utility of data sources is a comparatively understudied topic in outbreak prediction which has been rendered only more important by the changes to the forecasting environment induced by COVID-19.

3 METHODS

3.1 Data Sources and Processing

Data utilized in this study spanned from March 30th, 2020 to September 20th, 2020, representing the activity of populations in the US at the state level. Ultimately, 144 features were utilized across four major data sources, as described below. Unnormalized and unprocessed, the magnitude of data considered varied considerably over time, reflecting the exponential growth of the pandemic.

3.1.1 Participatory Survey Data. Crowd-sourced data for monitoring the spread of infectious diseases has been recognized as a valuable resource as early as 2013 [6]. For this study, we utilized data collected through the Outbreaks Near Me platform (previously COVID Near You, abbreviated CNY), which has anonymously surveyed over 5 million individuals about their COVID symptoms and behavior as of February 25th, 2021 [1].

The CNY data was represented as a percent of the total population of each state, as per the official 2020 United States Census Bureau estimates, to contextualize the magnitude of the aggregated responses against the total patient volume of a state [2].

3.1.2 Google Health Trends. Online search data was obtained from Google Health Trends, a private API that provides daily absolute query volume of keywords across geographic regions. Keywords were chosen via exploration through Google Trends with the criteria of significant volume in the US during the dates of the study. Due to the magnitude of the data, values were logarithmized prior to training.

COVID-19: General keywords related to the pandemic represent public interest in the virus, including: "social distancing", "mask mandate", "covid testing center", and "trump covid".

Symptoms: Keywords in this category included "loss of smell and taste", "fever", "persistent cough", and other symptoms of COVID-19 as defined by the Centers for Disease Control and Prevention (CDC) guidelines [9].

COVID-19 progression: Keywords included combinations of COVID-19 symptoms according to likely disease progression according to work conducted by Larsen et al. [18]. Notably, the progression of COVID-19, typically beginning with a fever, differs from other respiratory tract infections and may provide insight to undiagnosed cases [18]. Examples include "fever then cough" and "fever then cough then nausea and vomiting".

Panic: Previous research conducted on public behavior indicated the role of political statements, media coverage, and misinformation on panic buying [26, 39]. Relevant keywords to represent consumer behavior include "toilet paper shortage" and "bleach coronavirus".

Protection: Especially as the pandemic progressed, safety and interest in preventative measures, such as mask-wearing and social distancing, as well as treatments, including monoclonal antibodies and vaccines, shifted. Keywords relevant to protection and treatment included "n95 mask covid", "flu vaccine", and "vaccine covid19".

Other Keywords: As the scope of this study was limited to the US, other specific factors may represent the public's willingness to follow preventative measures, including: "stimulus bill", "covid election", "anthony fauci".

3.1.3 Mobility Data. The Delphi Epidata API provides data on the spread and impact of COVID-19 on the United States at various geographic levels [8]. This study utilized public behavior (mobility less than 3, 3-6, and 6+ hours a day away from home and median at-home time) and COVID-related doctors' visits data signals.

3.1.4 Population, Incidence, and Other Data. COVID-19 incidence data is publicly available from The New York Times, based on reports from state and local health agencies [38]. This information was utilized as the ground truth number of cases over time, as well as a feature for the number of confirmed deaths due to the virus.

3.2 Problem Formulation

Predicting incidence of COVID-19 cases in the US is formulated as a regression task with multiple time series as input features. Predicting incidence for each state constitutes its own regression task.

- $\mathbf{x}_t^i$ is the value of the ith input series at time step t.
- $\mathbf{y}_t^i$ is the value of the ground truth (i.e., the number of new cases) at time step t.
- m is the number of historical time steps (i.e., size of the input window) utilized to predict n time steps (i.e., size of the prediction window) in the future. T is the total number of time steps available for learning.

We now develop a representative example of a deep learning-based model for COVID-19 forecasting, which will be used to test the impact of including different sources of data.

3.3 Learnable Detrending

To remove long-term trends from the input series, we introduced four learnable parameters θ_i per feature i: the initial level a_0^i, the initial trend b_0^i, the level smoothing coefficient α_t^i, and the trend smoothing coefficient β_t^i [35]. These parameters were updated using Holt's equations [11]:

$$\begin{aligned} a_t^i &= \alpha^i \mathbf{x}_t^i + (1-\alpha^i)(a_{t-1}^i + b_{t-1}^i), \\ b_t^i &= \beta^i (a_t^i - a_{t-1}^i) + (1-\beta_i)(b_{t-1}^i), \end{aligned} \tag{1}$$

Utilizing these constants, the residual input series following detrending was

$$\hat{\mathbf{x}}_t^i = \mathbf{x}_t^i - a_t^i \tag{2}$$

A detrending module predicting the level and trend coefficients from the input series was applied to each state's series to produce coefficients θ^s for each US state s.

3.4 Normalization

The residual input series x_t^i were separated into rolling input and prediction windows of size m and n, respectively. The magnitude of the series in these windows still may vary in magnitude, so min-max normalization was applied to their cumulative time time series. Specifically, over a window from t_1 to t_2:

$$s_t^i = \sum_{a=t_1}^{t} \hat{\mathbf{x}}_a^i, \quad \hat{s}_t^i = \frac{s_t^i - s_{t_1}^i}{s_{t_2}^i - s_{t_2}^i} \tag{3}$$

This data transformation constrains the input series to a monotonically increasing series between 0 and 1 over the input window.

3.5 Joint Training

A LSTM model was trained on the normalized input series $\hat{s}_{t-m:t}$ to predict $\hat{\mathbf{y}}_{t:t+n}$, an estimate of the residual values for the ground truth over the prediction window. This prediction task was specific to each state, as state-specific detrending coefficients were predicted from each state's data.

These predicted residual values are added to $\bar{\mathbf{x}}^i_{t:t+n}$, the linear extrapolation of the long-term trend, calculated as a linear combination of the relevant level and trend coefficients:

$$\bar{\mathbf{x}}^i_{t+k} = a^i_t + h \cdot b^i_t \tag{4}$$

Ultimately, the loss function is represented as

$$L(y, \hat{y}) = E(\mathbf{x}_{t:t+n}, \ \ \bar{\mathbf{x}}^i_{t:t+n} + \hat{\mathbf{y}}_{t:t+n}) \tag{5}$$

where the error function E was chosen to be Mean Squared Error (MSE).

4 RESULTS

To best utilize the available data, we selected an input window of 14 days and prediction window of 7 days. Training data was generated through rolling windows across all 50 US states from Day 0 to 130 and the remaining days (131 to 178) were reserved as the testing dataset for evaluation and comparison.

Following a search over several characteristics, the bolded hyperparameters were chosen on the basis of lowest testing MSE.

- LSTM layers: 1, **2**, 3
- LSTM hidden size: 100, **200**, 300
- Learning rate: 0.001, **0.005**, 0.01
- Batch size: 25, **50**, 75

Due to the fundamentally different data distributions between the training and testing sets, we experimented across several methods for regularization and model design: batch normalization, dropout ($p = 0, 0.2, 0.3, 0.4$) and L2 regularization (1e-4, 1e-5) between linear layers, L2 Regularization on the weights of the LSTM and Linear layers, and Epochs of training (30-60).

The best model over the hyperparameters and models was selected using the lowest MSE loss on the test set (as opposed to a validation set) at the end of the training period. Specifically, due to the limited available data (178 days) and the bias towards more interesting, non-exponential case patterns towards the end of the dataset, it was not possible to temporally split the data into meaningful training, validation, and test sets. In fact, creating a validation set would bias the "best performing" model towards a simpler model fitting to a training set of predominantly exponentially growing daily case counts.

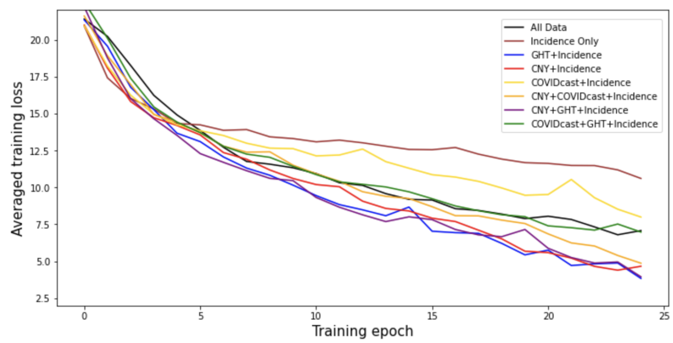
4.1 Ablation Study

To understand the relative importance of different features and learning components, components of the pipeline were selectively included and excluded to understand the overall effect on end predictive ability.

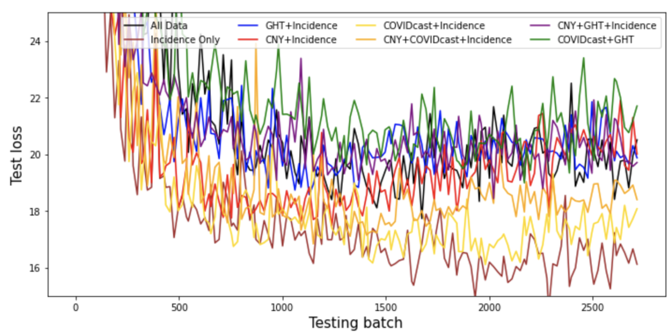
Table 1: Final train and test MSE for each trained model after 30 epochs (averaged across the batch size of 50).

Data Subset	*Train Test MSE*	*Final Test MSE*
All Data	7.0743	20.5065
Incidence Only	10.6059	16.1173
GHT+Incidence	3.8364	19.8716
CNY+Incidence	4.6599	20.0045
COVIDcast+Incidence	7.9918	18.0743
GHT+CNY+Incidence	3.645	20.417
CNY+COVIDcast+Incidence	3.9455	19.7069
GHT+COVIDcast+Incidence	6.9754	21.7014

Incidence data (specifically, case counts) was a key component of the learnable detrending process, and therefore was present in all the datasets.



(a) MSE on the train dataset for models trained on the different data subsets.



(b) MSE on the test dataset for models trained on the different data subsets. Every 20 training batches, the model was evaluated on the entire test set.

Figure 1: Training and test set loss curves for each of the experiments during training.

Figure 2 shows windows from the test set with results from all the models.

A few interesting trends are evident from the training and testing performance of the different models. The first is that more data didn't necessarily help the model achieve lower MSE on the prediction task: there is a consistent trend of data sources combined together to produce a trained model with equal to or worse performance than the original data source alone.

In addition, the models that perform best on the test set (Incidence, COVIDcast+Incidence) perform worse on the training set. Similarly, models that perform better on the train set significantly overfit to this subset of data as evidenced by the difference in train and test loss. This hypothesis of overfitting is strengthened by the data in Table 1.

Figure 2 shows four columns of example data from the test set where data is fluctuating, trending upward, trending downward, and showing other patterns, respectively , each with a different performance from the different models. There are numerous examples of agreement between the models, even when their collective prediction was far from the ground truth.

4.2 Feature Importance

We used the SHAP (SHapley Additive exPlanations) library, a prominant explainability method for deep learning which uses a game theoretic approach to explain the output of a machine learning model, to understand the relative importance of the 144 features in the model[22]. A SHAP value for a feature represents the model's expected prediction when conditioning on that feature [22]. Specifically, for this application we transformed our model using the GradientExplainer, which uses expected gradients to approximate SHAP values [22, 33, 37].

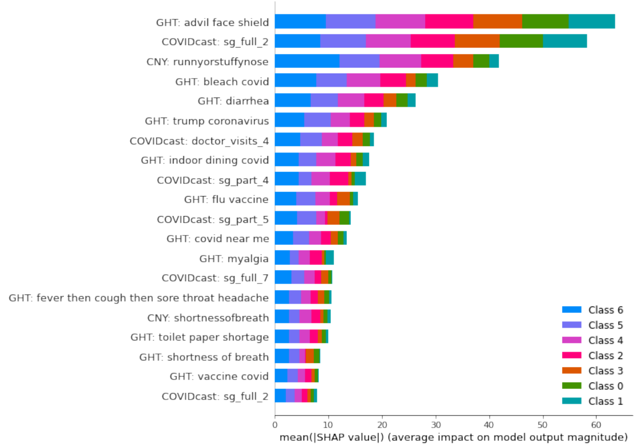


Figure 3: Mean SHAP values for each feature for each time step in the prediction window (denoted in the legend as "class") across the combined test set.

Figure 3 shows the relative SHAP values of the top 20 features explaining the output of the model across the 7-day predictive window. The features highlighted by SHAP generally fall into one of several general categories:

- Physical indicators (7 of 20): Mobility metrics, primarily fraction of the population at home 3-6 and 6+ hours of the day, dominated 5 of the 20 top features. Related physical indicators, such as doctors visits and
- COVID-19 symptoms (6 of 20): Features across all types of symptom features were significant to the model, from less common symptoms (myalgia) to common symptoms (runny or stuffy nose) to progression of symptoms (fever then cough then ...).
- Significant events (4 of 20): The presence of temporally-specific features (such as "trump covid", which spiked at the beginning of the pandemic and when President Trump was diagnosed with the virus in October and "bleach covid", which coincided with panic purchasing of disinfectants at the beginning of the pandemic) indicate the model's reliance on significant events for benchmarking. This reliance on features with high variance may allude to the reasons the model with all data sources overfit to the training set.
- Public health (3 of 20): Interest in preventative measures and general health (advil face shield, indoor dining, flu vaccine)

The importance of features across the 7-day prediction window incidates a reliance on different features for short-term trends versus longer-term trends. In particular, shortness of breath appears to be more important for predicting time steps further out from the input window based on the ratio of SHAP values in Figure 3. Shortness of breath (dyspnea) has been found to be positively associated with severe progression of COVID-19 and typically occurs later in the disease's course [31].

5 DISCUSSION AND CONCLUSION

This work contributes an analysis of the utility of different data types for epidemiological forecasting, and is notably the first to do so for COVID-19. We use the framework of a deep learning model with participatory surveys, mobility, and search trends data combined with traditional the case and death incidence.

Our results challenge the notion that having more data directly translates into better performance in an epidemiological prediction task. Typically, lack of synergy occurs due to the data generating processes of the data sources. Specifically, two data sources collected with similar generating processes (ex: search trends from different search engines across many overlapping keywords and themes) included together in a predictive model may result in lower model performance than if the more meaningful data source were included alone. There must also be alignment between the prediction task and the type of data utilized: since the prediction task in this study was tied to infection dynamics, it is intuitive that the most instructive features (and best-performing models) were those relating to incidence and mobility data. A review of the submissions to the COVID-19 Forecast Hub, a platform aggregating incidence predictions from across academic, industry, and independent research groups, shows that only two of the top 5 performing systems utilized data outside of daily new case and deaths [7]. Most of the high-performing submissions utilized mecahnistic or SEIR models rather than utilizing significant deep learning approaches [7].

The rigorous testing of the model was limited by the time horizon for which data was available. Expanding the available data and forming a validation set would allow for better analysis without sacrificing the generalizability of the model. Especially at the end of 2020, with spikes in cases following the Thanksgiving and Christmas holidays, as well as the ramp-up of vaccination campaigns

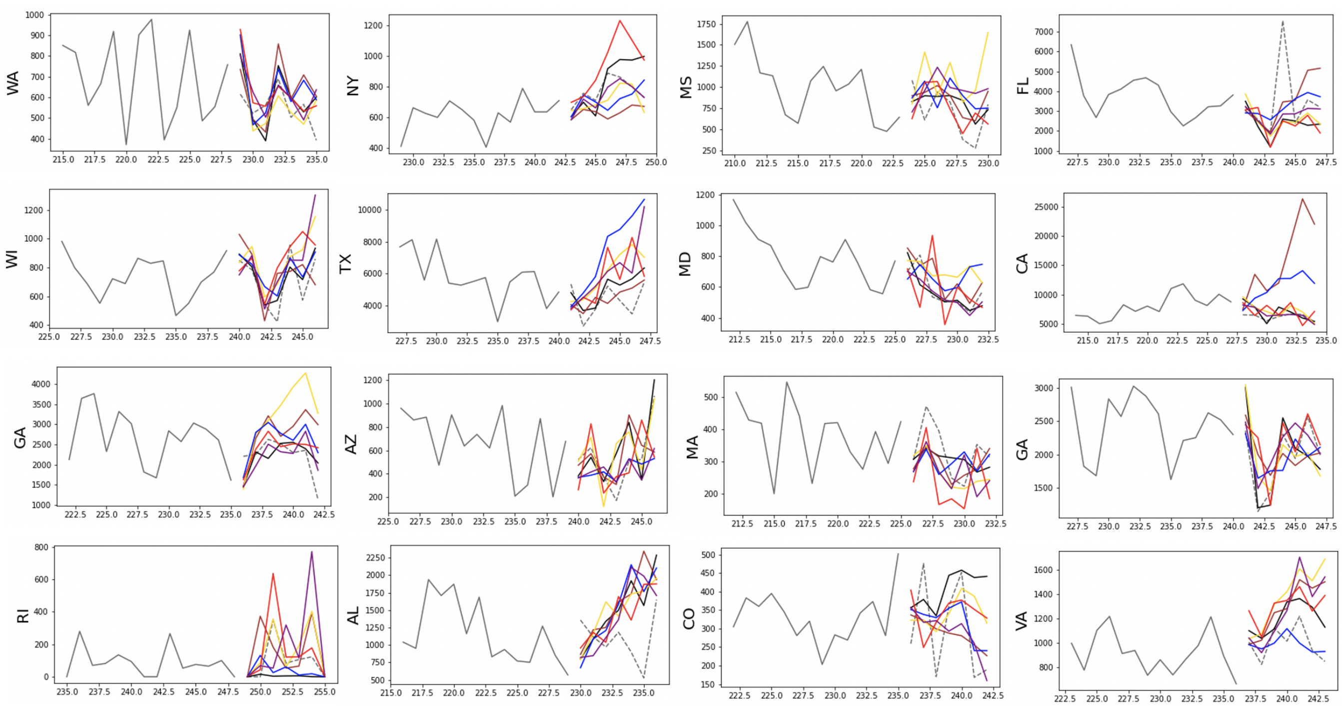


Figure 2: Example input window data with prediction window inferences for the 8 trained models. Models are color-coded with the same scheme as in Figure 1, with the input ground truth in solid grey and the ground truth of the prediction window in dotted grey.

across the US, challenging and interesting patterns likely emerge relating the input features and incidence.

5.1 Future Work

Given the framework for analyzing COVID-19-related data, next steps include identifying and studying the performance across different COVID-19 forecasting tasks, such as adherence to social distancing measures and daily deaths. In addition, given the nature of the models with high variance overfitting to extreme values, future work includes exploring debiasing approaches for the CNY and GHT data. Finally, other COVID-19 incidence metrics (hospitalizations, local testing results) as well as internet data sources (notably, social media) are of interest for further exploration.

REFERENCES

[1] 2021. Outbreaks Near Me. https://outbreaksnearme.org/us/en-US
[2] The U.S. Census Bureau. 2020. July 1, 2020 Estimates of Population and Housing Units. https://www.census.gov/newsroom/press-releases/2020/estimates-population-housing-units.html
[3] Rohitash Chandra, Ayush Jain, and Divyanshu Singh Chauhan. 2021. Deep learning via LSTM models for COVID-19 infection forecasting in India. *arXiv preprint arXiv:2101.11881* (2021).
[4] Serina Chang, Emma Pierson, Pang Wei Koh, Jaline Gerardin, Beth Redbird, David Grusky, and Jure Leskovec. 2021. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* 589, 7840 (2021), 82–87.
[5] Emily Chen, Kristina Lerman, and Emilio Ferrara. 2020. Tracking social media discourse about the covid-19 pandemic: Development of a public coronavirus twitter data set. *JMIR Public Health and Surveillance* 6, 2 (2020), e19273.
[6] Rumi Chunara, Mark S Smolinski, and John S Brownstein. 2013. Why we need crowdsourced data in infectious disease surveillance. *Current infectious disease reports* 15, 4 (2013), 316–319.
[7] Estee Y Cramer, Velma K Lopez, Jarad Niemi, Glover E George, Jeffrey C Cegan, Ian D Dettwiller, William P England, Matthew W Farthing, Robert H Hunter, Brandon Lafferty, et al. 2021. Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the US. *medRxiv* (2021).
[8] Aaron Rumack Ryan J. Tibshirani Roni Rosenfeld David C. Farrow, Logan C. Brooks. 2015. Delphi Epidata API. https://github.com/cmu-delphi/delphi-epidata
[9] Center for Disease Control and Prevention. 2021. Symptoms of Coronavirus. https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html
[10] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.
[11] Charles C Holt. 2004. Forecasting seasonals and trends by exponentially weighted moving averages. *International journal of forecasting* 20, 1 (2004), 5–10.
[12] Cornelia Ilin, Sébastien E Annan-Phan, Xiao Hui Tai, Shikhar Mehra, Solomon M Hsiang, and Joshua E Blumenstock. 2020. Public Mobility Data Enables COVID-19 Forecasting and Management at Local and Global Scales. (2020).
[13] Ali Imran, Iryna Posokhova, Haneya N Qureshi, Usama Masood, Muhammad Sajid Riaz, Kamran Ali, Charles N John, MD Iftikhar Hussain, and Muhammad Nabeel. 2020. AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. *Informatics in Medicine Unlocked* 20 (2020), 100378.
[14] Xiaoyong Jin, Yu-Xiang Wang, and Xifeng Yan. 2020. Inter-Series Attention Model for COVID-19 Forecasting. *arXiv preprint arXiv:2010.13006* (2020).
[15] 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).
[16] Juhyeon Kim and Insung Ahn. 2019. Weekly ILI patient ratio change prediction using news articles with support vector machine. *BMC bioinformatics* 20, 1 (2019), 1–16. https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-2894-2
[17] Vasileios Lampos, Maimuna S Majumder, Elad Yom-Tov, Michael Edelstein, Simon Moura, Yohhei Hamada, Molebogeng X Rangaka, Rachel A McKendry, and Ingemar J Cox. 2021. Tracking COVID-19 using online search. *NPJ digital medicine* 4, 1 (2021), 1–11.
[18] Joseph R Larsen, Margaret R Martin, John D Martin, Peter Kuhn, and James B Hicks. 2020. Modeling the Onset of Symptoms of COVID-19. *Frontiers in public health* 8 (2020), 473.

[19] Wendy Edelberg Lauren Bauer, Kristen E. Broady and Jimmy O'Donnell. [n.d.]. Ten facts about COVID-19 and the U.S. economy. https://www.brookings.edu/research/ten-facts-about-covid-19-and-the-u-s-economy/

[20] Bryan Lim and Stefan Zohren. 2020. Time series forecasting with deep learning: A survey. *arXiv preprint arXiv:2004.13408* (2020).

[21] Benjamin Lindemann, Timo Müller, Hannes Vietz, Nasser Jazdi, and Michael Weyrich. 2020. A Survey on Long Short-Term Memory Networks for Time Series Prediction. In *2020 14th CIRP Conference on Intelligent Computation in Manufacturing Engineering (ICME), Gulf of Naples*.

[22] Scott M Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.). Curran Associates, Inc., 4765–4774. http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf

[23] Samuel Mwalili, Mark Kimathi, Viona Ojiambo, Duncan Gathungu, and Rachel Mbogo. 2020. SEIR model for COVID-19 dynamics incorporating the environment and social distancing. *BMC Research Notes* 13, 1 (2020), 1–5.

[24] American Journal of Managed Care Staff. [n.d.]. A Timeline of COVID-19 Developments in 2020. https://www.ajmc.com/view/a-timeline-of-covid19-developments-in-2020

[25] Eleni Papoutsi, Vassilis G Giannakoulis, Vasiliki Ntella, Sofia Pappa, and Paraskevi Katsaounou. 2020. Global burden of COVID-19 pandemic on healthcare workers.

[26] Jessica Malaty Rivera, Shagun Gupta, Divya Ramjee, Ghinwa Y El Hayek, Nisrine El Amiri, Angel N Desai, and Maimuna S Majumder. 2020. Evaluating interest in off-label use of disinfectants for COVID-19. *The Lancet Digital Health* 2, 11 (2020), e564–e566.

[27] Daniel Romero-Alvarez, Nidhi Parikh, Dave Osthus, Kaitlyn Martinez, Nicholas Generous, Sara Del Valle, and Carrie A Manore. 2020. Google Health Trends performance reflecting dengue incidence for the Brazilian states. *BMC infectious diseases* 20 (2020), 1–15.

[28] Anca Rădulescu, Cassandra Williams, and Kieran Cavanagh. 2020. Management strategies in a SEIR-type model of COVID 19 community spread. *Scientific reports* 10, 1 (2020), 1–16.

[29] Loukas Samaras, Elena García-Barriocanal, and Miguel-Angel Sicilia. 2020. Comparing Social media and Google to detect and predict severe epidemics. *Nature Scientific reports* 10, 1 (2020), 1–11.

[30] Farah Shahid, Aneela Zameer, and Muhammad Muneeb. 2020. Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. *Chaos, Solitons & Fractals* 140 (2020), 110212.

[31] Li Shi, Ying Wang, Yadong Wang, Guangcai Duan, and Haiyan Yang. 2020. Dyspnea rather than fever is a risk factor for predicting mortality in patients with COVID-19. *Journal of Infection* 81, 4 (2020), 647–679.

[32] Junaid Shuja, Eisa Alanazi, Waleed Alasmary, and Abdulaziz Alashaikh. 2020. Covid-19 open source data sets: A comprehensive survey. *Applied Intelligence* (2020), 1–30.

[33] Daniel Smilkov, Nikhil Thorat, Been Kim, Fernanda Viégas, and Martin Wattenberg. 2017. Smoothgrad: removing noise by adding noise. *arXiv preprint arXiv:1706.03825* (2017).

[34] Mark S Smolinski, Adam W Crawley, Kristin Baltrusaitis, Rumi Chunara, Jennifer M Olsen, Oktawia Wójcik, Mauricio Santillana, Andre Nguyen, and John S Brownstein. 2015. Flu near you: crowdsourced symptom reporting spanning 2 influenza seasons. *American journal of public health* 105, 10 (2015), 2124–2130.

[35] Slawek Smyl. 2020. A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *International Journal of Forecasting* 36, 1 (2020), 75–85.

[36] Des Spence. 2016. Bad medicine: the worried hell. *British Journal of General Practice* 66, 651 (2016), 526–526.

[37] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. 2017. Axiomatic attribution for deep networks. In *International Conference on Machine Learning*. PMLR, 3319–3328.

[38] The New York Times. 2021. Coronavirus (Covid-19) Data in the United States. https://github.com/nytimes/covid-19-data

[39] Shu-Feng Tsao, Helen Chen, Therese Tisseverasinghe, Yang Yang, Lianghua Li, and Zahid A Butt. 2021. What social media told about us in the time of COVID-19: a scoping review. *arXiv preprint arXiv:2101.01688* (2021).

[40] Siva R Venna, Amirhossein Tavanaei, Raju N Gottumukkala, Vijay V Raghavan, Anthony S Maida, and Stephen Nichols. 2018. A novel data-driven model for real-time influenza forecasting. *IEEE Access* 7 (2018), 7691–7701.

[41] Svitlana Volkova, Ellyn Ayton, Katherine Porterfield, and Courtney D Corley. 2017. Forecasting influenza-like illness dynamics for military populations using neural networks and social media. *PloS one* 12, 12 (2017), e0188941.

[42] Lijing Wang, Xue Ben, Aniruddha Adiga, Adam Sadilek, Ashish Tendulkar, Srinivasan Venkatramanan, Anil Vullikanti, Gaurav Aggarwal, Alok Talekar, Jiangzhuo Chen, et al. 2020. Using Mobility Data to Understand and Forecast COVID19 Dynamics. *medRxiv* (2020).

[43] Timothy L Wiemken and Robert R Kelley. 2020. Machine learning in epidemiology and health outcomes research. *Annual review of public health* 41 (2020), 21–36.

[44] Chenfeng Xiong, Songhua Hu, Mofeng Yang, Hannah Younes, Weiyu Luo, Sepehr Ghader, and Lei Zhang. 2020. Mobile device location data reveal human mobility response to state-level stay-at-home orders during the COVID-19 pandemic in the USA. *Journal of the Royal Society Interface* 17, 173 (2020), 20200344.

[45] Novanto Yudistira. 2020. COVID-19 growth prediction using multivariate long short term memory. *arXiv preprint arXiv:2005.04809* (2020).

[46] Joshua Graff Zivin and Nicholas Sanders. 2020. The spread of COVID-19 shows the importance of policy coordination. *Proceedings of the National Academy of Sciences* 117, 52 (2020), 32842–32844.

[47] Bin Zou, Vasileios Lampos, and Ingemar Cox. 2018. Multi-task learning improves disease models from web search. In *proceedings of the 2018 world wide web conference*. 87–96. https://www.lampos.net/papers/www2018-google-flu-mtl.pdf