![THE UNIVERSITY of EDINBURGH]

Edinburgh Research Explorer

Structural elements promote architectural stripe formation and facilitate ultra-long-range gene regulation at a human disease locus

OPEN ACCESS

HOIST: Evidence-Driven Spatio-Temporal COVID-19 Hospitalization Prediction with Ising Dynamics

Junyi Gao[1,2], Joerg Heintz[3], Christina Mack[4], Lucas Glass[4], Adam Cross[5]*, Jimeng Sun[3]*

1 The University of Edinburgh, Edinburgh, Edinburgh, UK

2 Health Data Research UK, UK

3 University of Illinois Urbana Champaign, Champaign, IL, USA

4 IQVIA, Cambridge, Massachusetts, USA

5 University of Illinois, College of Medicine Peoria, Department of Research Services, Peoria, IL, USA

* Correspond to Adam Cross (arcross@uic.edu) and Jimeng Sun (jimeng@illinois.edu)

Abstract

In this work, we aim to accurately predict the number of hospitalizations during the COVID-19 pandemic by developing a spatio-temporal prediction model. We propose HOIST, an Ising dynamics-based deep learning model for spatio-temporal COVID-19 hospitalization prediction. By drawing the analogy between locations and lattice sites in statistical mechanics, we use the Ising dynamics to guide the model to extract and utilize spatial relationships across locations and model the complex influence of granular information from real-world clinical evidence. By leveraging rich linked databases, including insurance claims, census information, and hospital resource usage data across the U.S., we evaluate the HOIST model on the large-scale spatio-temporal COVID-19 hospitalization prediction task for 2,299 counties in the U.S. In the 4-week hospitalization prediction task, HOIST achieves 368.7 mean absolute error, 0.6 R^2 and 0.89 concordance correlation coefficient score on average. Our detailed number needed to treat (NNT) and cost analysis suggest that future COVID-19 vaccination efforts may be most impactful in rural areas. This model may serve as a resource for future county and state-level vaccination efforts.

Introduction

The COVID-19 pandemic has caused the enormous social and economic loss. With over 90 million confirmed cases and 1 million deaths in the U.S. by Aug 2022[1], the pandemic placed a heavy burden on national and global healthcare systems. The disparities in medical resource availability among U.S. counties – including ventilators, hospital beds, and critical care staff – can and, in some cases, have devastatingly impacted patient outcomes[2]. Nationwide vaccination efforts have primarily favored urban areas, and the urban-rural disparity in vaccination coverage has continued to worsen throughout the pandemic[3]. Rural vaccination efforts are technically challenging due to a combination of more limited access to healthcare, generally lower perception of the severity of COVID-19 by residents of rural communities, and usually higher levels of vaccine hesitancy[3]. This multi-year trend of relative rural under-immunization raises the question of whether further vaccination efforts might be more impactful in these under-immunized areas regarding hospitalization, death, and cost-effectiveness. Answering this question is technically challenging but clinically meaningful; if counties can be stratified in terms of improved outcomes per vaccination,

41 this may provide meaningful insight into how local and statewide efforts might be distributed to
42 maximize impact. To accomplish this task, we propose the following technical contributions:

43 1. Extracting inter-region similarities. The pandemic progression might be similar in regions with
44   similar geographical and socioeconomic properties. If so, extracting and utilizing these
45   similarities would help the model better predict pandemic progression in the future. Most
46   existing spatio-temporal pandemic prediction methods use pre-defined location graphs to
47   explicitly inform the spatial connections to the model[4-7]. In contrast, these manually defined
48   location graphs may differ significantly from the ground truth inter-regional relationships. Other
49   works learn the spatial connectivity purely based on the historical pandemic progressions[8,9],
50   regardless of the underlying social, economic, or geographical similarities, which cannot
51   capture the inter-region similarities in full view and lead to inferior prediction results. Therefore,
52   it is non-trivial to model spatial dependencies with multi-source background data in a flexible
53   way.

54 2. Utilizing complex inner-region influence factors. Local COVID-19 hospitalization rates are
55   likely affected by many factors, including the number of infected patients, high-risk patient
56   population size, and vaccination rates. Estimating these effects becomes more challenging with
57   more granular information, such as the number of COVID-19 immunizations per individual.
58   Leveraging these complex inner-region influence factors, such as medical claims and
59   vaccination statistics, may significantly improve hospitalization prediction performance.
60   Recent research combines the susceptible-infected-removed (SIR) and susceptible-exposed-
61   infected-removed (SEIR) differential models with deep learning models to improve predictions
62   by simulating real-world SIR dynamics[4-7]. However, when applied to hospitalization prediction
63   tasks, these models face the following issues: (1) Limited data sources and model parameters
64   in the traditional epidemiological model make it difficult to utilize these complex high-
65   dimensional inner-region factors effectively. Ignoring complex inner-region factors may cause
66   the model to fail to capture the underlying disease distribution, virus subvariants, and
67   vaccination effectiveness rates across the state, which could lead to biased results. While purely
68   deep-learning-based time series prediction models such as recurrent neural networks (RNN)
69   can handle high-dimensional data and extract complex nonlinear relationships, they ignore real-
70   world progression dynamics and have minimal interpretability compared to epidemiological
71   based models. (2) Most epidemiological models offer a simple increasing-decreasing trend with
72   different rates, while the hospitalization curves are more complex and may not follow this
73   simple trend. (3). Moreover, the SIR dynamics are not designed for spatio-temporal prediction
74   tasks, so these models cannot directly utilize for hospitalization prediction.

75   To address the above challenges, we propose an Ising dynamics-based deep learning model for
76 spatio-temporal COVID-19 hospitalization prediction. We have combined multiple data sources,
77 including disease and vaccination statistics from real-world medical claims, medical resource usage,
78 census, and geographical and mobility data, to create an evidence hub for the ability to train a
79 complex evidence-driven spatio-temporal prediction model, HOIST. The proposed HOIST model
80 can handle complex inner-regional influence factors from multiple data sources and adaptively learn
81 the inter-regional relationships using the locational census and geographical and inter-locational
82 mobility data without requiring a pre-defined location graph with fixed edges.

83   The model learning process is guided by the Ising dynamics, a mathematical statistical
84 mechanics model to estimate site spin configurations in square lattices[10]. By drawing the analogy

85   between locations and lattice sites, we find the Ising dynamics a natural choice to model the inner-
86   region factors and inter-region similarities simultaneously: (1) the Ising dynamics guide the model
87   to extract and utilize inter-region spatial relationships by taking prediction results from similar
88   locations as kinetic energy; (2) the Ising dynamics model the complex influence of granular inner-
89   region factors from the real-world clinical evidence as potential energy (i.e., external fields). Both
90   kinetic energy and potential energy jointly decided the energy of the location (predicted
91   hospitalization case to increase or decrease). The Ising dynamics in this paper are used as a
92   regularizer to guide the model learning and prediction process, which enables the model to learn
93   complex nonlinear patterns flexibly. We use recurrent neural networks (RNN) to learn the
94   parameters of the Ising model so our model can also extract and utilize temporal patterns in the data.
95   The overview of HOIST is shown in **Figure 1**. The HOIST model first uses static data to
96   calculate the distances between locations in the latent space. These distances are further normalized
97   into 0 and 1, which indicate the connectivity between locations learned by the model. The dynamic
98   data are used to estimate the External Fields (E.F.s) of locations and used to generate predictions
99   with LSTM. The estimated E.F.s and the learned connectivity are used to calculate the Ising dynamic
100   loss, which is then used to regularize the prediction results and better model the real-world
101   connectivity and effects of various influence factors.
102   We evaluate the HOIST model on the large-scale spatio-temporal COVID-19 hospitalization
103   prediction task for 2,299 counties in the United States. This scale is much larger and more granular
104   than existing COVID-19 spatio-temporal predictive works. In the 4-week hospitalization prediction
105   task, HOIST achieves 368.7 MAE, 0.6 R^2 and 0.89 CCC scores. We also conduct experiments with
106   different lengths of prediction windows from 1 week to 5 weeks. Compared to the best baseline
107   model, HOIST achieves 48% lower MAE, 65% lower MSE, 272% higher R^2 and 51% higher CCC
108   on average. The prediction performances under the temporal data split setting show that HOIST can
109   consistently achieve low prediction error regardless of the underlying data distribution shift caused
110   by new virus variants. These results suggest the HOIST model can accurately predict both long-
111   term trends and short-term variations, enabling broader real-world applications.
112   Unlike totally black-box deep learning time-series models, the fusion of real-world dynamics
113   in HOIST allows us to look deeper into the model to see how various influence factors affect the
114   model predictions. By analyzing the weights of external fields, we find that the booster vaccination
115   rate has a more significant negative correlation with future hospitalization cases (i.e., more profound
116   impact) compared to the first and second vaccinations in the series. We also find that the effects of
117   two major vaccination brands, Pfizer and Moderna, have no statistical difference in our model. By
118   simulating varying immunization rates, we conduct a detailed analysis of the marginal benefit of the
119   vaccination ratio by answering two questions: (1) How many more vaccinations are necessary to
120   prevent one hospitalization case for a specific location? (2) What is the cost ratio between these
121   vaccinations and the average COVID-19 hospitalization for a specific location? Quantitatively, our
122   model shows that increasing the vaccination ratio by 10% can reduce the number of current
123   hospitalization cases by 15% on average for all locations. We also find that the cost ratio is generally
124   highest in much of the rural Midwest and Rock Mountain regions, suggesting that prioritization of
125   vaccine efforts in these counties could most significantly reduce the overall statewide healthcare
126   financial burden of COVID-19. For 368 counties among 43 states all over the United States, rural
127   vaccination outreach efforts are likely cost-saving endeavors. We believe these results and our
128   model may inform clinicians, healthcare institutions, and policymakers to improve their decisions

129 and ultimately reduce the negative economic and health impacts caused by the pandemic. All
130 detailed county-level analysis results are available at the online visualization platform at
131 https://v1xerunt.github.io/HOIST/.

Results

Problem formulation and data sources

134   We develop the HOIST model to predict the total number of COVID-19 hospitalization cases
135 in an upcoming four-week period at a county level across the United States. Throughout this paper,
136 we use N to denote the number of locations (counties) and T to denote the length of timesteps
137 (days). The model uses static location background data to adaptively learn location connectivity and
138 then uses dynamic data to learn the temporal patterns. The model uses the learned spatio-temporal
139 patterns to conduct predictions. Concretely, we formulate model inputs and the prediction task as
140 follows:

141 **Input 1 (Static Data).** The static data include the distance matrices $\mathbf{D}, \mathbf{G} \in \mathbb{R}^{N \times N}$, population
142 demographics statistics $\mathbf{M} \in \mathbb{R}^{N \times 9}$ and economics and healthcare statistics $\mathbf{E} \in \mathbb{R}^{N \times 4}$. $\mathbf{D}$ is the
143 geographical distance matrix and $\mathbf{G}$ is the mobility distance matrix. The geographical distances are
144 the Haversine distances between N locations. The mobility distance includes the average mobility
145 flows between N locations during 2020 and 2021. The mobility scores are collected from the
146 Multiscale Dynamic Human Mobility Flow Dataset, which analyzed millions of anonymous mobile
147 phone users' visits to various places provided by SafeGraph[11]. The census features $\mathbf{M}, \mathbf{E}$ include
148 populations of different age and race groups and medical resource statistics, which are collected
149 from the county-level census dataset provided by[12].

150 **Input 2 (Dynamic Data).** All dynamic data are three-dimensional tensors, including the daily new
151 infected COVID-19 case counts $\mathbf{I} \in \mathbb{R}^{N \times T \times 1}$, medical claim statistics $\mathbf{C} \in \mathbb{R}^{N \times T \times 20}$, vaccination
152 statistics $\mathbf{V} \in \mathbb{R}^{N \times T \times 17}$, and medical resource usage statistics $\mathbf{U} \in \mathbb{R}^{N \times T \times 4}$. The daily new cases $\mathbf{I}$
153 are collected from the Johns Hopkins COVID-19 Data Repository[13]. Claim statistics $\mathbf{C}$ include
154 daily statistics of total patients, patients older than 65, and patients with certain comorbidities. The
155 age threshold and comorbidities are derived from the CDC COVID-19 guidelines[14] and the Charlson
156 Comorbidity Index[15]. The vaccination statistics $\mathbf{V}$ include daily statistics of 13 different
157 vaccination CPT codes. Both vaccination and claim statistics are collected from IQVIA's real-world
158 claims dataset[16]. The medical resource usage statistics $\mathbf{U}$ include the usage of inpatient and ICU
159 beds and their usage for COVID-19 patients, which are collected from HealthData.gov[17]. Though
160 $\mathbf{U}$ also provide the information for hospitalization cases, we use $\mathbf{C}$ instead of $\mathbf{U}$ as the ground-
161 truth case numbers. This is because $\mathbf{U}$ only provides weekly statistics and the statistics are collected
162 from fewer healthcare institutions compared to the claims data. We include medical resource usage
163 statistics $\mathbf{U}$ to help reduce the potential biases in the claims data.

164   The feature list and statistics can be found in the Supplementary Information.

165 **Task (Spatio-temporal COVID-19 Hospitalization Prediction).** Given all static and dynamic data
166 of N locations and T timesteps, our task is to predict the total number of COVID-19
167 hospitalization cases in future L days for each of the N locations, denoted as $\mathbf{y} \in \mathbb{R}^N$. Since the
168 claims data do not show the primary cause for hospitalizations, we define a hospitalization case as
169 a patient hospitalized within 35 days after the COVID-19 diagnosis in the claims dataset and we
170 calculate the total hospitalization cases for each location to get $\mathbf{y}$. This setting is inspired by

171   previous COVID-19 hospitalization prediction works[18,19].

Experiment settings

173   In the experimental phase, we extract all the required data from September 2020 to May 2022.
174   We set the input window τ to 5 weeks and the prediction window L to 4 weeks. We select counties
175   using the FIPS codes. The final number of counties are 2,299. We only exclude counties that have
176   zero cases for all timesteps. Note that our experiment scale is much larger and more granular than
177   existing works[4,5], which allows for a wider scope of possible applications.

178   We use the mean square error (MSE), mean absolute error (MAE), the coefficient of
179   determination (R^2), and the average concordance correlation coefficient (CCC) to evaluate model
180   prediction. The CCC and R^2 are computed as:

$$R^2 = 1 - \frac{\sum(\hat{y}_i - y_i)^2}{\sum(\hat{y}_i - \mu_y)^2} \tag{1}$$

$$CCC = \frac{2\rho\sigma_{\hat{y}}\sigma_y}{\sigma_{\hat{y}}^2 + \sigma_y^2 + (\mu_{\hat{y}} - \mu_y)^2} \tag{2}$$

183   where $\mu_{\hat{y}}$ and μ_y are the means for the predictions and ground-truth, and $\sigma_{\hat{y}}^2$ and σ_y^2 are the
184   corresponding variances. ρ is the correlation coefficient between the two variables $\hat{y}$ and y. Note
185   that the range of R^2 is $(-\infty, 1)$, so an extreme value may significantly affect the average value. In
186   contrast, the range of CCC is $(-1,1)$, so this will be less affected by extreme values.

187   We split the sequence into train, validation, and test sets in a 3:1:1 ratio. The training sequences
188   are from Sep 04, 2020 to Sep 02, 2021; the validation sequences are from Sep 03, 2021 to Dec 23,
189   2021, and the test sequences are from Dec 24, 2021 to Apr 15, 2022. We train the model on the
190   training set and save the model and hyper-parameters with the best performance on the validation
191   set. We then test the model on the testing set and report the performance. We train all models 5 times
192   with different random initializations and calculate the standard deviations. Due to diverse location
193   characteristics, the average hospitalization cases vary from a few to tens of thousands of cases. This
194   large variation in case numbers poses a challenge for the deep learning model to learn stable
195   parameters. We therefore conduct the log transformation on the prediction targets and scale the
196   model predictions back to the original scale, then calculate the performance metrics. We also provide
197   the prediction uncertainty at 90% confidence in all county-level and state-level prediction plots
198   using the conformal methods[20]. The conformal algorithms possess explicit and non-asymptotic
199   guarantees without distributional assumptions or model assumptions, so that they can be easily
200   applied to all trained models. The prediction interval α is estimated as:

$$\alpha = [\hat{y} - s\sigma, \hat{y} + s\sigma] \tag{3}$$

202   where $\hat{y}$ is the average prediction using the 5 models with different random initializations, σ
203   denotes the standard deviation of 5 models. The value of parameter s depends on the required
204   confidence. For example, if we aim to obtain the prediction interval with 90% confidence, we use
205   calculate the s to make the prediction interval α covers 90% ground truth on the validation set.
206   Then we applied the s on the test set to calculate the test prediction interval. We provide more
207   detailed model uncertainty analysis in the Supplementary Information.

208   We use Python 3.9, PyTorch 1.12[21], scikit-learn 1.2, NumPy 1.19 to collect the data and
209   implement the models. We use the mini-batch gradient descent strategy to train the models and the
210   batch size is set to 128. We use Adam optimizer with learning rate 0.01 for 300 epochs. We save the

model with the highest score on the validation set and report the prediction performances on the testing set. All the experiments are done on the server with Intel i9-13900K CPU, 64 GB RAM and one NVIDIA RTX 4090 GPU. The HOIST source code is publicly available on GitHub (https://github.com/v1xerunt/HOIST).

Baseline models

We compare HOIST with the following epidemiology and deep-learning methods.

1. **DELPHI-SEIR[22]**: This is a variant of the SEIR (susceptible, exposed, infectious, and recovered) epidemiology model. Compared with the traditional SEIR model, DELPHI-SEIR can model hospitalization trends and policy strategies. We use the trust region[23] optimization strategy to estimate the model's parameters.

2. **GRU**: We input all features into a gated recurrent unit (GRU) model and predict the future number of cases. The GRU model is a variant of RNN, widely applied in multiple pandemic prediction works[24-26]. The hidden dimension of GRU is set to 128.

3. **LSTM**: We input all features into a long short-term memory (LSTM) model and predict the future number of cases. The LSTM model is another variant of RNN. The hidden dimension of LSTM is also set to 128.

4. **ColaGNN[9]**: ColaGNN learns the location graph with sequential data to learn spatial relationships for pandemic progression. The hidden dimension of RNN is set to 128, and the convolution filter dimension is set to 64.

5. **ACTS[8]**: ACTS is a COVID-19 forecasting model which uses the inter-series attention mechanism to learn spatial relationships between locations. The convolution filter dimension is set to 64 and the segment length is set to 14.

6. **CovidGNN[27]**: CovidGNN uses a graph neural network with skip connections to predict future COVID-19 cases. We use a two-layer graph attention network, and the graph network dimension is set to 64.

7. **STAN[4]**: STAN fuses the SIR dynamics into a spatio-temporal prediction model for COVID-19 case prediction. The graph network dimension is set to 64, and the hidden dimension of GRU is set to 128. Since the SIR dynamics do not apply to the hospitalization prediction task, we remove the SIR constraints in the STAN model.

All models can access the same data sequences with the same input window and are all evaluated on the same testing set. For the spatial-temporal prediction models (i.e., ColaGNN, ACTS, CovidGNN, STAN), we use the static data to build the location graph or calculate the location similarities in their algorithms. For the GRU and LSTM model, we concatenate the static data with the original inputs at each timestep to the model. The SEIR model cannot take the static location data as inputs. All model hyper-parameters are decided by using grid search on the validation dataset. For the DELPHI-SEIR model, we use the deployed version of the DELPHI-SEIR model (i.e., DELPHI-SEIR V4) and recommended optimal parameters. To assess the performance improvement from the Ising dynamics, the adaptative connectivity learning, and the real-world evidence, we conduct an ablation study by comparing HOIST against the following ablation versions from both a data perspective and a method perspective:

1. **HOIST-Vaccination**: We remove the vaccination statistics data from the model input sequences of HOIST.

2. **HOIST-Risk**: We remove the high-risk patient statistics in the real-world claims data from

254   model input sequences of HOIST.

255   3. **HOIST-AC**: We remove the adaptative connection learning module from HOIST. The **S**
256   matrix is learned by calculating the sequence similarities instead of using location
257   background data.

258   4. **HOIST-Ising**: We remove the Ising dynamic loss and replace the E.F. modeling module
259   with a naïve LSTM network in HOIST.

Model performance analysis

261 We design experiments to answer the following research questions:

262   1. How well does HOIST perform in the hospitalization prediction task?
263   2. How well does HOIST perform with different lengths of prediction window?
264   3. How well does HOIST perform under temporal data split setting?
265   4. What is the analysis of the learned external field weights?
266   5. How does the HOIST model help to increase vaccination rates effectively?

28-day hospitalization prediction performance

268   **Table 1** shows the performance of the 28-day hospitalization prediction task. Compared to the
269 best baseline model, HOIST reaches 70% lower MSE and 50% lower MAE. It also achieves an R^2
270 score of 0.6 and a CCC of 0.89 while the best baseline model achieves 0.16 R^2 and 0.6 CCC. We
271 find ColaGNN and ACTS achieve better prediction performance among the compared models; this
272 may be because both CovidGNN and STAN require a fixed location graph structure as input, while
273 ColaGNN and ACTS can learn the spatial relationships based on the sequence similarity. Compared
274 to the infected case prediction task, the spatial patterns of hospitalization cases may be more
275 complex and thus cannot be pre-defined using a fixed location graph. Therefore, models that can
276 learn flexible connections commonly outperform all others in this task. By utilizing the location-
277 static background data, HOIST can better extract spatial patterns of the pandemic progression. We
278 also find that traditional epidemiology SEIR models fail to predict accurately in this task, probably
279 because the hospitalization curves do not follow a simple increasing-decreasing trend, which is the
280 underlying assumption of most epidemiological models. Additionally, the ground-truth curves may
281 have multiple peaks and complex short-term variations, which increases the difficulty of the
282 predictive task. Note that when calculating R^2 and CCC scores, the average value μ_y used in the
283 denominator is the average value in the test time phase, which is often difficult to beat since the
284 future μ_y is unavailable at the prediction time. Therefore, in all experiments, any positive R^2 or
285 high CCC scores can be impressive to achieve.

286   We conduct the student *t*-test to evaluate the significance of the performance differences. The
287 results show that the performance differences between HOIST and the best baseline models are
288 significant (p<0.001). HOIST also outperforms all ablation versions of HOIST by a large margin,
289 validating the effectiveness and necessity of all proposed modules. According to the performance
290 analysis, the Ising dynamics are the most critical component of HOIST. By modeling various
291 influence factors as external fields and using spatial-temporal dynamics to regularize the model
292 learning process, HOIST confirms that the Ising dynamics can improve spatio-temporal predictive
293 performance by more closely resembling the real-world curve. Furthermore, integrating real-world
294 evidence data, such as high-risk patient cohorts and vaccination statistics can reduce prediction error.
295 Using background data to learn adaptive spatial connectivity in HOIST also enables it to

outperform the reduced model, which only uses sequence similarities (i.e., HOIST-AC). The adaptative connectivity learning module also helps the HOIST learn clustered spatial embeddings using the demographics data. We provide visualization plots of the learned embeddings in the supplementary material. The results show that HOIST can not only extract geographical similarities between locations, but also can identify geographically distant but socio-economically similar locations. We also provide the predicted curve plots for all US states in the Supplementary Information.

Prediction performance with different lengths of prediction window

In this section, we evaluate each model's long-term and short-term prediction performance by changing the prediction window L from 1 week to 5 weeks and then train and evaluate all the models with the same strategy. The MAE, MSE, R^2, and CCC of HOIST and the 6 best baseline models are shown in **Figure 2**. Due to its poor performance, we exclude the SEIR model.

The results demonstrate that the MSE and MAE of all models increase as the length of the prediction window increases. This is expected since predicting further into the future becomes progressively more challenging. However, HOIST consistently achieves a much lower MAE and MSE than all baseline models. HOIST also achieves consistently high R^2 and CCC scores. Compared to the baseline models, HOIST achieves on average a 48% lower MAE, 65% lower MSE, 272% higher R^2, and 51% higher CCC. The results demonstrate that the HOIST model accurately predicts both long-term trends and short-term variations, enabling broader real-world applications.

Prediction performance under temporal data split setting

New variants of COVID-19 virus can lead to different disease severity. It is non-trivial for a model to accurately predict future hospitalization under such distribution transitions. We design experiments to evaluate how model performance evolves over time. We use a sliding window training setting by using 10 weeks data for training and using next 4 weeks data for testing. We split the time from Sep 2020 to April 2022 into 7 periods and the model are tested in 7 testing time phases. We compare HOIST against other 3 best baseline models (i.e., LSTM, ACTS and ColaGNN). The results are shown in **Figure 3**.

The results show that HOIST consistently outperform all baseline models in terms of MSE and MAE on all testing phases. HOIST consistently achieves low prediction errors and high CCC scores and is less affected by the temporal data distribution shifts. We notice that some baseline models experience low prediction performance on Feb 2021 and Jan 2022. This may be due to distribution shifts between training and testing data caused by emerging of new variants of the COVID-19 virus. However, HOIST achieves lower prediction error than baselines on these time phases.

Analysis of learned external fields

Compared with the baseline black-box models, modeling external fields with Ising dynamics allows us to analyze the importance of different factors for different locations, providing us with more insight into which features are most predictive of hospitalization. In the learned E.F. parameters $\hat{\mathbf{f}}$, each dimension denotes the weight of the corresponding input factor. To visualize the effect of the number of vaccinations on the hospitalization rate, we plot the corresponding weights of the following vaccination features: 1) total number of vaccinations, 2) number of first vaccination administrations, 3) number of second vaccination administrations, and 3) number of booster

337   administrations, as shown in **Figure 4**.

338   The negative EF weights occur because vaccination ratio is negatively correlated with
339   hospitalization case count. However, the results show that the E.F. weights of all four features are
340   increasing, meaning the predictiveness of vaccinations are decreasing over time. This is possibly
341   due to the increased prevalence of new variants of the virus, many of which are more resistant to
342   vaccinations, which is consistent with conclusions in recent medical research[28,29]. However, we can
343   still observe that the booster doses (the purple line) are more effective than the first and second
344   doses (p-value < 0.05). We also compare the model effect (i.e., the E.F. weights) of Pfizer (CPT
345   code 0001A~0004A) and Moderna (CPT code 0011A~0013A, 0064A) COVID-19 vaccines, and
346   we find that their effects do not have statistical difference in our model (p-value 0.1).

347   We also explore the E.F. weights of high-risk conditions, which are condition codes defined by
348   the CDC COVID-19 People with Certain Medical Conditions guideline[14] and the Charlson
349   Comorbidity Index[15]. These features are proved to have significant impacts on COVID-19 patients'
350   outcomes[30,31]. We find the feature with the highest E.F. weight is the number of patients older than
351   65. The top 5 conditions are: renal disease, dementia, immunodeficiency, malignancy, and chronic
352   lung disease. The results are consistent with CDC guidelines, and we do not observe statistically
353   significant differences in these five features.

Predicted hospitalization hotspot map and error analysis

355   We present the predicted hospitalization hotspot map in **Figure 5**. **Figure 5-(a)** is the hotspot
356   map colored by the number of hospitalization cases and **Figure 5-(b)** is colored by the ratio of
357   hospitalization cases over total population. There are some blank areas in the map in cases of no
358   valid data or zero case counts in these counties. These heatmaps can be accessed in an online
359   visualization platform at https://v1xerunt.github.io/HOIST/.

360   We find that the spatial patterns of hospitalization are similar in both figures. West coast,
361   Northeast coast and Midwest areas generally have high hospitalization case counts and per-person
362   occurrence. Interestingly, we also find that some locations – such as the San Diego county (A) and
363   Riverside county (B) along the West coast – have very high case counts but only medium-level
364   hospitalization ratios. A possible interpretation could be that the people from these counties tend to
365   travel to Los Angeles for hospitalization.

366   We also plot a heatmap to show the spatial patterns of the mean average percentage error
367   (MAPE). We only calculate the MAPE for locations where case counts are larger than 100 in the
368   testing phase because a small prediction error can lead to very large MAPE for locations with low
369   counts. The results are shown in **Figure 5-(c)**.

370   We find that HOIST achieves low MAPE for most locations, with only 15 locations having a
371   MAPE greater than 0.5. For these locations, HOIST tends to underestimate the case counts. This
372   may be because there are some pattern changes in the testing time window; for example, in
373   Effingham, Georgia, HOIST has 0.64 MAPE. When we plot the predicted hospitalization curve and
374   the number of infected cases (**Figure 6**), we find several peaks in the infected case count curve
375   during both the validation and testing phases, which is quite different from training patterns. These
376   sudden fluctuations in input features may cause HOIST to fail to accurately predict the surge of
377   hospitalization cases during the testing phase. We provide another two case examples for error
378   analysis in the Supplementary Information.

Number needed to treat prediction with HOIST

Increasing a community's vaccination rate is an effective method of reducing both the occurrence and severity of COVID-19[32]. However, there is still uncertainty surrounding how and whether this effectiveness is influenced by local vaccination rates. In other words, we ask the question, "Due to the disproportionately higher vaccination rates in urban areas, do inequalities exist regarding the number of additional vaccinations necessary to reduce a single hospitalization or death?" We also calculate the cost effectiveness of future vaccination efforts based on the county-level average cost of COVID-19 hospitalization. Accurately answering these questions at the county level may allow policymakers and healthcare institutions with limited distribution capabilities and finite medical resources to develop more targeted vaccination efforts.

The high predictive accuracy of HOIST enables us to answer these questions. By changing the vaccination rates in the input features, HOIST can simulate the changes of the predicted hospitalization curve. We use the predicted number of hospitalizations to calculate the Number Needed to Treat (NNT) with respect to how many vaccinations are needed to prevent one hospitalization at each county. For each location, we increase the vaccination rate by 10% (i.e., 10% total population) and calculate the predicted hospitalization case reduction between the original prediction $\hat{y}$ and the simulated prediction $\hat{y}_{sim}$. The NNT for hospitalization is calculated as:

$$\text{NNT}_h = \frac{1}{\text{CER} - \text{EER}} = \frac{1}{\left(\frac{\hat{y}}{0.1p} - \frac{\hat{y}_{\text{sim}}}{0.1p}\right)} = \frac{p}{10(\hat{y} - \hat{y}_{\text{sim}})} \tag{4}$$

where CER denotes the control event rate (original predicted hospitalization rate), EER denotes the experimental event rate (simulated hospitalization rate after the vaccination rate increases by 10%), and p denotes the population size of the location.

Furthermore, we acknowledge that morbidity and mortality among individuals hospitalized with COVID-19 occur disproportionately in certain ethnic and racial minority groups. To address these outcome disparities in our model, we adjust the NNT for each county based on its demographics and their respective risk ratios:

$$\text{NNT}_a = \frac{\text{NNT}_h}{1 * r_{\text{white,nh}} + 0.8 * r_{\text{asian,nh}} + 1.7 * r_{\text{black,nh}} + 1.8 * r_{\text{hispanic}}} \tag{5}$$

where $r_{\text{white,nh}}$, $r_{\text{asian,nh}}$ and $r_{\text{black,nh}}$ denote the population ratio of White, Asian, and Black non-Hispanic persons, and r_{hispanic} denotes the population ratio of Hispanic or Latino persons. All county-level population ratios are extracted from the census data. The risk adjustment factor for each race and ethnicity is based on national COVID-19 death risk ratios reported by the CDC[33,34]. While NNT_h represents the predicted number of vaccinations needed to prevent one hospitalization, NNT_a is not a direction calculation of NNT for death prevention; instead, its purpose is to illustrate how locations with similar NNT_h might be further stratified based on the predicted outcomes of those predicted hospitalizations. The resultant NNT_a favors locations with more Black/African American, Hispanic, and Latino populations. Future analytical and predictive efforts incorporating county-level death rates are warranted.

The heatmaps of NNT_h and NNT_a are shown in **Figure 7**. Note how the count and ratio maps identify hotspots primarily among urban and metropolitan areas. In contrast, the NNT values are generally lowest in the rural region, which are highly clustered in certain states, including North Dakota, South Dakota, Kansas, Nebraska, Montana, etc. Quantitatively, increasing the vaccination

419 ratio by 10% can reduce the number of current hospitalization cases by 15% on average for all
420 locations.

421 We further calculate the cost ratio between state average vaccination costs and state average
422 COVID-19 hospitalization costs. We collect the state-level average vaccine cost from the Centers
423 for Medicare and Medicaid Services (CMS)[35] and hospitalization cost for both complex and
424 noncomplex COVID-19 inpatients from the COVID-19 Cost Tracker[36]. The cost ratio is then
425 calculated as:

$$r_{\text{cost}} = \frac{r_{\text{complex}} * c_{\text{complex}} + r_{\text{noncomplex}} * c_{\text{noncomplex}}}{\text{NNT}_h * c_{\text{shot}}} \tag{6}$$

427 where r_{complex} and $r_{\text{noncomplex}}$ are the average complex and noncomplex COVID-19 inpatients
428 ratio extracted from the hospital resource data. c_{complex} and $c_{\text{noncomplex}}$ are hospitalization costs
429 for complex and noncomplex COVID-19 inpatients. c_{shot} is the cost of one vaccine. All
430 parameters to calculate the r_{cost} vary by location.

431 The cost ratio heatmap is shown in **Figure 7-(c)**. A cost ratio greater than 1 indicates that
432 increasing the number of vaccinations in that location by the NNT is less expensive than the cost of
433 the single prevented hospitalization. The heatmap demonstrates a ratio favoring future vaccination
434 efforts in primarily rural regions, most commonly among centrally located and landlocked counties.
435 We list the NNT_h and NNT_a of the top 15 locations with the highest (most favorable) cost ratios
436 in **Table 2**. The full county-level table can be accessed in the GitHub repository and the visualization
437 platform. These results suggest three meaningful conclusions. First, even small-scale vaccination
438 efforts in targeted counties are likely to prevent at least one hospitalization and its related sequelae.
439 Secondly, rural vaccination outreach efforts are likely cost-saving endeavors in 368 counties among
440 43 states (with cost ratio > 1). Thirdly, we find that locations that have high cost ratio also have high
441 adjust ratio. This may indicate large healthcare disparities for different race in these locations. More
442 life could be saved by improving the vaccination fairness in these locations.

Discussion

444 In this work, we propose an Ising dynamics-based deep learning model, HOIST, for spatio-
445 temporal COVID-19 hospitalization prediction. The HOIST model is built with multiple data
446 sources including disease and vaccination statistics from real-world medical claims, medical
447 resource usage, census, geographical and mobility data. HOIST can handle complex inner-region
448 influence factors and adaptively learn the inter-regional relationships without requiring a pre-
449 defined location graph with fixed edges. By drawing the analogy between locations and lattice sites,
450 we use the Ising dynamics as a regularizer to guide the model learning and prediction process, which
451 is a natural choice to model the inner-region factors and inter-region similarities simultaneously.

452 We evaluate the HOIST model on the large-scale spatio-temporal COVID-19 hospitalization
453 prediction task for 2,299 counties in the United States. In the 4-week hospitalization prediction task,
454 HOIST achieves 368.7 MAE, 0.6 R^2 and 0.89 CCC scores. For different lengths of prediction
455 window from 1 week to 5 weeks, HOIST achieves 48% lower MAE, 65% lower MSE, 272% higher
456 R^2 and 51% higher CCC on average compared to the best baseline models. We conducted a detailed
457 analysis of the model results and learned parameters. We find that the booster shot of vaccination
458 population percentage has a more significant negative correlation to future hospitalization cases
459 than the first and second vaccination shots. We also find that the effects of two major vaccination
460 brands, Pfizer and Moderna, have no statistical difference in our model.

461   We note that, in contrast to many previous models, our HOIST-based clinical and economic
462   predictions suggest a need to prioritize future vaccination efforts in rural areas over urban centers.
463   The results indicate that increasing the total vaccination ratio by 10% can make the number of
464   current hospitalization cases reduce by 15% on average for all locations. We believe these
465   predictions accurately reflect the growing disparity in vaccination rates between these communities.
466   Additionally, we note a positive correlation between vaccination rate and NNT; in fact, the adjusted
467   NNT map appears strikingly similar to maps depicting percentages of fully vaccinated residents by
468   county[37]. We surmise that communities with lower-than-average proportions of vaccinated
469   individuals are more likely to benefit from further vaccination efforts irrespective of rurality. For
470   368 counties among 43 states, increasing vaccination ratio in these counties can significantly reduce
471   the overall healthcare costs for COVID-19 patients. We hope these results and our model can inform
472   clinicians, healthcare institutions, and policymakers to improve their decisions and ultimately
473   reduce the negative economic and health impacts caused by the pandemic.

474   This research is not without limitations. In the data collection process, the number of
475   hospitalization cases may have bias. We define a hospitalization case as a patient hospitalized within
476   35 days after the COVID-19 diagnosis. We extract these statistics from the claim dataset, so we
477   cannot know if COVID-19 is the primary cause of hospitalization. Though this definition is the
478   same with existing COVID-19 prediction works[38,39], the model may conflate hospitalizations trends
479   due to other causes with those accurately attributed to COVID-19. This is especially true in the
480   pediatric population where COVID-19 infectivity is high but severe cases are uncommon. Despite
481   these limitations, we still select claims data as our primary data source as it provides valuable
482   granular county-level statistics.

483   The second limitation concerns data preprocessing. The original data distribution of
484   hospitalization cases is highly skewed. Our dataset is much larger than previous research; over 80%
485   of locations have fewer than 1,000 hospitalization cases in the 28-day prediction window, and only
486   2% have more than 10,000 hospitalization cases. To maintain the stability of the machine learning
487   models, we conduct a log transformation on the prediction target, followed by a z-score
488   transformation to normalize the sequence. Therefore, model prediction results require two reverse
489   transformations to return to the original scale. This transformation may cause a small error in the
490   prediction scale to grow exponentially as it converts to the original scale. As a result, we observe
491   that baseline models sometimes have abnormal large prediction errors and standard deviations (e.g.,
492   over 10,000 MAE), especially when the prediction window is long. We manually remove these
493   outliers from the performance table. Though we do not find these issues in HOIST, which further
494   demonstrates the stability of the model, we still believe this is an open research challenge in the
495   large-scale spatio-temporal prediction work. The data scale issue might be solved by proposing more
496   advanced preprocessing techniques or new scale-invariant models.

497   A third limitation exists at the methodology level. Though the spatial connectivity learning
498   module is more flexible than previous works requiring fixed graph structures, the learned
499   connectivity does not change with time. In real-world scenarios, spatial connectivity may be affected
500   by several factors, including weather, holidays, and travel restrictions. Therefore, future works may
501   include integrating more data sources such as real-time mobility data between locations to further
502   improve the extracted spatial patterns, interpretability analysis on the feature contributions and more
503   granular age information. Besides, the Ising dynamics maybe not the only choice to model the
504   spatial and social background factors for hospitalization prediction. The reason we choose the Ising

505 dynamics is that it can incorporate both inter-region spatial relationships and inner-region factors
506 and can handle these complex factors compared to naïve SIR and SEIR models. We are inspired by
507 previous sociology studies that use Ising models for human behavior tendencies[40]. Though the Ising
508 dynamics show good prediction performances in HOIST, we hope our exploratory work can provide
509 more inspiration for future studies using other real-world dynamics.

Methods

511   We follow the recommendations set out in the Global Code of Conduct for Research in
512 Resource-Poor Settings when designing, executing and reporting the research and this research does
513 not use individual-level data which may raise ethic issues. Our study complies with the
514 recommendations of the GATHER statement.

Background

516   Machine learning and deep learning models have been widely applied in pandemic predictions.
517 Statistical epidemic prediction models, including SIR, SEIR, and their variants, have achieved some
518 success in infection, hospitalization, and mortality prediction tasks[25,41,42]. To extract complex
519 temporal patterns, some works have applied the recurrent neural network (RNN) and its variants,
520 such as long short-term memory network (LSTM) and gated recurrent unit network (GRU), to
521 predict future infected and hospitalization cases[24-26].
522   To further improve the spatio-temporal prediction performance, a significant line of research
523 focuses on extracting and utilizing spatial dependencies. Graph neural networks (GNN) [4,27,43,44] and
524 metapopulation analysis models[45-47] have achieved remarkable success in solving this issue. In these
525 works, counties and states are modeled as nodes in the graph, and the edges are defined using
526 geographical and sociological similarities or mobility scores. By combining the GNNs and RNNs,
527 these models can extract spatio-temporal patterns in the data and make better predictions. However,
528 most of these models rely on pre-defined location graphs. These manually defined location graphs
529 may differ greatly from the ground truth inter-regional relationships and spatial pandemic
530 transmission patterns. Besides, some metapopulation models are only applicable for infectious
531 diseases since the underlying assumption is the population flow, which makes it difficult to extract
532 other spatial patterns such as hospitalizations. Some works aim to solve this issue by adaptively
533 learning the connection weights between locations. Jin et al. [8] and Deng et al. [9] utilize the attention
534 mechanism to predict future infected cases in pandemics. Their work learns the connection weights
535 based on the similarities of historical case curves at different locations but largely ignores the
536 underlying social, economic, medical, and geographical similarities, which may have a profound
537 impact on the inter-location relationships. Therefore, further exploration is needed to discern how
538 best to model spatial dependencies with multi-source background data in a flexible way.
539   Another line of research focuses on integrating real-world disease transmission dynamics into
540 deep learning models. Though deep learning networks can extract complex temporal patterns, they
541 can only predict known data patterns and thus have worse long-term prediction accuracy. Integrating
542 epidemiology models such as SIR and SEIR can help deep learning models predict curves
543 resembling real-world transmission patterns. Gao et al.[4] used an SIR model as a regularizer term in
544 the loss function to help the GAT-GRU model predict county-level COVID-19 infection case
545 numbers in the United States. Storlie et al.[6] proposed a stochastic SIR model to predict future
546 COVID-19 cases. Though these models can be applied in the COVID-19 hospitalization prediction,

547   their main approach involves dividing the infected population into hospitalized and non-hospitalized
548   cohorts and then modifying the transmission equation accordingly. However, as mentioned in the
549   introduction, the relatively few parameters in most epidemiological models may not fully describe
550   the influences of complex underlying disease distribution, viral variants, and vaccination effects.
551   Additionally, these epidemiological models are not designed for general spatio-temporal predictions
552   and thus cannot extract and utilize inter-regional interactions. Therefore, integrating appropriate
553   real-world dynamics that can naturally model spatio-temporal hospitalizations may improve the
554   prediction performance. The Ising model is a statistical mechanics model for site spin configuration
555   estimations; it has been applied in sociology research to model social behavior by considering each
556   individual as a site[40]. In this work, instead of directly using statistical methods to estimate the Ising
557   parameters in a low-dimensional and static-variable-only dataset, we combine the Ising dynamics
558   with deep learning techniques to model the spatio-temporal pandemic hospitalization patterns from
559   complex, high-dimensional multi-source data.

Learning adaptative connectivity in the latent space

561   Pandemic progression patterns have spatial similarity depending on the underlying population,
562   economic, geographic, and mobility factors in various locations. Recent works suggest that utilizing
563   these spatial similarities can guide the model to better extract progression patterns and make more
564   accurate predictions[4,27,43,44]. In this work, instead of pre-defining a connected location graph with
565   fixed edges, we aim to learn the adaptative connectivity between locations. Concretely, we measure
566   the latent distances between locations from multiple perspectives from the static data, and then
567   normalize the distances to similarities.

568   In HOIST, each location will be embedded in a multi-dimensional latent space. The latent space
569   is created by considering the set of socio-demographic variables as dimensions so that each location
570   will have a unique position, and locations that are similar in these socio-demographic factors will
571   be embedded closely in the space (also known as the Blau space in sociology study)[48]. To calculate
572   the distance of two locations in the latent space, previous works directly use the sum of Euclidean
573   distance over all dimensions[40]. This is applicable in small-scale settings, but our setting is large-
574   scale and high-dimensional, containing all counties in the United States. The scale differences in
575   some factors are substantial (e.g., population size and geographic distance). Using Euclidean
576   distance may lead to biased connectivity. Furthermore, we also aim to capture the complex nonlinear
577   similarity, which is difficult for scale-invariant distance metrics such as Mahalanobis distance. In
578   this work, we calculate the distances using the graph attention mechanism[49]. Concretely, for location
579   i and j, the latent distance l_{ij} is calculated as:

$$l_{ij} = \sum_{\forall \mathbf{K} \in \{\mathbf{M}, \mathbf{E}\}} \sigma(\mathbf{W}_{K,2}(\mathbf{W}_{K,1}\mathbf{k}_i || \mathbf{W}_{K,2}\mathbf{k}_j)) + \sigma(\mathbf{W}_{D,2}(\mathbf{W}_{D,1}(d_{ij} || g_{ij}))) \tag{7}$$

581   where $\mathbf{k}_i$ denotes the i-th row vector of the $\mathbf{K}$ matrix (i.e., the population demographics $\mathbf{M}$ and
582   the economics and healthcare statistics $\mathbf{E}$), d_{ij} and g_{ij} denote the value at i-th row and j-th
583   column in the matrices $\mathbf{D}$ and $\mathbf{G}$ (i.e., the distance between location i and j),
584    $\mathbf{W}_{K,2}, \mathbf{W}_{K,1}, \mathbf{W}_{D,1}, \mathbf{W}_{D,2}$ are weight matrices, σ denotes the sigmoid activation function, and $(\cdot || \cdot$
585    $)$ denotes the concatenation operation. Using this attention mechanism, the model can adaptatively
586   learn the distance between two locations from multiple perspectives including demographics,
587   populations, economics, healthcare, geographic distance, and mobility. The distances are further
588   normalized to obtain the similarity s_{ij} as:

$$s_{ij} = \frac{\exp(l_{ij})}{\sum_k^N \exp(l_{ik})} \qquad (8)$$

where s_{ij} indicates the normalized adaptative similarity between location i and j learned by the model. The higher the s_{ij}, the more similar the two locations. The similarity scores of all locations formulate the matrix $\mathbf{S} \in \mathbb{R}^{N \times N}$ which is used to inform how the model utilizes the learned spatial connectivity in the following sections.

Modeling the external fields in Ising dynamics

In the Ising model, the energy of a spin configuration $\mathbf{y}$ of lattice sites is given by the Ising Hamiltonian I, which takes the form:

$$I = -\sum_i \left(e_i y_i + \sum_{<i,j>} s_{ij} y_i y_j \right) \qquad (9)$$

where s_{ij} denotes the interaction between two sites i and j, and e_i denotes the external field (E.F.) interacting with site i. The Ising Hamiltonian I can be considered as the sum of the energy related to the interactions between sites and the energy related to the external field. This Ising Hamiltonian form is applied to capture pandemic hospitalization patterns. First, the hospitalization patterns are highly correlated with the location-specific distributions on comorbidity, vaccination, and hospital resource usage. This can be modeled as the first term in the Ising Hamiltonian $e_i y_i$. Secondly, the hospitalization patterns are highly correlated across locations depending on the underlying socioeconomic factors. The second term in the Ising Hamiltonian matches location-correlation. We propose to use the Ising dynamics to guide the model learning process in this manner.

To fuse this dynamic equation into the HOIST model, the initial step is to model the external fields using Ising dynamics. More specifically, we model the external fields of a specific location from multiple perspectives. The idea is to utilize historical statistics to infer future hospitalization cases. For example, if the number of infected cases among high-risk cohorts has risen in the past few months, the number of hospitalization cases is also likely to rise soon. This temporal delay is expected since many infected patients, even those at high risk for critical illness, do not develop severe manifestations immediately. Of course, we also take vaccinations into consideration; as the number of immunized individuals increases, the hospitalization rate decreases. All sequence vectors are flattened into vectors in a historical window τ instead of just the values at timestep t (e.g., $\mathbf{u}'_t = [\mathbf{u}_{t-\tau+1}, \mathbf{u}_{t-\tau+2}, \dots, \mathbf{u}_t]$). Likewise, we create the historical new infected case sequence $\mathbf{i}'_t$, claim statistics sequence $\mathbf{c}'_t$, vaccination statistics $\mathbf{v}'_t$, and medical resource usage sequence $\mathbf{u}'_t$ at timestep t. Note that we omit the location index i in this section because we are only discussing one specific location. The parameters of external fields are modeled as:

$$\mathbf{f}_I = \sigma(\phi_I(\mathbf{i}'_t)) \qquad (10)$$
$$\mathbf{f}_C = \sigma(\phi_C(\mathbf{c}'_t)) \qquad (11)$$
$$\mathbf{f}_V = \sigma(\phi_V(\mathbf{v}'_t)) \qquad (12)$$
$$\mathbf{f}_U = \sigma(\phi_U(\mathbf{u}'_t)) \qquad (13)$$

where $\phi: \mathbb{R}^F \to \mathbb{R}^F$ denotes the projection function to generate scaling factors for each sequential factor. By multiplying the scaling factors and the sequential factors, the final predictions are calculated as:

$$\hat{\mathbf{f}} = [\mathbf{f}_I \cdot \mathbf{i}'_t || \mathbf{f}_C \cdot \mathbf{c}'_t || - \mathbf{f}_V \cdot \mathbf{v}'_t || \mathbf{f}_U \cdot \mathbf{u}'_t] \qquad (14)$$
$$\hat{\mathbf{h}}_t = \text{LSTM}(\hat{\mathbf{f}}, \hat{\mathbf{h}}_{t-1}) \qquad (15)$$

629 Here LSTM($\cdot$) denotes a multi-layer perceptron to capture temporal progressions, the ($\cdot$)
630 operation denotes the Hadamard product and ($\cdot \, || \, \cdot$) denotes the concatenation operation. This
631 process is essentially learning weights (i.e., $\mathbf{f}_I$, $\mathbf{f}_C$, $\mathbf{f}_V$ and $\mathbf{f}_U$) for different influence factors. The
632 larger the weight, the greater the influence of the corresponding factors on the prediction result. This
633 is also why we add the negative symbol before $\mathbf{f}_V$, since the vaccinations should be negatively
634 correlated with hospitalization cases. By concatenating all the scaling factors, we obtain the E.F.
635 strength vector $\hat{\mathbf{f}}$. We further use an LSTM network to model the temporal progression of the E.F.
636   To generate the final predictions for location i (here we bring back the location subscript as
637 $\hat{\mathbf{h}}_{i,t}$), we calculate the weighted sum of the final hidden state from the LSTM network from other
638 locations. We use the adaptative connectivity $\mathbf{S}$ as the connection weight as:

$$\mathbf{z}_{i,t} = \hat{\mathbf{h}}_{i,t} + \sum_{j}^{N} s_{ij} \hat{\mathbf{h}}_{j,t} \tag{16}$$

$$\hat{y}_{i,t} = \text{MLP}(\mathbf{z}_{i,t}) \tag{17}$$

641 In this way, the final prediction result $\hat{y}_{i,t}$ considers not only the temporal progression of the
642 pandemic and influence factors (Equation 3), but also the relevant spatial relationships (Equation 2
643 and 4).

End-to-end learning with the spatio-temporal Ising loss function

645   Our loss function consists of a prediction loss and an Ising dynamics constraint loss. First, the
646 prediction loss is calculated by using the mean square error:

$$\mathcal{L}_p = \sum_{i}^{N} \sum_{t}^{T} (\hat{y}_{i,t} - y_{i,t})^2 \tag{18}$$

648   The second loss term is the Ising constraint loss $\mathcal{L}_d$, which takes the form:

$$e_i = \sum_{k}^{|\hat{f}_i|} \hat{f}_{i,k} \tag{19}$$

$$\hat{y}_{i,t}^d = e_i \hat{y}_{i,t} + \sum_{j}^{N} s_{ij} \hat{y}_{i,t} \hat{y}_{j,t} \tag{20}$$

$$\mathcal{L}_d = \sum_{i}^{N} \sum_{t}^{T} (\hat{y}_{i,t}^d - y_{i,t})^2 \tag{21}$$

652 The term e_i denotes the summation over all dimensions of $\hat{\mathbf{f}}_i$, and $\hat{f}_{i,k}$ denotes the k-th
653 dimension in the $\hat{\mathbf{f}}_i$ vector, which turns the E.F. vector into a scalar, denoting the E.F. strength. s_{ij}
654 is the value at the i-th row, j-th column in the $\mathbf{S}$ matrix, which is the similarity score between
655 location i and j. $\mathcal{L}_d$ is essentially another mean square error loss between the ground truth value
656 and the estimated value using the Ising dynamics (i.e., $\hat{y}_{i,t}^d$). In the original Ising dynamics setting,
657 the system's energy should be as low as possible to keep the system stable. Here, we use the same
658 idea to reduce the prediction error $\mathcal{L}_d$ as part of our loss function. It is worth noting that $\hat{y}_{i,t}$ is the
659 final prediction result and $\hat{y}_{i,t}^d$ is only used as an auxiliary output to optimize the Ising constraint
660 parameters. The final loss takes the form:

$$\mathcal{L} = \mathcal{L}_d + \mathcal{L}_p \tag{22}$$

By optimizing the Ising constraint loss and the prediction loss together, the model can extract the temporal hospitalization progression patterns as well as optimize the spatial adaptive connectivity. Using the Ising dynamics as a constraint instead of directly using $\hat{y}^d$ as the prediction result and only optimizing $\mathcal{L}_d$ allows the model to have more flexibility to learn spatio-temporal patterns from high-dimensional data.

Data Availability

The mobility scores are collected from the Multiscale Dynamic Human Mobility Flow Dataset[11] (https://github.com/GeoDS/COVID19USFlows). The census features are collected from the county-level census dataset provided by the US County-level Dataset[12] (https://github.com/JieYingWu/COVID-19_US_County-level_Summaries). The daily new cases are collected from the Johns Hopkins COVID-19 Data Repository[13] (https://github.com/CSSEGISandData/COVID-19). The medical resource usage statistics are collected from HealthData.gov[17] (https://healthdata.gov/Hospital/COVID-19-Reported-Patient-Impact-and-Hospital-Capa/anag-cw7u). All the processed public data are available at https://github.com/v1xerunt/HOIST. The claims data contains sensitive healthcare information and are extracted from https://www.iqvia.com/solutions/real-world-evidence/, which can be accessed on request. Our model can be trained without using claims data. Source data of tables and figures are provided with this paper.

Code Availability

The codes for model construction, training and inference used in this paper are publicly available at https://github.com/v1xerunt/HOIST. The visualization results are available at https://v1xerunt.github.io/HOIST/.

References

1. Centers for Disease Control and Prevention. CDC COVID Data Tracker. Accessed May 01, 2022. https://covid.cdc.gov/covid-data-tracker/#datatracker-home.

2. Hassan, E.M. & Mahmoud, H.N. Impact of multiple waves of COVID-19 on healthcare networks in the United States. *PloS one* **16**, e0247463 (2021).

3. Centers for Disease Control and Prevention. Disparities in COVID-19 Vaccination Coverage Between Urban and Rural Counties. Accessed May 2022. https://www.cdc.gov/mmwr/volumes/71/wr/mm7109a2.htm.

4. Gao, J., Sharma, R., Qian, C.*, et al.* STAN: spatio-temporal attention network for pandemic prediction using real-world evidence. *Journal of the American Medical Informatics Association* **28**, 733-743 (2021).

5. Kargas, N., Qian, C., Sidiropoulos, N.D.*, et al.* Stelar: Spatio-temporal tensor factorization with latent epidemiological regularization. in *the 2021 AAAI Conference on Artificial Intelligence*, Vol. 35 4830-4837 (2021).

6. Storlie, C.B., Rojas, R.L., Demuth, G.O.*, et al.* A Hierarchical Bayesian Model for Stochastic Spatiotemporal SIR Modeling and Prediction of COVID-19 Cases and Hospitalizations. *arXiv preprint arXiv:2104.04033* (2021).

7. Fan, C., Meng, Y., Sun, X.*, et al.* Parameter Estimation for the SEIR Model Using Recurrent Nets. *arXiv preprint arXiv:2105.14524* (2021).

703 8. Jin, X., Wang, Y.-X. & Yan, X. Inter-series attention model for covid-19 forecasting. in *the 2021*
704   *SIAM International Conference on Data Mining (SDM)* 495-503 (SIAM, 2021).
705 9. Deng, S., Wang, S., Rangwala, H., Wang, L. & Ning, Y. Cola-GNN: Cross-location attention
706   based graph neural networks for long-term ILI prediction. in *the 29th ACM International*
707   *Conference on Information & Knowledge Management* 245-254 (2020).
708 10. Cipra, B.A. An introduction to the Ising model. *The American Mathematical Monthly* **94**, 937-
709   959 (1987).
710 11. Kang, Y., Gao, S., Liang, Y.*, et al.* Multiscale dynamic human mobility flow dataset in the US
711   during the COVID-19 epidemic. *Scientific data* **7**, 1-13 (2020).
712 12. Killeen, B.D., Wu, J.Y., Shah, K.*, et al.* A county-level dataset for informing the United States'
713   response to COVID-19. *arXiv preprint arXiv:2004.00756* (2020).
714 13. Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real
715   time. *The Lancet infectious diseases* **20**, 533-534 (2020).
716 14. Centers for Disease Control and Prevention. People with Certain Medical Conditions. Accessed
717   Aug 1, 2022. https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-
718   with-medical-conditions.html.
719 15. Sundararajan, V., Henderson, T., Perry, C.*, et al.* New ICD-10 version of the Charlson
720   comorbidity index predicted in-hospital mortality. *Journal of clinical epidemiology* **57**, 1288-
721   1294 (2004).
722 16. IQVIA. Harness the power of Real World Data. Accessed Aug 1, 2022.
723   https://www.iqvia.com/solutions/real-world-evidence/real-world-data-and-insights.
724 17. HealthData. COVID-19 Reported Patient Impact and Hospital Capacity by State. Accessed Aug
725   1, 2022. https://healthdata.gov/dataset/COVID-19-Reported-Patient-Impact-and-Hospital-
726   Capa/6xf2-c3ie/data.
727 18. Gao, J., Yang, C., Heintz, J.*, et al.* MedML: Fusing medical knowledge and machine learning
728   models for early pediatric COVID-19 hospitalization and severity prediction. *Iscience* **25**,
729   104970 (2022).
730 19. The BARDA Community Challenge. BARDA Community Challenge - Pediatric COVID-19
731   Data Challenge. Accessed May 04, 2022.
732   https://www.synapse.org/#!Synapse:syn25875374/wiki/.
733 20. Lin, Z., Trivedi, S. & Sun, J. Conformal Prediction Intervals with Temporal Dependence. *arXiv*
734   *preprint arXiv:2205.12940* (2022).
735 21. Paszke, A., Gross, S., Massa, F.*, et al.* Pytorch: An imperative style, high-performance deep
736   learning library. in *Advances in neural information processing systems*, Vol. 32 (2019).
737 22. Li, M.L., Bouardi, H.T., Lami, O.S.*, et al.* Forecasting COVID-19 and analyzing the effect of
738   government interventions. *Operations Research* (2022).
739 23. Byrd, R.H., Gilbert, J.C. & Nocedal, J. A trust region method based on interior point techniques
740   for nonlinear programming. *Mathematical programming* **89**, 149-185 (2000).
741 24. Kavouras, I., Kaselimi, M., Protopapadakis, E.*, et al.* COVID-19 Spatio-Temporal Evolution
742   Using Deep Learning at a European Level. *Sensors* **22**, 3658 (2022).
743 25. Yang, Z., Zeng, Z., Wang, K.*, et al.* Modified SEIR and AI prediction of the epidemics trend of
744   COVID-19 in China under public health interventions. *Journal of thoracic disease* **12**, 165
745   (2020).
746 26. Nikparvar, B., Rahman, M., Hatami, F. & Thill, J.-C. Spatio-temporal prediction of the COVID-

747   19 pandemic in US counties: modeling with a deep LSTM neural network. *Scientific reports* **11**,
748   1-12 (2021).

749   27.   Kapoor, A., Ben, X., Liu, L.*, et al.* Examining covid-19 forecasting using spatio-temporal graph
750   neural networks. *arXiv preprint arXiv:2007.03113* (2020).

751   28.   Callaway, E. Fast-spreading COVID variant can elude immune responses. *Nature* **589**, 500-502
752   (2021).

753   29.   Eyre, D.W., Taylor, D., Purver, M.*, et al.* Effect of Covid-19 vaccination on transmission of
754   alpha and delta variants. *New England Journal of Medicine* **386**, 744-756 (2022).

755   30.   Kuswardhani, R.T., Henrina, J., Pranata, R.*, et al.* Charlson comorbidity index and a composite
756   of poor outcomes in COVID-19 patients: A systematic review and meta-analysis. *Diabetes &*
757   *Metabolic Syndrome: Clinical Research & Reviews* **14**, 2103-2109 (2020).

758   31.   Christensen, D.M., Strange, J.E., Gislason, G.*, et al.* Charlson comorbidity index score and risk
759   of severe outcome and death in Danish COVID-19 patients. *Journal of general internal*
760   *medicine* **35**, 2801-2803 (2020).

761   32.   Tenforde, M.W., Self, W.H., Adams, K.*, et al.* Association between mRNA vaccination and
762   COVID-19 hospitalization and disease severity. *Jama* **326**, 2043-2054 (2021).

763   33.   Centers for Disease Control and Prevention. Risk for COVID-19 Infection, Hospitalization, and
764   Death By Age Group. Accessed Dec 2022. https://www.cdc.gov/coronavirus/2019-ncov/covid-
765   data/investigations-discovery/hospitalization-death-by-age.html.

766   34.   Centers for Disease Control and Prevention. Risk for COVID-19 Infection, Hospitalization, and
767   Death By Race/Ethnicity. Accessed Dec 2022. https://www.cdc.gov/coronavirus/2019-
768   ncov/covid-data/investigations-discovery/hospitalization-death-by-race-ethnicity.html.

769   35.   Services, C.f.M.M. Medicare COVID-19 Vaccine Shot Payment. Accessed Dec 2022.
770   https://www.cms.gov/medicare/covid-19/medicare-covid-19-vaccine-shot-payment.

771   36.   FAIRHealth. COVID-19 Cost Tracker. Accessed Dec 2022. https://www.fairhealth.org/states-
772   by-the-numbers/covid19-heatmap.

773   37.   The New York Times. COVID-19 Vaccinations: County and State Tracker. Accessed Dec 2022.
774   https://www.nytimes.com/interactive/2020/us/covid-19-vaccine-doses.html.

775   38.   Gao, J., Yang, C., Heintz, G.*, et al.* MedML: Fusing Medical Knowledge and Machine Learning
776   Models for Early Pediatric COVID-19 Hospitalization and Severity Prediction. *arXiv preprint*
777   *arXiv:2207.12283* (2022).

778   39.   Bennett, T.D., Moffitt, R.A., Hajagos, J.G.*, et al.* Clinical characterization and prediction of
779   clinical severity of SARS-CoV-2 infection among US adults using data from the US National
780   COVID Cohort Collaborative. *JAMA network open* **4**, e2116901-e2116901 (2021).

781   40.   Godoy-Lorite, A. & Jones, N.S. Inference and influence of network structure using snapshot
782   social behavior without network data. *Science Advances* **7**, eabb8762 (2021).

783   41.   Pei, S. & Shaman, J. Initial simulation of SARS-CoV2 spread and intervention effects in the
784   continental US. *MedRxiv* (2020).

785   42.   Albani, V.V., Velho, R.M. & Zubelli, J.P. Estimating, monitoring, and forecasting COVID-19
786   epidemics: a spatiotemporal approach applied to NYC data. *Scientific Reports* **11**, 1-15 (2021).

787   43.   Panagopoulos, G., Nikolentzos, G. & Vazirgiannis, M. Transfer graph neural networks for
788   pandemic forecasting. in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35
789   4838-4845 (2021).

790   44.   Skianis, K., Nikolentzos, G., Gallix, B., Thiebaut, R. & Exarchakis, G. Predicting COVID-19

791   positivity and hospitalization with multi-scale graph neural networks. *Preprint* (2022).

792   45.   Colizza, V. & Vespignani, A. Epidemic modeling in metapopulation systems with heterogeneous

793   coupling pattern: Theory and simulations. *Journal of theoretical biology* **251**, 450-467 (2008).

794   46.   Calvetti, D., Hoover, A.P., Rose, J. & Somersalo, E. Metapopulation network models for

795   understanding, predicting, and managing the coronavirus disease COVID-19. *Frontiers in*

796   *Physics* **8**, 261 (2020).

797   47.   Coletti, P., Libin, P., Petrof, O.*, et al.* A data-driven metapopulation model for the Belgian

798   COVID-19 epidemic: assessing the impact of lockdown and exit strategies. *BMC infectious*

799   *diseases* **21**, 1-12 (2021).

800   48.   McPherson, M. A Blau space primer: prolegomenon to an ecology of affiliation. *Industrial and*

801   *Corporate Change* **13**, 263-280 (2004).

802   49.   Veličković, P., Cucurull, G., Casanova, A.*, et al.* Graph attention networks. *arXiv preprint*

803   *arXiv:1710.10903* (2017).

Acknowledgements

Author Contributions Statement

Junyi Gao processed the feature and built the model. George Heintz, Dr. Jimeng Sun, and Dr. Adam Cross provided technical and clinical guidance. All authors participated in report writing. All authors declare that they have no conflicts of interest.

Competing Interests Statement

We declare that none of the authors have competing financial or non-financial interests.

Tables

Table 1 Performance of 28-day hospitalization prediction. The performance numbers are mean (std). The bold values denote the best results. The asterisk * denotes the performance differences between HOIST and the best baseline models (ACTS) are significant based on the two-sided t-test results (p<0.001). Source data are provided as a Source Data file.

Model	MSE ($\times 10^5$)	MAE	R^2	CCC
DELPHI-SEIR	400.3 (-)	1,647 (-)	<-1	0.02 (-)
GRU	149.2 (6.6)	897.7 (20.9)	-0.04 (0.04)	0.27 (0.20)
LSTM	133.9 (6.9)	826.6 (30.0)	0.06 (0.04)	0.48 (0.10)
ColaGNN	138.5 (9.5)	801.9 (32.1)	0.10 (0.13)	0.52 (0.09)
ACTS	127.9 (14.1)	732.1 (45.3)	0.19 (0.15)	0.60 (0.10)
CovidGNN	181.0 (46.2)	1,048.0 (126.2)	-0.26 (0.18)	0.18 (0.16)

STAN	148.3 (30.7)	948.2 (68.4)	-0.06 (0.46)	0.18 (0.12)
HOIST-Risk	65.2 (14.7)	530.6 (59.5)	0.41 (0.21)	0.71 (0.08)
HOIST-Vaccination	62.1 (13.2)	515.3 (42.3)	0.45 (0.24)	0.78 (0.07)
HOIST-AC	73.4 (38.4)	562.5 (87.9)	0.27 (0.30)	0.68 (0.15)
HOIST-Ising	135.0 (15.4)	735.5 (28.6)	0.17 (0.14)	0.39 (0.07)
HOIST	**38.5 (10.2)***	**368.7 (18.7)***	**0.60 (0.16)***	**0.89 (0.02)***
p-value	3e-6	2e-7	1e-5	7e-4

824

825

826 **Table 2** NNT for hospitalization and death of the top 15 locations ranked by the cost ratio. Source data
827 are provided as a Source Data file.

County	State	Hospitalization NNT	Adjusted NNT	Cost ratio
Butte	Idaho	130	73	60.3
Woodbury	Iowa	76	44	34.5
Toole	Montana	230	130	34.5
Sheridan	Montana	204	115	34.0
Broadwater	Montana	186	109	33.7
Mineral	Nevada	273	159	32.1
Garfield	Washington	156	89	31.8
Stonewall	Texas	71	42	31.8
Lincoln	Nevada	201	114	31.2
Stanton	Kansas	101	67	29.2
Oneida	Idaho	287	163	28.7
Bowman	North Dakota	142	81	28.1
Phillips	Montana	234	132	27.4
Ness	Kansas	145	84	25.2
Teton	Montana	308	172	25.0

828

829 # Figure Captions

830 **Figure 1** The proposed HOIST model.
831 We use static data to calculate the latent distances between locations. The latent distances are then used
832 to calculate the adaptative connectivity of the location graph. The dynamic data are used to estimate the
833 External Fields (E.F.s) and then generate predictions using the LSTM network. We use the Ising
834 dynamics to regularize the spatio-temporal prediction results using the estimated E.F.s and the
835 adaptative connectivity.

836

837 **Figure 2** Prediction performance plots under different lengths of prediction window L.
838 Data are presented as mean values +/- standard deviations. The error bars are standard deviations over
839 5 experiments with random initializations with n=2,299 locations. Source data are provided as a Source
840 Data file.

841

842 **Figure 3** Prediction performance under different time split in MSE (A), MAE (B) and CCC (C).
843 Each dot denotes a testing phase. Source data are provided as a Source Data file.

844

Figure 4 Temporal change of external field weights for vaccination features.

Source data are provided as a Source Data file.

847

Figure 5 Predicted hospitalization hotspot and error map by HOIST.

a. Predicted hospitalization hotspot map in future 28 days by HOIST, colored by case count. West coast, Northeast coast and Midwest areas generally have high hospitalization case counts. **b**. Predicted hospitalization hotspot map in future 28 days by HOIST, colored by hospitalization ratio. We find two example locations San Diego county (A) and Riverside county (B) along the West coast – have very high case counts but only medium-level hospitalization ratios. **c**. Prediction errors for locations with more than 100 cases in the testing time window, colored by mean average percentage error. HOIST achieves low prediction errors on most locations. Source data are provided as a Source Data file.

856

Figure 6 Predicted hospitalization and infected case count curves for Effingham, Georgia.

a. Predicted curve by HOIST. The line denotes the mean value and the shadowed area denotes the prediction interval with 90% confidence. **b**. Curve of infected case count. Source data are provided as a Source Data file.

861

Figure 7 County-level NNT and cost ratio heatmap.

a. Predicted NNT heatmap by HOIST, colored by NNT_h. Red locations indicate that only few vaccine shots can help reduce a hospitalization case. **b**. Race-adjusted NNT heatmap by HOIST, colored by NNT_a. The spatial pattern is similar to the unadjusted map but some locations are affected by the race percentage. **c**. Heatmap of cost ratio between hospitalization costs and vaccination costs. Red locations indicate that they can save healthcare costs by giving more vaccinations (i.e., cost ratio > 1). Source data are provided as a Source Data file.

869

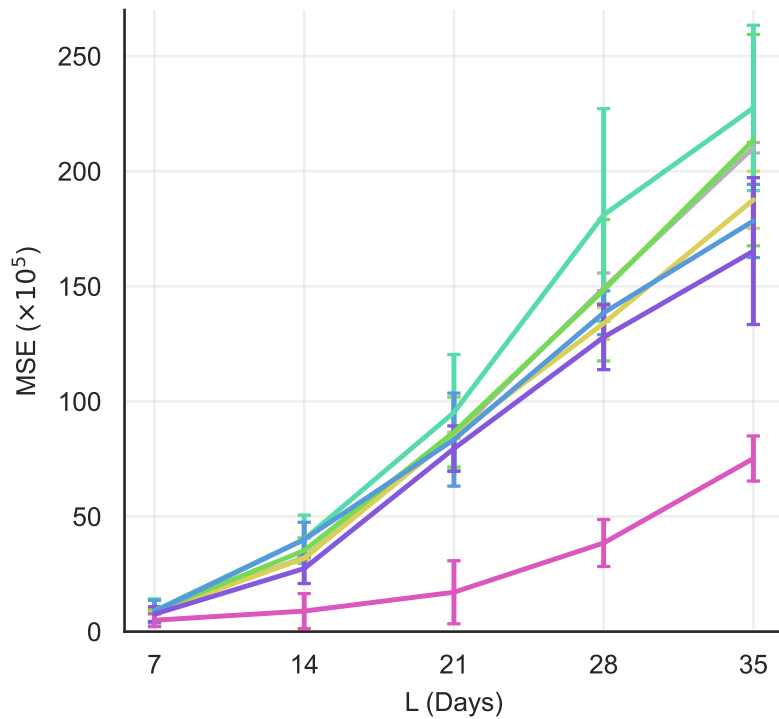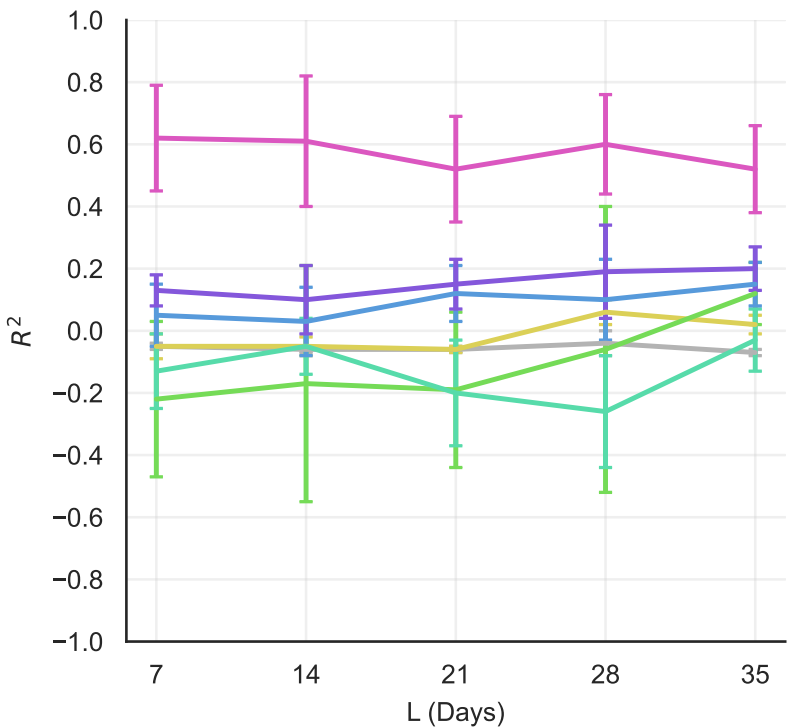**Static Data**

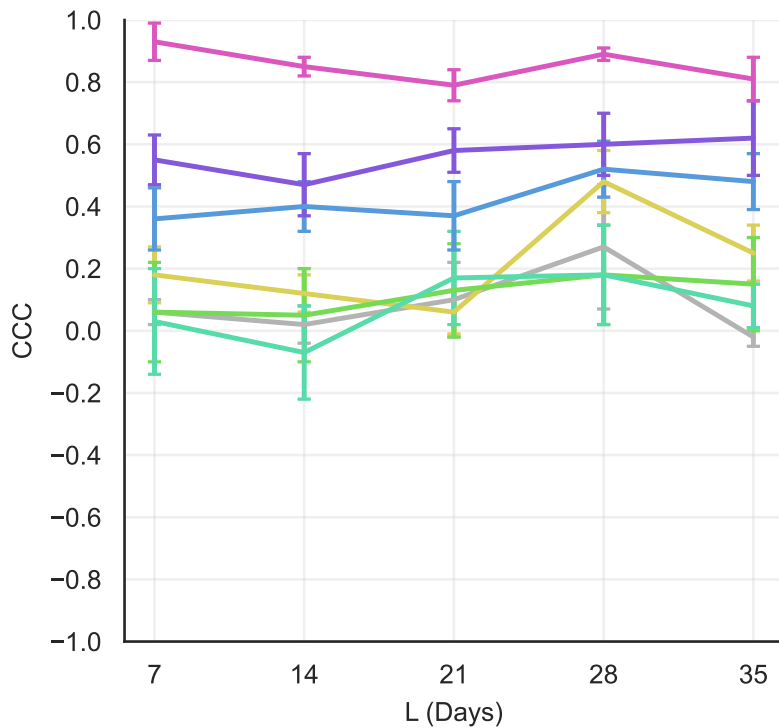Distance Matrices **D**, **G** Demograhpics **M**

Healthcare & Economics **E**

Latent Similarities

Adaptative Connectivity

Ising Dynamics

S

LSTM

$\hat{y}$

$\mathcal{L}_d$ Ising Dynamics Loss

$\mathcal{L}_p$ Prediction Loss

External Fields

$\mathbf{f}_I$ $\mathbf{f}_C$ $\mathbf{f}_V$ $\mathbf{f}_U$

Infected Cases **I** Medical Claims **C** Vaccinations **V** Medical Resource Usages **U**
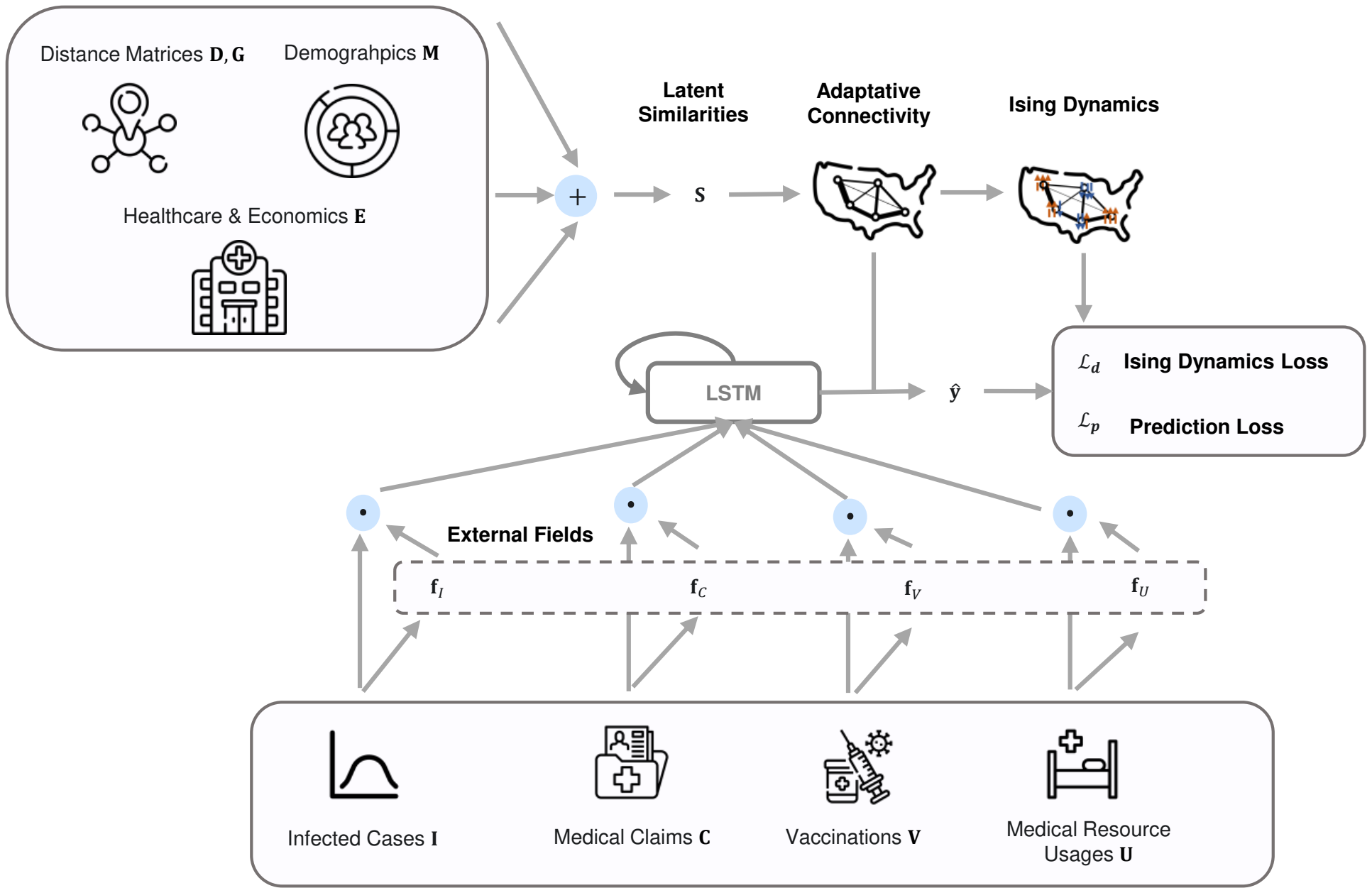
Dynamic Data
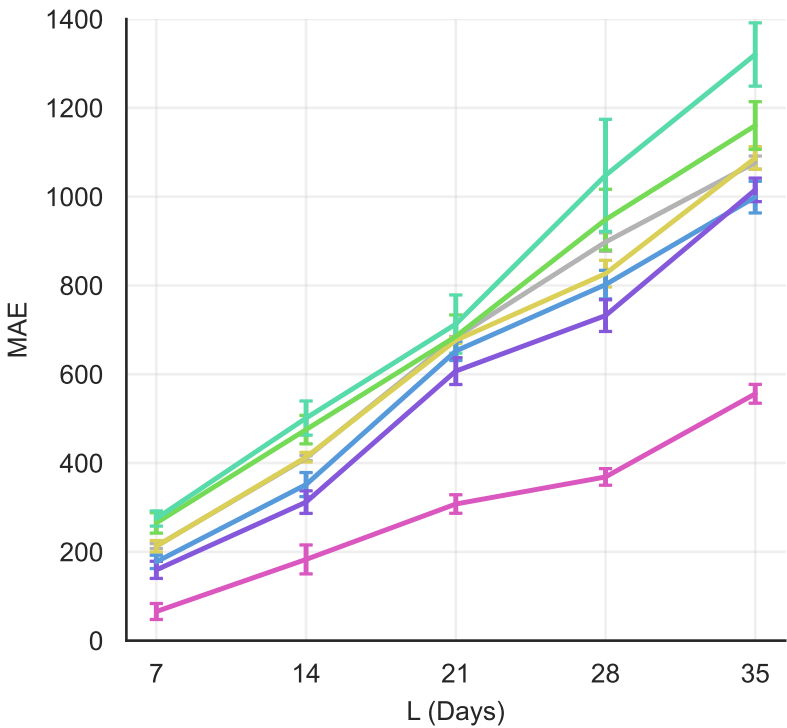
(a). MAE

(b). MSE

(c). R^2

(d). CCC

(A). MSE

(B). MAE

(C). CCC