HG-DCM: History Guided Deep Compartmental Model for Early Stage Pandemic Forecasting

Anonymous authors
Paper under double-blind review

Abstract

We introduce the History-Guided Deep Compartmental Model (HG-DCM), a novel framework for early-stage pandemic forecasting that synergizes deep learning with compartmental modeling to harness the strengths of both interpretability and predictive capacity. HG-DCM employs a Residual Convolutional Neural Network (CNN) to learn temporal patterns from historical and current pandemic data while incorporating epidemiological and demographic metadata to infer interpretable parameters for a compartmental model to forecast future pandemic growth. Experimental results on early-stage COVID-19 forecasting tasks demonstrate that HG-DCM outperforms both standard compartmental models (e.g., DELPHI) and standalone deep neural networks (e.g., CNN) in predictive accuracy and stability, particularly with limited data. By effectively integrating historical pandemic insights, HG-DCM offers a scalable approach for interpretable and accurate forecasting, laying the groundwork for future real-time pandemic modeling applications.

1 Introduction

Pandemics have historically caused catastrophic losses, from the Bubonic Plague in the 14th century (McEvedy) to the smallpox outbreak in the 18th century (Eyler), and most recently, the COVID-19 pandemic in 2020 (Holshue et al.). Despite significant advances in medical science, technology, and epidemiology, COVID-19 alone resulted in millions of deaths worldwide from 2020 to 2023. Accurate early-stage estimation of pandemic severity remains a crucial topic - Studies suggest that with improved forecasting and prompt interventions, early pandemic mortality could be reduced by as much as 90% (Piovani et al.; Li et al.). Yet accurate early-warning prediction is fundamentally challenging, with the lack of high-quality data being a major challenge. Mispredictions of pandemic severity lead to significant consequences: underestimating an outbreak risks overwhelming healthcare systems and delaying crucial interventions, thereby increasing mortality and transmission rates. Conversely, overestimations can lead to inefficient use of resources and societal disruptions, including panic buying (Islam et al.; Chua et al.) and social unrest (Barnard et al.; Reicher & Stott).

A significant number of current pandemic forecasting models are compartmental models, in which the incidence of each location is fit separately and completely relies on data specific to the current outbreak. The limited data source of compartmental models leads to unsatisfactory performance on early pandemic forecasting tasks. Past pandemics can provide significant information on the likely severity of the current pandemic at the early stage, but compartmental models lack the ability to integrate past pandemic information into forecasting. The wealth of historical pandemic data, which, though costly in terms of human lives, remains underutilized and represents a missed opportunity to enhance predictive accuracy. Therefore, in this study, we present the History-Guided Deep Compartmental Model (HG-DCM), which leverages historical data and meta-data to enhance forecasting accuracy by incorporating insights from previous pandemics and early-stage pandemic meta-data.

HG-DCM combines a residual convolutional neural network (He et al.) with a novel compartmental model DELPHI (Li et al.) to create a powerful tool for early pandemic warning. The neural network within HG-DCM allows cross-learning among different pandemics and different locations when fitting the DELPHI model, incorporating data from prior pandemics and metadata to improve incidence curve fitting. This approach preserves the interpretability and epidemiological grounding of the DELPHI model while leveraging historical data through neural network guidance to improve early-stage pandemic forecasting accuracy.

We applied HG-DCM to early COVID-19 forecasting across 258 locations globally, demonstrating that it consistently outperforms the original DELPHI model in early-stage COVID-19 forecasting. This study provides strong evidence that integrating historical data into compartmental models through neural networks can significantly enhance the accuracy and stability of early pandemic forecasting. Furthermore, our comparative analysis reveals that HG-DCM surpasses both state-of-the-art deep learning-based models and compartmental models in early-case forecasting tasks.

To our knowledge, this is the first study to develop a forecasting framework that leverages data from multiple prior pandemics to predict the trajectory of a newly emerging pandemic. While previous work has borrowed parameter priors from earlier outbreaks (Tindale et al.) or transferred models between related epidemics (Roster et al.), no prior research has systematically integrated information across different pandemics to improve forecasting during the emergence of a new one. As part of this effort, we constructed a new pandemic dataset including time-series case and death data, along with associated pandemic- and country-level meta-data, from major global outbreaks since 1990, including COVID-19, Ebola, SARS, dengue, monkeypox, and seasonal influenza.

1.1 Literature Review

Compartmental models have been used to forecast the trend of pandemics since the early 20th century. Starting with the simplest SIR (Susceptible, Infectious, Removed) model (Ross), various compartmental models with different states have shown satisfactory performance in forecasting seasonal pandemics (Schlickeiser & Kröger; KERR). One of the core strengths of compartmental models is their high interpretability - each parameter in a compartmental model usually corresponds to a physical quantity, which provides valuable insights into the pandemic. However, compartmental models also have limitations. Given the inevitable noisiness of the data, compartmental models can significantly overfit during the earliest stage of the pandemic when limited data is available. Furthermore, since compartmental models are inherently modeled for a pandemic in a certain area, it is also not obvious how to incorporate information from other pandemics to augment a compartmental model.

From another direction, machine learning is widely used in time-series forecasting fields such as stock prediction, weather forecasting, tourism (Law et al.), etc. However, most machine learning time-series models are not designed for early-stage pandemic prediction. There are attempts to use advanced deep learning models for pandemic forecasting (Rodríguez et al.; Tariq & Ismail; Devaraj et al.), but these models have been limited to modeling a single pandemic within a single region. Furthermore, these models suffer from the lack of interpretability, which makes the resulting predictions difficult to understand, especially during the early phase of a pandemic.

Overall, there have been few attempts to combine compartmental and deep learning models (Janssen et al.). Recently, there has been some research that integrates mobility data into the compartmental model through deep learning (Deng & Wang) or utilizes deep learning to estimate the time-varying parameter for the compartmental model (Ning et al.). However, these models assume that a significant amount of training data is available for the current pandemic, which makes it unsuitable for early-stage pandemic forecasting.

2 Methods

2.1 Model Construction

We introduce the History Guided Deep Compartmental Model (HG-DCM), which integrates a deep neural network with a compartmental model to combine the expressivity of a deep learning model and the interpretability of a compartmental model. The model architecture is defined as:

$$\hat{\theta} = f(T, M) \tag{1}$$

$$\hat{y} = h(\hat{\theta}). \tag{2}$$

Here T and M are the time-series and the metadata for the pandemic, which is combined through a deep learning model $f(\cdot)$ to create predictions $\hat{\theta}$ for the parameters for the compartmental model.

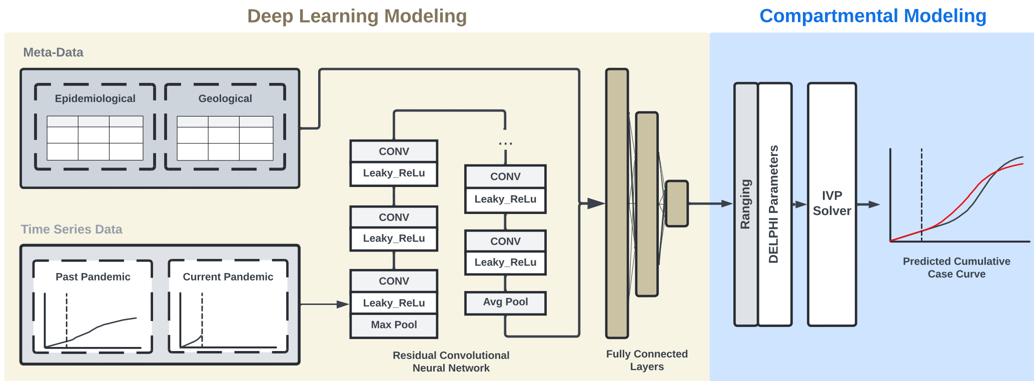

Figure 1: **Model Architecture of HG-DCM** HG-DCM consists of two parts: deep learning modeling and compartmental modeling. The deep learning modeling part predicts the compartmental model parameters, and the compartmental modeling section uses the predicted parameters to construct the predicted cumulative case curve for the pandemic.

The final predictions $\hat{y}$ are then calculated by solving an Initial Value Problem (IVP) to map the predicted parameters to a cumulative incidence curve. The key idea is that different pandemics could share a mapping between how the pandemic behaves (T, M) and the underlying parameters θ, which is captured by the deep learning model f. A graphical illustration of the model architecture is showcased in Figure 1. In the following paragraphs, we detail each of the specific structures.

Residual Convolutional Neural Network (CNN) We use a ResNet architecture to predict pandemic parameters from time-series data across past and current pandemics. Because the number of available pandemics is limited, relying on raw historical data is insufficient for training. To address this, we design two complementary data augmentation strategies: window-shift augmentation for past pandemics and masking augmentation for the current pandemic.

WINDOW-SHIFT AUGMENTATION (PAST PANDEMICS) For each historical pandemic, we generate additional training samples by shifting the start of the input time series forward one day at a time. The shifting stops when the start date reaches the peak of the first epidemic wave, which we identify as follows: Compute daily incidence from cumulative counts and smooth with a centered 7-day rolling average. Detect the first prominent peak that exceeds 25% of the global maximum using find_peaks (Scipy v1.12.0) and mark the subsequent local minimum as the end of the first wave. Define the last day of augmentation (LFoA) as the day of maximum daily incidence within this interval. Locations without a detectable first wave are excluded to ensure reliability. A graphical illustration of these three steps is shown in Figure 2.

MASKING AUGMENTATION (CURRENT PANDEMIC) Because future observations are unavailable for an unfolding pandemic, we apply a masking strategy instead. Specifically, we randomly replace consecutive 7-day segments of the observed input sequence with zeros, forcing the model to learn robust temporal patterns even when data are partially missing.

To account for variations in case numbers across locations, daily case numbers are log-transformed to enhance model stability. For time series with weekly reporting frequencies or missing data, linear interpolation is used for input data during training. We did not apply interpolation when evaluating model performance. Since the collected pandemic case data occasionally contained noisy entries with negative daily case counts, we set these values to zero to prevent spurious effects on the model. In addition, we excluded locations with fewer than 12 consecutive weeks of data where the cumulative number of cases was above 100. The ResNet input dimension is $[L, N, D]$, where L represents the lengths of training windows, N is the batch size, and D is the number of input features (e.g., daily cases, daily deaths). Due to limited data availability, only case numbers are used in this study. We also modify the ResNet implementation by removing batch normalization, as differences in batch statistics between past and current pandemics can lead to unstable predictions.

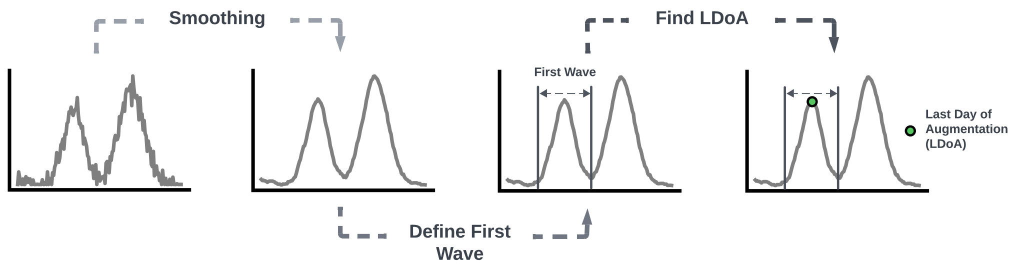

Figure 2: **Data Augmentation Methods** Window shift data augmentation method for past pandemic time series data

Fully Connected Layers The learned embeddings of the time-series data are concatenated with epidemiological metadata (e.g., transmission pathways) and demographic metadata (e.g., healthcare expenditure). A full table of the metadata is provided in A.1. Meta-data are normalized using min-max normalization to the range [0, 1], and passed through two fully connected layers before concatenating with the time series embeddings output from the CNN. The concatenated embeddings are passed through fully connected layers to produce parameters for the DELPHI model. To ensure that the produced parameters lie within physical bounds, we utilize a sigmoid ranging function to normalize the predicted parameters between 0 and 1.

Compartmental Modeling We utilize DELPHI (Li et al.) as the compartmental model for prediction in this framework. DELPHI is a compartmental epidemiological model that extends the widely used SEIR model to account for under-detection, societal response, and epidemiological trends including changes in mortality rates. The model is governed by a system of ordinary differential equations (ODEs) across 11 states: susceptible (S), exposed (E), infectious (I), undetected cases who will recover (U^R) or die (U^D), hospitalized cases who will recover (H^R) or die (H^D), quarantined cases who will recover (Q^R) or die (Q^D), recovered (R) and dead (D). The transition rates between the 11 states are defined with 12 parameters, which we predict as $\hat{\theta}$ in the HG-DCM framework. To generate the final incidence curve, the estimated parameters are passed through torchODE(Lienen & Günnemann), a parallel Initial Value Problem (IVP) Solver, to output the predicted cumulative case curve. We used Tsit5 with $a_{tol} = 1 \times 10^{-8}$, $r_{tol} = 1 \times 10^{-4}$ as the ODE solver. We refer the readers to (Li et al.) for details on the DELPHI model and its performance.

Objective Function The objective function of HG-DCM is to minimize the loss between the predicted incidence curve and the actual incidence curve of past and current pandemics. The loss of past pandemics includes both the loss of the length-t training window and the length-v forecasting window (Eqn. 3). The current pandemic loss contains only the training window due to the inaccessibility of the forecasting window in practice (Eqn. 4). Both losses of the past and current pandemics are calculated through a sum of mean absolute error (MAE) and mean absolute percentage error (MAPE) weighted by α to balance the effect of the population. The overall loss is calculated by a mean weighted by β to balance between past pandemic losses and the current pandemic loss (Eqn. 5). The weight determines the amount of information inherited from past pandemics in predicting the current pandemic. Concretely, the formula for the loss function can be written as:

$$L_P = \frac{1}{n_P(t+v)} \sum_{i=0}^{n_P} \sum_{j=0}^{t+v} (|C_{ij} - \hat{C_{ij}}| + \alpha | \frac{C_{ij} - \hat{C_{ij}}}{C_{ij}} |) \tag{3}$$

$$L_C = \frac{1}{n_C t} \sum_{i=0}^{n_C} \sum_{j=0}^{t} (|C_{ij} - \hat{C_{ij}}| + \alpha | \frac{C_{ij} - \hat{C_{ij}}}{C_{ij}} |) \tag{4}$$

$$L = L_P + \beta L_C \tag{5}$$

where n_p/n_c is the number of samples in the past/current pandemic data, and $C_{ij}/\hat{C_{ij}}$ is the actual/predicted cumulative cases of the ith pandemic at the jth time point.

3 EXPERIMENTS

3.1 EXPERIMENTAL SETUP

3.1.1 DATA

We were unable to find a publicly available database that contained pandemic data from the past. Therefore, we constructed a pandemic dataset, which contains case and death (if available) time series data, pandemic meta-data, and country meta-data for major pandemic outbreaks and seasonal pandemics that have occurred worldwide since 1990. Only pandemics with significant (more than 100) and frequent (daily or weekly) reported incidences are included in the dataset. The dataset includes country-level and domain-level data on the following outbreaks: the 2020 COVID-19 pandemic, the 2014 Ebola pandemic, the 2003 SARS pandemic, the Peru (2000 - 2010) and Puerto Rico (1990 - 2008) Dengue Fever outbreak, the 2022 Worldwide Monkeypox outbreak, and world-wide seasonal influenza outbreaks (2009-2023).

The time series dataset contains daily or weekly reported cases for each pandemic. The start date of pandemics differs for each location and is set by the first day when the cumulative case number exceeds 100. Epidemiological meta-data with uncertainties that were available at the early stage of the pandemic for each location are collected. The geological meta-data includes 13 country development indicators from the World Bank data (WorldBank) for each location in the dataset. The list of meta-data is available in A.1

3.1.2 SETUP AND COMPARISON METHODS

Comparison Methods We evaluate the model performances on early-stage forecasting tasks, where HG-DCM is used to forecast the cumulative case curve of 12 weeks based on 2/4/6/8 weeks of daily case data. Due to the lack of death data in pandemics prior to COVID-19, only case numbers are used to fit and evaluate the models in the experiments. Locations with no new daily cases reported during the training window are removed from the dataset. The mean and median MAE of the forecasting window between the predicted incidence and the true incidence are used to evaluate model performance. HG-DCM is compared to state-of-the-art compartmental models DELPHI (Li et al.) and its component deep neural network models, Residual Convolution Neural Network (CNN) (Chung et al.), on the early-stage pandemic forecasting tasks. We attempted to benchmark HG-DCM against other models used for COVID-19 forecasting, specifically those included in the COVID-19 Forecast Hub (Cramer et al., 2022). However, most of these models lack publicly available, reproducible code bases, and the shared forecasting outputs do not include early-stage results (training windows ≤ 8 weeks), thereby limiting direct comparison. Moreover, the majority of models in the Forecast Hub are compartmental models. In contrast, the present study aims to demonstrate that the HG-DCM architecture outperforms both stand-alone compartmental models and its component neural network. We do not aim to compare the performance of different compartmental models, as such performance is highly contingent on pathogen-specific biological characteristics across different pandemics.

HG-DCM Setup Four HG-DCM models are trained using the 2/4/6/8-week training window respectively and predict for 12 weeks. Each HG-DCM is trained separately using the Adam optimizer with a stable learning rate of 1×10^{-5}. Given the large variation of case numbers among different locations, we use a customized geographic-pandemic sampler, where in each batch, one sample from each geographic-pandemic pair in the training data was sampled among all the augmentations. This approach accelerates the convergence by avoiding the turbulent loss curve caused by large variations in incidence numbers among different locations. Dropout or weight decay is not used when training the model. The models are trained for 100 epochs, and the checkpoint with the lowest mean MAE for the training window of the current pandemic for each model is used in the comparison.

Residual-CNN Setup We also train a ResNet-50 model for each training window using the same set of past pandemic data as the HG-DCM to prove that HG-DCM outperforms its component Neural Network. We utilize a learning rate of 1×10^{-5} as optimized through a grid search. No dropout or weight decay is used.

DELPHI Setup For fitting DELPHI models, the cumulative case curves are fitted separately for each location and each training window. Dual annealing (DA) (Xiang et al.) is used as the optimizer for parameter search. The same default parameter ranges as training HG-DCM are used to fit the DELPHI curve.

3.2 RESULTS

Table 1: Model Performance on Covid-19 Early Forecasting. Bold indicates the best-performing models for each training window.

		2 Weeks	4 Weeks	6 weeks	8 Weeks
Mean MAE					
	CNN	15600.4	**11238.1**	11012.5	10211.2
	DELPHI	342686.3	813807.8	29745.6	45140.7
	T-DCM	**15049.2**	17691.2	20571.1	24322.2
	HG-DCM	18602.6	110452.4	**7112.5**	**4643.1**
Median MAE					
	CNN	2963.4	2301.7	**1187.8**	871.8
	DELPHI	3609.1	2619.7	1249.2	**537.7**
	T-DCM	2745.8	2799.1	3101.0	4335.2
	HG-DCM	**2231.1**	**1770.9**	1275.6	796.0

To assess the effectiveness of integrating prior pandemic results into compartmental models, we benchmark HG-DCM against three strong baselines: (i) DELPHI (Li et al.), a state-of-the-art compartmental model that has demonstrated strong performance in early-stage COVID-19 prediction, (ii) CNN, a purely end-to-end deep learning model that forecasts case counts directly from historical pandemic data without using the DELPHI architecture, and (iii) T-DCM, a Deep Compartmental Model (T-DCM) with the same architecture as HG-DCM but excluded the historical pandemic data and meta-data. The aim of the three comparisons is to independently show the value of (i) deep learning, (2) physics-driven modeling, and (ii) past pandemic data.

HG-DCM Outperforms DELPHI Generally, both HG-DCM and DELPHI achieve higher accuracy as the length of training data increases. However, HG-DCM consistently outperforms DELPHI across forecasting horizons, particularly in the crucial early stages when only limited data are available. With 2 weeks of training data, HG-DCM reduces median MAE by 38.2% relative to DELPHI; with 4 weeks, the reduction is 32.4%. When 6 weeks of data are available, HG-DCM and DELPHI achieve comparable accuracy in terms of median error, but HG-DCM forecasts remain more stable across locations (Table 1, Figure 3). HG-DCM addresses a central limitation of compartmental models such as DELPHI, which is the tendency to overshoot case counts when trained on limited data. Overshoot arises from overfitting to the limited training data available, leading to forecasts that substantially deviate from observed trajectories. To formally quantify overshooting, we define it as occurring when the predicted cumulative case count in the final week of the forecasting window exceeds the corresponding observed value by more than fivefold. Across evaluation settings, HG-DCM exhibited markedly fewer overshooting events than DELPHI (Figure 4a). For example, in the case of the United States with an 8-week training window, DELPHI forecasts substantially overshoot true case numbers, whereas HG-DCM, by leveraging historical pandemic information, reduced overfitting and produced predictions that were more consistent with real-world epidemic dynamics (Figure 4b). These results demonstrate the value of incorporating prior pandemic information to enhance early-stage forecasts when outbreak-specific data are scarce.

HG-DCM Outperforms End-to-end CNN Model We next compare HG-DCM to a purely end-to-end CNN model. Unlike HG-DCM, which uses CNN to predict parameters of an epidemiologically grounded model, the CNN baseline bypasses mechanistic structure and directly predicts case trajectories from data. Despite its greater expressiveness, CNN generally underperforms HG-DCM across all training horizons. The performance gap is largest in the early stage (2–4 weeks of training data), where HG-DCM's integration of historical knowledge and compartmental dynamics yields

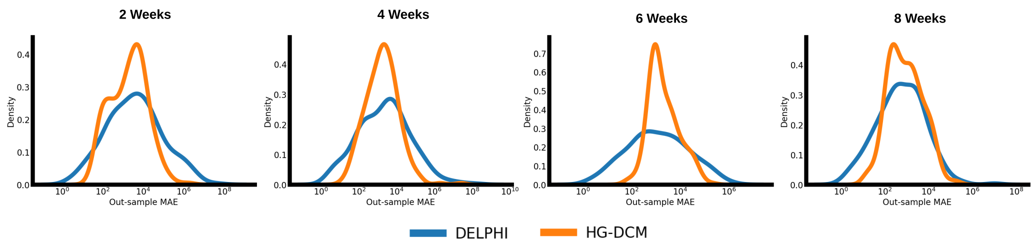

Figure 3: **Forecasting Window MAE Distribution** Forecasting window mean absolute error distribution for DELPHI and HG-DCM on COVID-19 12 Weeks Early Forecasting Tasks using 2 weeks, 4 weeks, 6 weeks, and 8 weeks of available data.

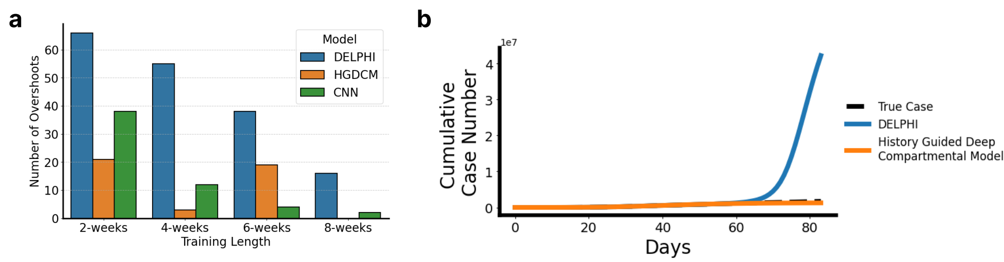

Figure 4: **Forecasting Example (a)** Number of overshooting predictions in different training window length for DELPHI, HGDCM, and CNN. **(b)** United States 8-week-training-window example where DELPHI suffers from overshooting caused by overfitting, while HG-DCM mitigates the overshooting by leveraging historical pandemic data.

markedly lower forecasting error (Table 1). These findings indicate that epidemiological inductive bias provides critical structure for learning, enabling HG-DCM to achieve both stronger predictive performance and greater interpretability than a black-box end-to-end model.

HG-DCM Outperforms T-DCM We further conducted an ablation study by training a Truncated Deep Compartmental Model (T-DCM) that excluded historical pandemic data and meta-data. The T-DCM was trained on datasets with 2, 4, 6, or 8 weeks of observations and evaluated on a 12-week forecasting task.

Table 1 shows that T-DCM consistently underperformed HG-DCM across all training window lengths with respect to median MAE. Notably, HG-DCM achieved significant improvements in median MAE, with the gap widening as training data length increased. This result underscores the importance of incorporating historical context and structured meta-data for reliable forecasting in the early stages of pandemics.

Taken together, these results establish that HG-DCM effectively leverages historical pandemic data to guide compartmental modeling, producing more accurate and stable forecasts than both a leading compartmental model (DELPHI) and a purely data-driven end-to-end model (CNN). By combining mechanistic interpretability with neural network flexibility, HG-DCM represents a significant step forward in reliable early-stage pandemic forecasting.

Parameter Inference One of the key advantages of employing HG-DCM over traditional deep neural networks for pandemic forecasting is its interpretable parameterization. Unlike black-box models, the epidemiologically meaningful parameters predicted by HG-DCM can be extracted before being passed to the Initial Value Problem (IVP) solver, which offers actionable insights.

To illustrate this advantage, we analyzed the parameters inferred by HG-DCM compared to the traditional DELPHI model in an early-stage COVID-19 forecasting task using four weeks of data

(Figure 5). The DELPHI model's parameters exhibited a wide distribution, often leading to unstable forecasts and an overshooting problem. This instability arises because DELPHI fits models independently for each location, amplifying sensitivity to minor noise in the data. In contrast, HG-DCM leverages historical pandemic data and geospatial meta-data, ensuring more robust and consistent parameter estimation.

Statistical analysis using the *Wilcoxon Signed-rank Test* (Woolson) confirmed significant differences in all parameters, including the infection rate (α), median day of action (t_{med}), and rate of action (r_s), with p-values < 0.05. Specifically, HG-DCM predicted a lower infection rate, median day of action, and death rate, while exhibiting a higher rate of action. These findings suggest that, by adapting knowledge from past pandemics, HG-DCM avoids overfitting to the initial boost in case numbers and produces more conservative and realistic estimates, reducing biases that may otherwise arise from noise introduced by heterogeneous factors, including a lack of standardized case identification criteria in the early stage of data collection. The complete parameter analyses for all 12 DELPHI parameters can be found in Appendix A.3.

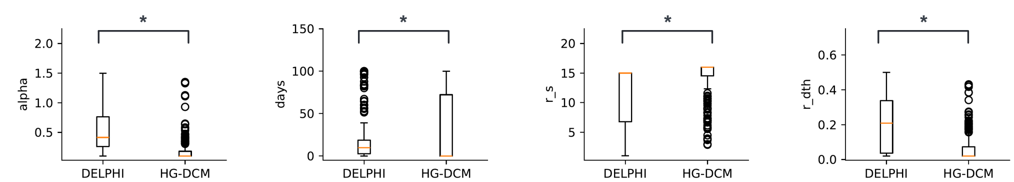

Figure 5: **Comparison of fitted parameters in DELPHI and HG-DCM models.** Box plots show the distribution of selected predicted parameters for DELPHI and HG-DCM. The central line represents the median, the box bounds the interquartile range, and whiskers extend to 1.5× IQR. Outliers are shown as points. Asterisks indicate statistically significant differences between methods (Wilcoxon signed-rank test, p < 0.05).

3.3 Discussion

In this work, we proposed HG-DCM, a hybrid architecture that bridges compartmental models with deep neural networks for early-stage pandemic forecasting. This framework synergizes the interpretability and domain-grounded rigor of compartmental models with the representational power of deep learning. Specifically, HG-DCM leverages the structured epidemiological insights of compartmental models to ensure plausible predictions while harnessing neural networks to integrate auxiliary information from historical, geographical, and meta-pandemic data. This integration effectively mitigates the pitfalls of overfitting and instability, which often plague individual modeling approaches, particularly during the early phases of a pandemic when data is sparse and noisy.

Our results demonstrate that HG-DCM outperforms both standalone compartmental models and purely deep learning-based models on early forecasting tasks. By offering a more robust and accurate early-stage estimation of pandemic trends, HG-DCM addresses critical challenges in public health response, such as resource allocation and policy planning. In particular, its ability to produce stable, noise-robust predictions reduces erratic shifts in trend forecasts, enabling more confident decision-making and minimizing the risks of resource misallocation caused by extreme over- or underestimation.

A key strength of HG-DCM is its interpretability, which remains a cornerstone for pandemic forecasting applications. While deep learning methods often function as opaque black boxes, HG-DCM retains the parameter-driven transparency of traditional compartmental models, with fitted parameters offering actionable epidemiological insights. For instance, in our implementation with the DELPHI compartmental model, the extracted parameters maintain clinical relevance, providing healthcare providers with early and interpretable guidance on the potential trajectory of a pandemic. This interpretability is invaluable for building trust with stakeholders and ensuring actionable insights.

Beyond its strong predictive performance and interpretability, HG-DCM exhibits significant architectural flexibility. In our experiments, we employed a ResNet-based module for temporal representation learning and the DELPHI model for cumulative case curve estimation, selected based on empirical evaluations. However, the modular design of HG-DCM allows for seamless integration of more advanced compartmental models or neural network architectures as they emerge. This adaptability positions HG-DCM as a forward-compatible framework capable of evolving alongside advances in epidemiology and machine learning.

In summary, HG-DCM provides a practical, interpretable, and extensible solution for early-stage pandemic forecasting. By demonstrating the utility of combining epidemiological and deep learning methodologies, our work highlights the potential of hybrid approaches to address complex forecasting challenges in the face of limited data and high uncertainty. Future research may explore augmenting HG-DCM with additional data modalities, enhancing its generalizability to a broader spectrum of infectious diseases, and extending its application to real-time adaptive forecasting.

4 Limitation

While HG-DCM demonstrates strong performance in early-stage pandemic forecasting, it is not without limitations. One notable challenge lies in handling the high variability of incidence rates across different geographical regions. This variability renders conventional normalization techniques, such as batch normalization, unsuitable for the stable estimation of model parameters. Empirical experiments revealed that incorporating batch normalization resulted in unstable predictions, while layer normalization caused critical information loss, impeding model convergence. As a result, no normalization technique was employed in HG-DCM, which, while stabilizing predictions, increased the overall training time due to slower convergence.

Another limitation relates to the availability and quality of historical pandemic data. The COVID-19 pandemic provided the first instance of high-resolution, daily time series data, which proved instrumental in enabling robust model training and evaluation. In contrast, earlier pandemics, such as Ebola, SARS, and Dengue Fever, often lack comparable data granularity. These datasets are frequently reported in weekly aggregates, requiring interpolation to align with HG-DCM's daily prediction framework. Linear interpolation, while a practical workaround, introduces approximation errors, particularly during the critical early stages of a pandemic when precise trend estimation is most needed. This limitation highlights the dependency of HG-DCM on the quality and resolution of input data, which directly impacts its forecasting accuracy.

Furthermore, while the COVID-19 pandemic raised awareness of the importance of robust pandemic data collection, the availability of high-quality, real-time data remains inconsistent across regions and diseases. Recent outbreaks, such as Monkeypox in 2022, demonstrate progress in this area, with improved public access to daily incidence and mortality data. However, disparities in data quality and completeness persist globally, posing ongoing challenges for comprehensive model training.

Despite these constraints, HG-DCM's modular and flexible design ensures its applicability to evolving data landscapes. As the availability and fidelity of historical pandemic datasets improve, future iterations of HG-DCM can leverage these advancements to further enhance its capabilities. Addressing the aforementioned limitations will be critical for developing more generalizable and efficient forecasting frameworks for infectious disease outbreaks.

5 Conclusion

In this work, we introduced HG-DCM, a novel deep compartmental architecture designed to enhance early-stage pandemic forecasting. Our approach integrates historical pandemic data and metadata through a deep learning framework coupled with a compartmental modeling component that generates interpretable forecasts. We demonstrated that HG-DCM outperforms both traditional compartmental models and standalone deep learning models in early-stage forecasting tasks.

These results highlight the promise of deep compartmental models for pandemic forecasting and underscore the value of incorporating historical pandemic data. Future work could focus on integrating additional data sources, such as mobility patterns, policy interventions, or other metadata, to further improve forecasting accuracy. Moreover, adapting HG-DCM for real-time applications represents an exciting avenue for research. We believe this work establishes a foundation for leveraging past pandemics through deep learning to inform future forecasting efforts.

REFERENCES

Martha Barnard, Radhika Iyer, Sara Y. Del Valle, and Ashlynn R. Daughton. Impact of COVID-19 policies and misinformation on social unrest. URL http://arxiv.org/abs/2110.09234.

Grace Chua, Kum Fai Yuen, Xueqin Wang, and Yiik Diew Wong. The determinants of panic buying during COVID-19. 18(6):3247. ISSN 1660-4601. doi: 10.3390/ijerph18063247. URL https://www.mdpi.com/1660-4601/18/6/3247.

Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. URL http://arxiv.org/abs/1412.3555.

Estee Y Cramer, Yuxin Huang, Yijin Wang, Evan L Ray, Matthew Cornell, Johannes Bracher, Andrea Brennen, Alvaro J Castro Rivadeneira, Aaron Gerding, Katie House, Dasuni Jayawardena, Abdul H Kanji, Ayush Khandelwal, Khoa Le, Jarad Niemi, Ariane Stark, Apurv Shah, Nutcha Wattanachit, Martha W Zorn, Nicholas G Reich, and US COVID-19 Forecast Hub Consortium. The united states covid-19 forecast hub dataset. *Scientific Data*, 2022. doi: 10.1101/2021.11.04.21265886. URL https://doi.org/10.1038/s41597-022-01517-w.

Qi Deng and Guifang Wang. A deep learning-enhanced compartmental model and its application in modeling omicron in china. 11(9):906. ISSN 2306-5354. doi: 10.3390/bioengineering11090906. URL https://www.mdpi.com/2306-5354/11/9/906.

Jayanthi Devaraj, Rajvikram Madurai Elavarasan, Rishi Pugazhendhi, G.M. Shafiullah, Sumathi Ganesan, Ajay Kaarthic Jeysree, Irfan Ahmad Khan, and Eklas Hossain. Forecasting of COVID-19 cases using deep learning models: Is it reliable and practically significant? 21:103817. ISSN 22113797. doi: 10.1016/j.rinp.2021.103817. URL https://linkinghub.elsevier.com/retrieve/pii/S2211379721000048.

John M. Eyler. Smallpox in history: the birth, death, and impact of a dread disease. 142(4):216–220. ISSN 00222143. doi: 10.1016/S0022-2143(03)00102-1. URL https://linkinghub.elsevier.com/retrieve/pii/S0022214303001021.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778. IEEE. ISBN 978-1-4673-8851-1. doi: 10.1109/CVPR.2016.90. URL http://ieeexplore.ieee.org/document/7780459/.

Michelle L. Holshue, Chas DeBolt, Scott Lindquist, Kathy H. Lofy, John Wiesman, Hollianne Bruce, Christopher Spitters, Keith Ericson, Sara Wilkerson, Ahmet Tural, George Diaz, Amanda Cohn, LeAnne Fox, Anita Patel, Susan I. Gerber, Lindsay Kim, Suxiang Tong, Xiaoyan Lu, Steve Lindstrom, Mark A. Pallansch, William C. Weldon, Holly M. Biggs, Timothy M. Uyeki, and Satish K. Pillai. First case of 2019 novel coronavirus in the united states. 382(10):929–936. ISSN 0028-4793, 1533-4406. doi: 10.1056/NEJMoa2001191. URL http://www.nejm.org/doi/10.1056/NEJMoa2001191.

Tahir Islam, Abdul Hameed Pitafi, Vikas Arya, Ying Wang, Naeem Akhtar, Shujaat Mubarik, and Liang Xiaobei. Panic buying in the COVID-19 pandemic: A multi-country examination. 59:102357. ISSN 09696989. doi: 10.1016/j.jretconser.2020.102357. URL https://linkinghub.elsevier.com/retrieve/pii/S0969698920313655.

Alexander Janssen, Frank W. G. Leebeek, Marjon H. Cnossen, Ron A. A. Mathôt, and the OPTI-CLOT study group and SYMPHONY consortium. Deep compartment models: A deep learning approach for the reliable prediction of time-series data in pharmacokinetic modeling. 11(7):934–945. ISSN 2163-8306, 2163-8306. doi: 10.1002/psp4.12808. URL https://ascpt.onlinelibrary.wiley.com/doi/10.1002/psp4.12808.

C. KERR. The mathematical theory of infectious diseases and its applications. by norman t. j. bailey, m.a., d.sc.; second edition, 1975. london: Charles griffin & co. ltd. 9*x6*, pp. 430, with diagrams. price: £14.00. 1(18):674–674a. doi: https://doi.org/10.5694/j.1326-5377.1976.tb140951.x. URL https://onlinelibrary.wiley.com/doi/abs/10.5694/j.1326-5377.1976.tb140951.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.5694/j.1326-5377.1976.tb140951.x.

Rob Law, Gang Li, Davis Ka Chio Fong, and Xin Han. Tourism demand forecasting: A deep learning approach. 75:410–423. ISSN 01607383. doi: 10.1016/j.annals.2019.01.014. URL https://linkinghub.elsevier.com/retrieve/pii/S0160738319300143.

Michael Lingzhi Li, Hamza Tazi Bouardi, Omar Skali Lami, Thomas A. Trikalinos, Nikolaos Trichakis, and Dimitris Bertsimas. Forecasting COVID-19 and analyzing the effect of government interventions. 71(1):184–201. ISSN 0030-364X, 1526-5463. doi: 10.1287/opre.2022.2306. URL https://pubsonline.informs.org/doi/10.1287/opre.2022.2306.

Marten Lienen and Stephan Günnemann. torchode: A parallel ODE solver for PyTorch. URL http://arxiv.org/abs/2210.12375.

Colin McEvedy. The bubonic plague.

Xiao Ning, Linlin Jia, Yongyue Wei, Xi-An Li, and Feng Chen. Epi-DNNs: Epidemiological priors informed deep neural networks for modeling COVID-19 dynamics. 158:106693. ISSN 00104825. doi: 10.1016/j.compbiomed.2023.106693. URL https://linkinghub.elsevier.com/retrieve/pii/S0010482523001580.

Daniele Piovani, Maria Nefeli Christodoulou, Andreas Hadjidemetriou, Katerina Pantavou, Paraskevi Zaza, Pantelis G. Bagos, Stefanos Bonovas, and Georgios K. Nikolopoulos. Effect of early application of social distancing interventions on COVID-19 mortality over the first pandemic wave: An analysis of longitudinal data from 37 countries. 82(1):133–142. ISSN 01634453. doi: 10.1016/j.jinf.2020.11.033. URL https://linkinghub.elsevier.com/retrieve/pii/S0163445320307519.

Stephen Reicher and Clifford Stott. On order and disorder during the COVID-19 pandemic. 59(3):694–702. ISSN 0144-6665, 2044-8309. doi: 10.1111/bjso.12398. URL https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bjso.12398.

Alexander Rodríguez, Nikhil Muralidhar, Bijaya Adhikari, Anika Tabassum, Naren Ramakrishnan, and B. Aditya Prakash. Steering a historical disease forecasting model under a pandemic: Case of flu and COVID-19. URL http://arxiv.org/abs/2009.11407.

Ronald Ross. An application of the theory of probabilities to the study of a priori pathometry.—part i. 92(638):204–230. doi: 10.1098/rspa.1916.0007. URL https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1916.0007. _eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.1916.0007.

Kirstin Roster, Colm Connaughton, and Francisco A. Rodrigues. Forecasting new diseases in low-data settings using transfer learning. 161:112306. ISSN 09600779. doi: 10.1016/j.chaos.2022.112306. URL https://linkinghub.elsevier.com/retrieve/pii/S0960077922005161.

Reinhard Schlickeiser and Martin Kröger. Analytical modeling of the temporal evolution of epidemics outbreaks accounting for vaccinations. 3(2):386–426. ISSN 2624-8174. doi: 10.3390/physics3020028. URL https://www.mdpi.com/2624-8174/3/2/28.

Muhammad Usman Tariq and Shuhaida Binti Ismail. Deep learning in public health: Comparative predictive models for COVID-19 case forecasting. 19(3):e0294289. ISSN 1932-6203. doi: 10.1371/journal.pone.0294289. URL https://dx.plos.org/10.1371/journal.pone.0294289.

Lauren C Tindale, Jessica E Stockdale, Michelle Coombe, Emma S Garlock, Wing Yin Venus Lau, Manu Saraswat, Louxin Zhang, Dongxuan Chen, Jacco Wallinga, and Caroline Colijn. Evidence for transmission of COVID-19 prior to symptom onset. 9:e57149. ISSN 2050-084X. doi: 10.7554/eLife.57149. URL https://elifesciences.org/articles/57149.

R. F. Woolson. Wilcoxon signed-rank test. In Ralph B. D'Agostino, Lisa Sullivan, and Joseph Massaro (eds.), *Wiley Encyclopedia of Clinical Trials*, pp. 1–3. Wiley, 1 edition. ISBN 978-0-471-35203-7 978-0-471-46242-2. doi: 10.1002/9780471462422.eoct979. URL https://onlinelibrary.wiley.com/doi/10.1002/9780471462422.eoct979.

WorldBank. World development indicators | DataBank. URL https://databank.worldbank.org/source/world-development-indicators.

Y Xiang, D.Y Sun, W Fan, and X.G Gong. Generalized simulated annealing algorithm and its application to the thomson model. 233(3):216–220. ISSN 03759601. doi: 10.1016/S0375-9601(97)00474-X. URL https://linkinghub.elsevier.com/retrieve/pii/S037596019700474X.