Forecasting Epidemic Spread with Recurrent Graph Gate Fusion Transformers

Minkyong Kim¹, Jae Heon Kim¹, and Beakcheol Jang¹

¹Affiliation not available

June 07, 2024
Abstract—Predicting the unprecedented, nonlinear nature of COVID-19 presents a significant public health challenge. Recent advances in deep learning, such as graph neural networks (GNNs), recurrent neural networks (RNNs), and Transformers, have enhanced predictions by modeling regional interactions, managing autoregressive time series, and identifying long-term dependencies. However, prior works often feature shallow integration of these models, leading to simplistic graph embeddings and inadequate analysis across different graph types. Additionally, excessive reliance on historical COVID-19 data limits the potential of utilizing future data, such as lagged policy information. To address these challenges, we introduce ReGraFT, a novel sequence-to-sequence (Seq2Seq) model designed for robust long-term forecasting of COVID-19. ReGraFT integrates multigraph-gated recurrent units (MG-GRU) with adaptive graphs, leveraging data from individual states, including infection rates, policy changes, and interstate travel. First, ReGraFT employs adaptive MGGRU cells within an RNN framework to capture inter-regional dependencies, dynamically modeling complex transmission dynamics. Second, the model features a self-normalizing priming (SNP) layer using Scaled Exponential Linear Units (SeLU) to enhance stability and accuracy across short, medium, and long-term forecasts. Third, ReGraFT systematically compares and integrates various graph types, such as fully connected layers, pooling mechanisms, and attention-based structures, to provide a nuanced representation of inter-regional relationships. By incorporating lagged COVID-19 policy data, ReGraFT refines forecasts, demonstrating a 2.79% reduction in the root mean square error (RMSE) compared to state-of-the-art models. This work provides accurate long-term predictions, aiding in better public health decisions. Our code is available at https://github.com/mfriendly/ReGraFT.

Index Terms—COVID-19 forecasting, spatiotemporal forecasting, graph-gated recurrent neural networks, long-term prediction

I. INTRODUCTION

COVID-19 is the fifth deadliest pandemic in recorded history, having claimed approximately 7.05 million lives worldwide and 1.19 million in the U.S. alone [1], [2]. Due to the unprecedented, nonlinear, and complex nature of COVID-19, timely and effective policy measures were not initially implemented, while the complexity of the infection’s spread further impeded recovery efforts, requiring a long time to become controllable and predictable. Therefore, prior studies related to COVID-19 took various measures to understand the progression of the pandemic. For example, a change in flight frequency between different areas during the pandemic likely indicates a shift in their spatiotemporal dynamics [3]. Another objective has been to make long-term, rather than short-term, predictions based on available data [3]–[5] to prepare for future pandemics in a timely manner.

With the advent of deep learning techniques in time-series analysis, improved spatiotemporal and long-term forecasts have become feasible. Central to enhancing predictive accuracy are techniques such as GNNs [6]–[8], sequence models including RNNs and Transformers [9]–[14]. For instance, GNNs are effective in modeling regional interactions [7], [15]; RNNs [16], [17] are adept at managing the autoregressive properties of time-series, and Transformer architectures with Multi-Head Attention (MHA) for sequence-to-sequence (Seq2Seq) tasks are highly effective in identifying long-term dependencies [18]–[20]. Given these distinct advantages, research has increasingly focused on hybrid models that combine GNNs with RNNs or Transformers to predict the shift of spatiotemporal dynamics among states, leveraging temporal information from connected regions to enhance forecasting accuracy [8], [9], [21]–[23].

However, these prior works have several shortcomings. First, the integration of GNNs with sequence models, such as RNN and Transformer architectures, is challenging due to the multi-faceted spatiotemporal dynamics and the rapid changes in interactions among states. Second, most prior works have calculated the edge weights among states of inter-regional graphs with a single method without utilizing various approaches such as fully connected (FC) layers, pooling, and attention mechanisms. Third, existing studies predominantly utilize historical COVID-19 infection data and fail to recognize the importance of including lagged or future data for long-term forecasting. This leads to the insufficient use of policy data and future inputs [24]. For instance, Fig. [1]a shows that a decrease in stringency score is highly associated with a subsequent surge in COVID-19 infection rates.

To address the aforementioned gaps and enhance the robustness of forecasts, we propose ReGraFT, a model combining advanced techniques to capture the multifaceted spatiotemporal dynamics and rapid changes in interactions among states. We comprise the architecture with a graph convolutional network (GCN) and a gated recurrent unit (GRU) that incorporates multiple graph gates that reflect regional transmissions at
the gate level of each GRU cell. Each GRU cell includes three graph gates to combine static graphs with a dynamically fused graph created using multi-method adaptive graphs, including fully connected (FC) layer pools and attention mechanisms. This approach allows for a nuanced assessment of inter-regional similarities represented by the nodes and edges, as presented in Fig. 1 (b) and Fig. 1 (c). Our study also addresses the lack of comprehensive comparisons among graph architectures in the literature by methodically analyzing and merging various types to achieve more robust outcomes. Furthermore, based on the hypothesis that using data that reflect the effects over multiple weeks, such as policies, can improve the long-term forecasting accuracy, we incorporate known future variables into our model inputs. Although the impact of policy effectiveness has often been overlooked, [25]–[27] has demonstrated that including these factors can enhance the predictive capabilities of models designed to forecast the spread of COVID-19. This improvement is largely attributable to the significant time lag between the implementation of policies and their effects. Moreover, the comprehensive policy data available for each state provides a valuable resource for spatiotemporal predictions.

ReGraFT represents each state in the U.S. as a node in a graph, utilizing daily infection cases, policy score data, and temporal features (e.g., the day of the week and the week after a holiday) as node characteristics at each time stage. Distances between states and flight frequency data are utilized as two static adjacency matrices, which can be also referred to as edge weights of graphs. Within each multigraph-gated GRU (MGGRU) cell, an adaptive graph is constructed to model daily time steps and capture temporal changes. This adaptive graph, combined with the static graph, is weighted and applied within the RNN framework to facilitate the modeling of temporal dependencies, enabling the identification of complex, nonlinear spatiotemporal dependencies over time. The GCN enables each node to learn the importance of all other nodes, which is then used to update the current node by aggregating features (i.e., infection cases, policies, and time variables) of the neighboring nodes. Additionally, applying future known inputs, including policy delays, to the Seq2seq MHA enhances long-term horizon forecast accuracy. Our approach also incorporates a self-normalizing priming (SNP) layer based on the Scaled Exponential Linear Unit (SeLU) to enhance forecasting accuracy and stability over short (1 to 14 days), medium (14 to 28 days), and long (28 to 42 days) periods. In summary, the primary contributions of ReGraFT are as follows:

- **ReGraFT framework:** We introduce ReGraFT, a Seq2Seq model enhanced with the integration of RNN, MHA, and GNN for spatiotemporal modeling. The model includes adaptive MGGRU cells, which capture and utilize regional interdependencies within the data accurately, reflecting dynamic connections and interactions over time through adaptive similarity graphs.

- **Systematic comparison and fusion of graphs:** We present a comprehensive comparison of various graph types—namely, FC layer, pooling-, and attention-based structures—and implement their fusion to effectively represent inter-regional relationships and achieve robust outcomes.

- **SNP layer:** We develop a novel neural network layer that integrates self-normalizing properties using the SeLU activation function. The SNP layer is added to enhance the model stability during training and overall performance.

- **Integration of known future inputs for long-term forecast accuracy:** We incorporate lagged features, particularly those pertaining to the implementation of COVID-19-related policy data, into the model to enhance the understanding and forecasting ability of ReGraFT, thereby improving the accuracy of long-term forecasts.

- **Evaluation of model performance:** We conducted several tests on COVID-19 Tracking Project dataset and Ox-
Fig. 2. Our proposed model, ReGraFT. (a) Multigraph-gated recurrent units (MGGRU) capture and exploit regional interrelationships that reflect the dynamic connectivity and interactions among U.S. regions over time. (b) Inter-Regional Transmission (IRT) layer and Multi-Graph module dynamically model complex transmission dynamics through weighted fusion of static and adaptive graphs. (c) Self-normalizing priming (SNP) layer uses Scaled Exponential Linear Units (SeLU) to improve stability and accuracy over short, medium, and long-term predictions. (d) ReGraFT uses past inputs as keys and future known inputs as queries, with values representing the target variable. The model calculates the future target by evaluating the similarity between past and future inputs, and then aggregates these features into the final forecast using a fully connected layer.

CGRT COVID-19 policy dataset, in order to demonstrate that ReGraFT reduces the root mean square error (RMSE) by 2.79% compared to prior state-of-the-art models.

The remainder of this paper is organized as follows: Section II reviews related studies. Section III describes the materials and methods used in this study. Experimental setup and results are discussed in Section IV-F and Section V. Sections VI and VII present the discussion and conclusions, respectively.

II. RELATED WORK

A. Transformers in Time-Series Forecasting

Recent advancements in deep learning have markedly improved long-term time-series forecasting capabilities. Initially developed for natural language processing, Transformers leverage self-attention mechanisms and have been successfully adapted for various tasks, including video, audio, and image processing [28]. These models have demonstrated promising results in time-series forecasting applications [18], [29]. Seq2seq models, which are renowned for their ability to capture long-term dependencies, generate one sequence from another, thereby maintaining context over extended data sequences. This capability is essential for accurately predicting future sequences and managing complex patterns over time. An exemplary model, the Temporal Fusion Transformer (TFT), is effective at handling multivariate inputs and incorporates known future data for accurately forecasting pandemics [20]. However, conventional Transformers encounter difficulties with spatiotemporal dynamics and regional connectivity, which are critical for precise forecasting. These limitations underscore the necessity for further innovations in Transformer architecture to enhance their efficacy in complex forecasting scenarios.

B. Forecasting of Pandemics with Advanced Graph Neural Networks

Traditional and machine-learning models often encounter difficulties in accurately representing the dynamic nature of
pandemics and the complex regional connectivity [30]–[32]. Deep-learning models such as RNNs, long short-term memory (LSTM) [17], and gated recurrent units (GRUs) [16] have been employed to address this challenge, yet have been found to be ineffective in capturing spatiotemporal relationships. GNNs such as STGNN [22] and FIGI-Net [23], combine LSTM with a multi-layered GNN to model the transmission dynamics of COVID-19. STSGT [8] and CausalGNN [33] employ GCNs and attention mechanisms to enhance accuracy by capturing spatiotemporal dependencies. However, the majority of models utilize static graph structures with shallow embeddings, which fails to account for dynamic regional connections. To address this, we use an improved multigraph-gated recurrent neural network inspired by [34]–[36] with adaptive graphs to model evolving relations over time across diverse aspects. Consequently, we address the deficiency in long-term forecasting and the failure to utilize policy data or known future inputs. This enables us to overcome the limitations of prior models’ explanatory power and utility in public health planning.

**C. Leveraging Multivariate Features for COVID-19 Prediction**

Prior research has investigated various factors that influence the spread of COVID-19. Studies such as that of Wibbens et al. [26] have demonstrated the seasonality of the pandemic, linking it to other respiratory viruses and emphasizing the importance of epidemiological data such as hospitalization rates and mobility. In addition, environmental and demographic factors, including data from search engine queries [21], have proven to be valuable in refining the prediction accuracy.

In 2021, the Oxford COVID-19 Government Response Tracker (OxCGRT) released comprehensive indices of global policy responses, aiding in the evaluation of pandemic policies worldwide [4]. However, the impact of these policies on the trajectory of the pandemic has not been explored in depth. Wibbens et al. [27] shed light on the significant role of various policies and provided essential insights for understanding effective pandemic responses.

This study builds upon the insights of Wibbens et al. [27] by analyzing how specific policies can significantly alter the outcomes of the COVID-19 pandemic. This focus is vital for developing more accurate pandemic response strategies. While previous studies have largely overlooked the integration of local policy data into GNNs for pandemic modeling, which restricts our understanding of how regional policy discrepancies influence the spread and management of COVID-19, we incorporate diverse data sources across multiple domains, thereby enhancing the comprehensiveness of COVID-19 forecasting.

**III. PROPOSED MODEL**

Fig. 2 illustrates the proposed framework. Within this framework, we introduce a COVID-19 forecasting model that combines Multi-head Attention [28] and MGGRUs. The Seq2Seq-based model with MHA handles known future time step inputs, making it suitable for long-term predictions. The following subsections narrate the model’s details.

**A. Self-Normalizing Priming (SNP) Layer**

The SNP layer utilizes SeLU [37] activation functions for self-normalization. These functions play a pivotal role in the performance of deep learning networks, with the Rectified Linear Unit (ReLU) being one of the most prevalent functions due to its effectiveness in enhancing neural network operations [38]. The selection of the activation function can have a significant impact on the network’s performance and stability [39]. Despite its prevalent use, ReLU is vulnerable to the “dying ReLU” issue, where many neurons become non-functional and stop learning because of vanishing gradients. To address this problem, ELU [40] and SeLU have been proposed. SeLU is noted for its self-normalizing properties, ensuring each network layer has a mean of zero and a variance of one, which improves convergence and reduces the need for external normalization. The SeLU function is defined as SeLU\((x) = \xi x\) if \(x > 0\), and \(\gamma e^x - \gamma\) if \(x \leq 0\), where \(\xi\) and \(\gamma\) are parameters that can enhance the function’s self-normalizing capability. This self-normalization feature of SeLU helps to steer the activations towards a normal distribution, thereby stabilizing the gradients [37].

After this module, layer normalization [41] is used to control the SeLU output by calculating the statistics across all inputs in a layer instead of across a batch. This ensures that the activation scales are constant across various inputs in the same layer. This output is combined with FC layers to create the constituent block instead of the gated residual networks used in the TFT [37]. In addition, we replace all activations in the architecture with SeLU, except for the adaptive graphs and gated linear unit (GLU) in the final layers, where ReLU and Sigmoid are used.

**B. Variable Priming Network**

We employ a variable priming network, as in TFT, but use our SNP layer instead of a gated residual network, which uses ELU. Compared to ELU and ReLU, SeLU offers the advantage of self-normalization, which ensures consistent variance across network layers and prevents vanishing or exploding gradient problems, thereby promoting more stable and reliable training in deep learning architectures. These features make SeLU particularly effective for models that require robust feature selection and long-term dependency modeling.

**C. Multigraph-gated Recurrent Units**

We construct multigraph-gated recurrent units which are based on the GCN-gated RNN [34] in Seq2seq models. We effectively capture and utilize regional inter-relationships, reflecting the dynamic connectivity and interactions among states over time.

1) **Gating Operations within a Recurrent Cell:** Fig. 2 (a) shows the architecture of the MGGRU cells. In the MGGRU, GCN layers are added in front of each gate of the traditional MGGRU, which allows the model to capture both the spatial and temporal dependencies that exist in the sequential data. In the proposed model, the encoder and decoder RNN consist of multiple MGGRU cells instead of GRU or LSTM. In the Seq2seq for time series forecasting, the encoder is first fed
a set of input signals $X_1, \ldots, X_t$ and repeatedly updates the hidden state $h_t$. The final encoder state $h_{t-1}$ is then passed to the decoder. This decoder predicts the future signal sequence of $X_{t+1}, \ldots, X_{t+q}$ through recursive prediction. In the decoding phase, the final cell of the decoder produces a signal prediction. The decoder uses the prediction as input to the next prediction step for testing and validation. Within the cell, the input features at time $t$, $X_t$, and previous hidden state $h_{t-1}$ are concatenated and fed to both the reset gate $r_t$ and update gate $u_t$. The amount that $h_{t-1}$ and $X_t$ contribute to $c_t$ is determined by the reset gate, $r_t$. Subsequently, as $c_t$ updates the hidden state $h_t$, the update gate controls the flow of new data into it. The operation of an MGGRU cell can be described as follows:

- **Update gate:**
  $$z_t = \sigma(I\!R(T(w_z X_t + u_z h_{t-1}) + b_z))$$

- **Reset gate:**
  $$r_t = \sigma(I\!R(T(w_r X_t + u_r h_{t-1}) + b_r))$$

- **Candidate hidden state:**
  $$c_t = \tanh(I\!R(T(w_h X_t + u_h (r_t \odot h_{t-1})) + b_h))$$

- **Final hidden state:**
  $$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot c_t,$$

Here, the variables $r_t$, $u_t$, $c_t$, and $h_t$ represent the reset gate, update gate, cell states, and final hidden state, respectively. The $X_t$ represents the input of the MGGRU at time $t$. The bias terms are denoted by $b_r$, $b_u$, and $b_c$.

2) **Inter-Regional Transmission Layer:** In Equation 2 where each GCN gate uses IRT, Inter-Regional Transmission Layer (IRT layer), or the weighted fusion of the static and adaptive graphs, is defined as follows:

$$IRT(x) = SNP \left[ w_0 \odot x + w_1 \odot GCN_{distance}(x) + w_2 \odot GCN_{flight}(x) + w_3 \odot GCN_{Pool}(x) + w_4 \odot GCN_{Attention}(x) \right]$$

Using this multigraph approach, the model in each gate enables each vertex in the network to determine the importance of every other vertex. It then modifies the properties of the hub node by combining the features (infection rates, policy measures, and temporal features) from its neighboring nodes.

3) **Static Graph Calculation:** As described in [H-C.T] each GCN layer utilizes two static adjacency matrices and adaptive matrices for its computations. Based on the method used in [36], the static matrices, $G_{distance}$ and $G_{flight}$, are based on geographic distance and flight frequency, respectively. $G_{distance}$ is calculated from historical flight data, where each element $G_{distance_{\alpha,\beta}}$ is defined as $\exp \left( -\frac{distance_{\alpha,\beta}^2}{\sigma^2} \right)$ if $distance_{\alpha,\beta} > \omega_{distance}$ and 0 otherwise. Here, $\sigma$ is the standard deviation controlling the decay of interaction with distance, and $distance_{\alpha,\beta}$ is the geographic distance between locations $\alpha$ and $\beta$. Similarly, $G_{flight}$ is calculated as $\frac{flight_{\alpha,\beta}}{\sum_{\beta} \text{flight}_{\alpha,\beta}}$, which represents normalized flight frequencies between locations, where $flight_{\alpha,\beta}$ represents the flight frequency from location $\alpha$ to location $\beta$ and $E$ is the number of edges.

4) **Adaptive Graph Construction:** As shown in Equation 3 and Fig. 1, the adaptive similarity graphs compute the node similarities in the hidden states to dynamically generate graphs within the hidden state in the GRU cells. This module fuses the input features $X_t$ and the hidden states $h_t$ into a single vector, which then goes through an FC layer. Our experiments included three types of graphs: those based on pooling, FC layers, and attention mechanisms, with configurations where pooling layers were replaced by FC layers and multi-head attention, as well as combinations of these. For the pooling variant, the vector is split for processing by average and maximum pooling, then combined and normalized. The FC variant uses FC layers within a GRU cell to dynamically capture and quantify the evolving relationships between the U.S. states. It combines the input and hidden states of regions, which then pass through FC layers, applying a ReLU activation to introduce nonlinearity and a dropout layer to prevent overfitting. The output and its transpose undergo matrix multiplication to form a similarity matrix, which is then adjusted by a scaling factor and activated by a hyperbolic tangent function modulated by a learnability parameter. This similarity matrix is normalized to produce an adjacency matrix, representing the strength of inter-regional dependencies, which is important for understanding spatiotemporal dynamics in predictive models. The MHA variant replaces the FC layer with MHA. SNP units, self-attention units with SeLU at the end, are added at the beginning and end of each MGGRU cell to focus on important information in the time step. Experiments have shown that this approach is effective in focusing on important information.

D. **Multi-head Attention (MHA) for Final Prediction**

We employ MHA by using the decoder output as the key, and the encoder output as the query. Seq2seq with MHA employs future-known inputs in a special architecture that is designed for long-term prediction, whereas MGGRU is tailored for more abrupt dynamic changes in short-term forecasting. In this setup, the keys represent past inputs, the queries represent future known inputs, and the values represent the target variable. This setup projects the future target using the similarity between past and future known inputs. An FC layer aggregates these features to produce the final forecast. Several measures such as norm, GLU and skip connections were implemented.
Fig. 3. Table of all variables used for study

### IV. Experimental Setup

<table>
<thead>
<tr>
<th>Table I</th>
<th>Data Segmentation for Training, Validation, and Testing Datasets.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training</td>
</tr>
<tr>
<td>Start Date</td>
<td>2020-03-01</td>
</tr>
<tr>
<td>End Date</td>
<td>2021-07-24</td>
</tr>
<tr>
<td>Days</td>
<td>511</td>
</tr>
<tr>
<td>Number of Samples</td>
<td>51 × 511</td>
</tr>
<tr>
<td>Mean</td>
<td>948.39</td>
</tr>
<tr>
<td>Median</td>
<td>485.57</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>1394.02</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.00</td>
</tr>
<tr>
<td>Maximum</td>
<td>9806.71</td>
</tr>
</tbody>
</table>

**A. Datasets**

1) **COVID-19 Dataset**: We use data of newly confirmed COVID-19 case reports from the COVID Tracking Project [5]. The dataset includes all 50 U.S. states and Washington, D.C.

2) **COVID-19 Policy Dataset**: The OxCGRT dataset [24] offers extensive information on COVID-19 policy responses globally, encompassing three years of policy data across more than 180 countries and 200 regional jurisdictions. This dataset features 23 policy indicators, including measures like school closures, travel bans, and vaccination strategies, as depicted in Fig. 3. See Section IV-A.2 for variable selection methodology.

Table I details the data distribution: 70% for training, 15% each for validation and testing. Training covers 511 days from March 1, 2020, to July 24, 2021, with 51 daily samples. Testing spans 110 days from November 12, 2021, to March 1, 2022, during the Omicron outbreak [23]. The test data showed increased mean values and standard deviations, indicating the need for careful scaling strategies.

**TABLE II**

<table>
<thead>
<tr>
<th>Model</th>
<th>1-7</th>
<th>7-14</th>
<th>14-21</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>2021.71</td>
<td>76.91</td>
<td>2022.77</td>
</tr>
<tr>
<td>ColaGNN</td>
<td>2295.38</td>
<td>79.75</td>
<td>2388.05</td>
</tr>
<tr>
<td>DCRNN</td>
<td>2391.74</td>
<td>79.76</td>
<td>2388.05</td>
</tr>
<tr>
<td>TFT (wo. SE)</td>
<td>2290.65</td>
<td>79.59</td>
<td>2388.05</td>
</tr>
<tr>
<td>TFT (w. SE)</td>
<td>2287.75</td>
<td>79.78</td>
<td>2388.05</td>
</tr>
<tr>
<td>ReGraFT (wo. SA)</td>
<td>2298.22</td>
<td>79.65</td>
<td>2388.05</td>
</tr>
<tr>
<td>ReGraFT (Attn)</td>
<td>2292.51</td>
<td>79.76</td>
<td>2388.05</td>
</tr>
<tr>
<td>ReGraFT (Pool)</td>
<td>2284.38</td>
<td>79.91</td>
<td>2388.05</td>
</tr>
<tr>
<td>ReGraFT (FC)</td>
<td>2281.64</td>
<td>79.76</td>
<td>2388.05</td>
</tr>
</tbody>
</table>

**TABLE III**

<table>
<thead>
<tr>
<th>Model</th>
<th>21-28</th>
<th>28-35</th>
<th>35-42</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>2021.71</td>
<td>76.91</td>
<td>2022.77</td>
</tr>
<tr>
<td>ColaGNN</td>
<td>2295.38</td>
<td>79.75</td>
<td>2388.05</td>
</tr>
<tr>
<td>DCRNN</td>
<td>2391.74</td>
<td>79.76</td>
<td>2388.05</td>
</tr>
<tr>
<td>TFT (wo. SE)</td>
<td>2290.65</td>
<td>79.59</td>
<td>2388.05</td>
</tr>
<tr>
<td>TFT (w. SE)</td>
<td>2287.75</td>
<td>79.78</td>
<td>2388.05</td>
</tr>
<tr>
<td>ReGraFT (wo. SA)</td>
<td>2298.22</td>
<td>79.65</td>
<td>2388.05</td>
</tr>
<tr>
<td>ReGraFT (Attn)</td>
<td>2292.51</td>
<td>79.76</td>
<td>2388.05</td>
</tr>
<tr>
<td>ReGraFT (Pool)</td>
<td>2284.38</td>
<td>79.91</td>
<td>2388.05</td>
</tr>
<tr>
<td>ReGraFT (FC)</td>
<td>2281.64</td>
<td>79.76</td>
<td>2388.05</td>
</tr>
</tbody>
</table>

**1) Preprocessing**: To address irregular COVID-19 case reporting, we applied moving average smoothing and differencing to standardize the infection data, new cases, and policy data. Standard scaling was used to normalize the data and transform the values into standardized deviations from the means. This process included subtracting the means. This process included subtracting the means of variables, including demographic, economic, epidemiological, geographic, and health-related variables, as well as hospitalization rates and mobility indices, we focused on policy-relevant variables such as lockdowns, travel restrictions, and vaccination drives owing to their significant impact on...
pandemic transmission. The inputs were categorized into three groups: holiday variables, time categorical variables, and policy variables. The variable importance analysis highlighted 42-day lag variables for public event cancellations, contact tracing, international travel controls, and gathering restrictions as key predictors. Fig. [3] lists the variables used in the prediction. Final variables were a total of 14, including “cancel public events lag 42,” “categorical day of week,” “categorical month,” “categorical week,” “contact tracing lag 42,” “days from start,” “international travel controls lag 42,” “is holiday,” “is holiday lag 1,” “is holiday lag 2 is holiday lag 3,” “new confirmed,” “restrictions on gatherings lag 42,” “week from the start.”

B. Performance Measures

Root mean square error (RMSE) and Pearson correlation coefficient (PCC) are employed to evaluate the accuracy of COVID-19 infection forecasting. The RMSE is calculated as

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (y_n - \hat{y}_n)^2},$$

quantifying the average magnitude of the forecasting error. The PCC, measuring the linear correlation between the observed and forecasted values, is defined by

$$\text{PCC} = \frac{\sum_{n=1}^{N} (y_n - \bar{y})(\hat{y}_n - \bar{\hat{y}})}{\sqrt{\sum_{n=1}^{N} (y_n - \bar{y})^2} \sqrt{\sum_{n=1}^{N} (\hat{y}_n - \bar{\hat{y}})^2}}.$$

Here, $y_n$ and $\hat{y}_n$ represent the individual observed and forecasted sample points, respectively, while $\bar{y}$ and $\bar{\hat{y}}$ are the means of the observed and forecasted values.

C. Baseline Methods

Numerous spatiotemporal frameworks, such as ColaGNN, have been initially configured to process univariate input for each region. When subjected to multivariate data, these models exhibited suboptimal performance. Thus, it has been concluded that they are unsuitable for comparative analysis in this research and have been omitted from the performance evaluation tables.

The set of baseline models along with their descriptions are provided below, with the corresponding hyperparameters detailed in Table III.

- **GRU** [16]: A Seq2seq model using GRUs to predict future values based on past sequences. The parameters are not shared among different regions.
- **DCRNN** [34]: A diffusion convolution RNN that combines GCNs with RNNs in Seq2seq.
- **ColaGNN** [7]: A GNN and Transformer for modeling the spatiotemporal spread of COVID-19.
- **TFT with static embedding** [20]: As suggested in the original study, a static covariate encoder was used. Although it would be appropriate to use a GRU version for a fair comparison, we conducted experiments using LSTM, as in the original study. We used 14 variables, which significantly increased the number of parameters because the embedding dimension for each variable is 4.
- **TFT without static embedding** [20]: TFT without static embedding.

D. ReGraFT Model Variations

- **ReGraFT (wo SA)**: ReGraFT without self-attention-based SNP Unit in the recurrent cell.
- **ReGraFT (Pool)**: ReGraFT with pooling operations.
- **ReGraFT (Fc)**: ReGraFT with FC layers.
- **ReGraFT (Attn)**: ReGraFT with attention-based operations.
- **ReGraFT (FAP)**: ReGraFT with FC, attention, and pooling fusion.

E. Training Procedure and Hyperparameters

For the U.S. regional data, the first 70% of the timeline was designated as the training set, followed by 15% for validation and the remaining 15% for testing. The data were split prior to pre-processing steps, such as normalization, standardization, missing value imputation, and feature selection, to avoid data leakage. The hyperparameters were determined using the validation dataset. Table III lists the hyperparameters used for the ReGraFT and baseline models.

F. Environmental Setup

The proposed model was developed in Python, version 3.9. Evaluations were conducted on a high-performance computer equipped with an NVIDIA GeForce RTX 3080 GPU, an Intel(R) Xeon(R) E5-2686 v4 @ 2.30 GHz CPU, and 128GB of RAM. Both the proposed model and the deep-learning baseline models were implemented using PyTorch [42].

V. EXPERIMENTAL RESULTS

The simple average was used to calculate the national aggregate metrics, as we used normalization and the states were equally weighted. We aggregated the predicted and actual values for each week for errors across different horizons. For example, a one-week horizon included errors from the first seven days and a two-week horizon included errors from days 7 to 14.

A. Performance Analysis

The average results are presented in Table III and Fig. 5 with windowed prediction plots in Fig. 5. Variations in PCC and RMSE were observed across forecast horizons ranging from 1 to 42 days. Section III-C details experiments with different adaptive graphs employing attention, convolution, and pooling, as well as their combinations. Experiments were conducted with several random seeds, with results averaged from three median values.

The most ReGraFT variants generally exhibited excellent RMSE and the PCC in all intervals, meaning that the adaptive mechanism that captures inter-regional correlation significantly improved prediction accuracy.

Nevertheless, as shown in Fig. 4, two variants stood out in particular: one using solely the FC layer-based graph, and another that combined FC with attention and pooling for graph construction.

To illustrate, both in short- and mid-term forecasting (1–28 days), ReGraFT (FC), which utilizes an adaptive graph
Fig. 4. Performance comparison between ReGraFT and baseline models. While ReGraFT outperforms all baseline models in all horizons, the relatively small absolute differences in model performance can be attributed to the use of small batch sizes, differencing, and standard scaling, coupled with a small learning rate, which was applied to train all models.

<table>
<thead>
<tr>
<th>Model</th>
<th>RMSE 1-14d</th>
<th>RMSE 14-28d</th>
<th>RMSE 28-42d</th>
<th>PCC 1-14d</th>
<th>PCC 14-28d</th>
<th>PCC 28-42d</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRU</td>
<td>85.62</td>
<td>85.92</td>
<td>86.29</td>
<td>90.12</td>
<td>86.28</td>
<td>90.32</td>
</tr>
<tr>
<td>ColaGNN</td>
<td>85.62</td>
<td>85.92</td>
<td>86.29</td>
<td>90.12</td>
<td>86.28</td>
<td>90.32</td>
</tr>
<tr>
<td>TFT (w/o. SE)</td>
<td>85.62</td>
<td>85.92</td>
<td>86.29</td>
<td>90.12</td>
<td>86.28</td>
<td>90.32</td>
</tr>
<tr>
<td>DCRNN</td>
<td>85.62</td>
<td>85.92</td>
<td>86.29</td>
<td>90.12</td>
<td>86.28</td>
<td>90.32</td>
</tr>
<tr>
<td>ReGraFT (Attn)</td>
<td>85.62</td>
<td>85.92</td>
<td>86.29</td>
<td>90.12</td>
<td>86.28</td>
<td>90.32</td>
</tr>
<tr>
<td>ReGraFT (FAP)</td>
<td>85.62</td>
<td>85.92</td>
<td>86.29</td>
<td>90.12</td>
<td>86.28</td>
<td>90.32</td>
</tr>
<tr>
<td>ReGraFT (wo. SA)</td>
<td>85.62</td>
<td>85.92</td>
<td>86.29</td>
<td>90.12</td>
<td>86.28</td>
<td>90.32</td>
</tr>
</tbody>
</table>

Based on FC layers, showed the best RMSE and PCC metric. Meanwhile, ReGraFT (FAP) demonstrated its long-term forecasting ability by achieving the best long-term RMSE of 1258.44 and, at the same time, the best PCC of 86.10 over 35 to 42 days. This indicates that the fusion of multiple graphs reflects the interregional relationship of multiple aspects well by synthesizing various data-based insights. Existing models such as GRU and DCRNN generally do not reach the level of performance achieved by the ReGraFT variant.

Ablative testing of the SNP units at both the start and end of each recurrent cell revealed that ReGraFT (wo. SA) experienced a minor performance drop. This demonstrated the attention mechanism’s effectiveness within the RNN cell for capturing dynamic correlations.

Fig. 5. Windowed prediction plots (35- to 42-day horizon) for 8 states showing the two best model predictions. While both ReGraFT and the TFT model exhibited a close approximation to the actual values, ReGraFT showed a quicker adjustment to dynamics in curvature.
Fig. 6. Significance analysis of variables. Encoder and decoder weights were averaged; units are in %.

Fig. 7. Correlation heatmap of significant variables.

Overall, all ReGraFT variants except for ReGraFT (Attn) consistently yielded lower RMSE across both short- and long-term forecasts than baselines, demonstrating the FC layers’ effectiveness in capturing interregional dynamics for various forecasting horizons.

While all ReGraFT variants except for ReGraFT (Attn) consistently achieved lower RMSE across both short- and long-term forecasts than baseline models, highlighting the FC layers’ capability in capturing interregional dynamics, the FAP variant, while exhibiting the best RMSE and PCC for the longest forecast horizon, the variability in the corresponding PCC for shorter horizons suggested that focusing only on the RMSE during training—using it as the sole objective function—may not have adequately optimized the models to achieve a holistic correlation performance.

In addition, the suboptimal performance observed in the ReGraFT (Attn) using only the attention might stem from using just two heads in the attention module. Given the hidden dimension (number of features) of 14, which allows only one or two heads—since 14 is not divisible by 4 or 8—using the commonly employed four heads was not feasible.

In future work, we could benefit from adjusted training objectives that give equal weight to RMSE and the PCC, potentially stabilizing prediction consistency across different metrics and improving the model’s ability to generalize across different data sets.

B. Variable Importance Analysis

As illustrated in Fig. 6, the attention weight analysis revealed that “restrictions on gatherings” was the most effective variable for forecasting. “Contact tracing,” “cancel public events,” and “international travel controls” also showed significant scores, reflecting their effectiveness in limiting public interactions and the cross-regional spread of diseases. Based on these results, the four variables with the highest scores were included in the final model: “restrictions on gatherings,” “contact tracing,” “cancel public events,” and “international travel control.”

Fig. 7 shows the correlation heatmap for the significant variables. “New confirmed cases” and “cancel public events lag 42” exhibited a noticeable negative correlation. “Contact tracing lag 42” and “international travel controls lag 42” were strongly positively correlated, indicating that rigorous contact tracing often coincides with strict travel controls. The strongest correlation was observed between “cancel public events lag 42” and “restrictions on gatherings lag 42,” suggesting these interventions are commonly implemented together. These observations align with Fig. 1.

VI. DISCUSSION

A. Research Summary

The ReGraFT framework models each U.S. state as a node within a graph, incorporating daily infection numbers, policy scores, and temporal attributes such as weekdays and the periods following holidays. This forecasting framework enhanced the modeling of spatiotemporal dependencies between regions by integrating adjacency matrices, constructed through fusions of the adaptive graph using various methods such as pooling, FC layers, and attention mechanisms, to capture the multifaceted nature of relationships between U.S. states. Our findings indicate that the ReGraFT model significantly improves forecasting accuracy, achieving a 2.79% reduction in RMSE compared to the current state-of-the-art models.

B. Contribution to the Field

Barros et al. [43] highlighted the lack of a universal model for graph embeddings that could be applied across a range of dynamic graphs and emphasized the importance of exploring embeddings for graphs with multilayer structures at various temporal and connectivity levels. Addressing these
gaps, our study enhances the capabilities of GNNs, RNNs, and Transformer models in epidemic forecasting through the development of our ReGraFT framework, which systematically fuses various graph types—such as fully connected layers, pooling mechanisms, and attention-based structures—to provide a nuanced representation of inter-regional relationships, thereby addressing the issues of simplistic graph embeddings and shallow integration seen in most contemporary works [8], [22], [23]. This methodology allows for a more precise representation of fluctuating conditions and interactions over time, which is crucial for the dynamic nature of pandemics.

Additionally, our approach advances time-series forecasting by introducing a novel SNP layer using SeLU instead of ReLU or ELU, significantly improving upon traditional time-series transformers like those described by Lim et al. [20]. This innovation not only bridges the gaps highlighted by previous studies but also extends the functionalities of time-series models in handling complex epidemic data.

Thus, the flexibility and robustness of our model make it a potentially universal tool for modeling various infectious diseases, significantly broadening its applications in public health and epidemiological research.

C. Limitations and Future Work

While the ReGraFT model represents a significant step forward in forecasting capabilities, it has some limitations that can be addressed in future research, as described in Section V. One such limitation involves revising the training objectives to balance the weights between RMSE and PCC. Adjusting these weights could enhance prediction consistency across different metrics and improve the model’s applicability to various datasets.

Additionally, due to constraints on resources and time and considering the utility of our findings, we primarily focused our analysis on policy-related factors, which are vital for effective pandemic management. However, there is potential to broaden the scope of our research to include variables such as environmental factors and demographic information. Exploring these factors as well as their interactions could yield more comprehensive insights into the external factors’ impact on epidemic dynamics.

Addressing these limitations could further enhance the robustness and utility of the forecasting model for future epidemic scenarios.

VII. Conclusion

Accurate and interpretable epidemic modeling is crucial for informed policy decision-making and resource allocation. In particular, unpredictable cases of COVID-19 spreading across multiple regions pose serious forecasting challenges owing to rapid spatiotemporal transmission dynamics.

The proposed ReGraFT framework represents each state in the U.S. as a node on a graph, using daily infection cases, policy scores, and temporal features (such as weekdays and the week following holidays). It models daily time steps and constructs an adaptive graph for each MGGRU cell to capture temporal changes. The RNN framework facilitates temporal dependency modeling by applying weights to the adaptive graph in conjunction with a static graph. The adaptive graph was tested using various methods, including pooling, FC layers, and attention mechanisms, thereby providing a novel contribution. This approach significantly improves forecasting precision and provides nuanced insights for the public, potentially making substantial contributions to health policies over time.

REFERENCES


