

# Inter- and Intra-Series Embeddings Fusion Network for Epidemiological Forecasting

Feng Xie  
National University of Defense  
Technology  
Changsha, China  
xiepeng@nudt.edu.cn

Zhong Zhang  
National University of Defense  
Technology  
Changsha, China  
zhangzhong@nudt.edu.cn

Xuechen Zhao  
National University of Defense  
Technology  
Changsha, China  
zhaoxuechen@nudt.edu.cn

Bin Zhou\*  
National University of Defense  
Technology  
Changsha, China  
binzhou@nudt.edu.cn

Yusong Tan  
National University of Defense  
Technology  
Changsha, China  
ystan@nudt.edu.cn

## ABSTRACT

The accurate forecasting of infectious epidemic diseases is the key to effective control of the epidemic situation in a region. Most existing methods ignore potential dynamic dependencies between regions or the importance of temporal dependencies and inter-dependencies between regions for prediction. In this paper, we propose an Inter- and Intra-Series Embeddings Fusion Network (SEFNet) to improve epidemic prediction performance. SEFNet consists of two parallel modules, named Inter-Series Embedding Module and Intra-Series Embedding Module. In Inter-Series Embedding Module, a multi-scale unified convolution component called Region-Aware Convolution is proposed, which cooperates with self-attention to capture dynamic dependencies between time series obtained from multiple regions. The Intra-Series Embedding Module uses Long Short-Term Memory to capture temporal relationships within each time series. Subsequently, we learn the influence degree of two embeddings and fuse them with the parametric-matrix fusion method. To further improve the robustness, SEFNet also integrates a traditional autoregressive component in parallel with nonlinear neural networks. Experiments on four real-world epidemic-related datasets show SEFNet is effective and outperforms state-of-the-art baselines.

## KEYWORDS

deep learning; epidemiological forecasting; time series

## 1 INTRODUCTION

The outbreak of an epidemic will bring huge disasters to a region and even a country. The World Health Organization (WHO) estimates that influenza annually causes approximately 3-5 million severe cases and 290,000-650,000 deaths.<sup>1</sup> In recent years, the COVID-19 pandemic has spread to more than 200 countries and territories around the world,<sup>2</sup> and the number of infections and deaths in almost all affected countries is increasing at an alarming rate. Accurately forecasting epidemics plays an essential role in allocating healthcare resources and promoting administrative planning.

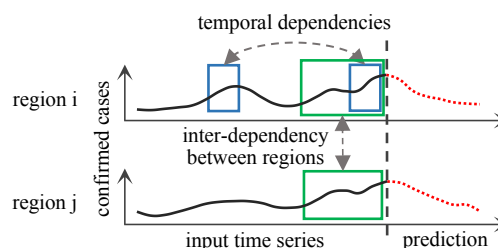
\*Corresponding Author: binzhou@nudt.edu.cn

<sup>1</sup>[https://www.who.int/en/news-room/fact-sheets/detail/influenza-\(seasonal\)](https://www.who.int/en/news-room/fact-sheets/detail/influenza-(seasonal))

<sup>2</sup><https://covid19.who.int/>

DOI reference number: 10.18293/SEKE2022-109

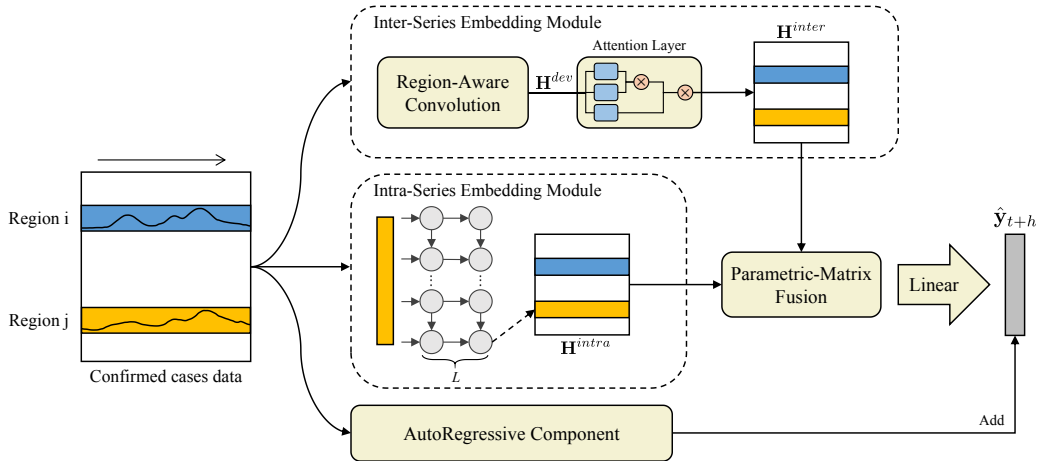
The epidemic prediction is similar to the multivariate time series forecasting task, but there are also significant differences. Multivariate time series forecasting methods inherently assume inter-dependencies among variables [21], while epidemic prediction needs to deal with unknown and complex patterns in the spread of epidemics and dynamic correlations between regions. The epidemic situation of a region at a certain time step is correlated with both its previous confirmed cases and other regions' epidemic situation. Therefore, two types of dependencies can be utilized in time series as demonstrated by Figure 1 and the epidemic time series modeling of regions can be decomposed into two parts:



**Figure 1: The blue boxes indicate the temporal dependency between time points, while the green boxes indicate the inter-dependency between regions.**

- **Inter-series embedding modeling.** There are dynamic dependencies among regions/time series. When epidemics spread in different geographic regions, it is highly likely that similar progression patterns are shared among multiple regions owing to various factors [9] (e.g., similar geographic topology or climate), and these similar patterns can aid in prediction.
- **Intra-series embedding modeling.** There are temporal dependencies within a region/time series (e.g., seasonal influenza). The epidemic development trend of one region also can be distinguished from others, this is due to region-specific factors such as government intervention, healthcare quality, climate, etc.

To date, various methods have been proposed for epidemic forecasting, but they suffer from some limitations that are bad for performance. First, using vanilla Recurrent Neural Networks (RNNs) [2,



**Figure 2: The overview of SEFNet. The original time series for each region are copied to three components: (1) Inter-Series Embedding module (top); (2) Intra-Series Embedding module (middle); and (3) AutoRegressive component (bottom).**

10] or single-scale Convolution Neural Networks (CNNs) [9, 20] is hard to capture multi-scale and complex patterns, thus resulting in a certain degree of distortion, making it difficult to extract dynamic dependencies between time series. Second, some methods are dedicated to capturing dependencies between regions by introducing the "Attention Mechanism" [2, 9, 10], but these dependencies may misguide the final prediction because the progression patterns or data distribution of different regions is not fully consistent. Therefore, we believe both inter-series dependencies and intra-series dependencies jointly contribute to epidemic forecasting, and their influence degree on prediction results varies by region.

To tackle these challenges, we propose a novel deep learning model called Inter- and Intra-Series Embeddings Fusion Network (SEFNet) that extract inter- and intra-series embeddings through two parallel modules respectively and fuse them using parametric-matrix fusion [22]. To further improve the robustness, we also integrate autoregressive component parallel to the model. Our contributions are summarized below:

- We propose a new model that extracts inter-series correlations and intra-series temporal dependencies through two separate neural networks and uses parametric-matrix fusion to emphasize the importance of each information for epidemic prediction.
- We propose a multi-scale unified convolution component called Region-Aware Convolution that is capable of extracting local, periodic, and global patterns to better obtain feature representation and capture potential dependencies between regions.
- We conduct extensive experiments on four real-world epidemic-related datasets. The results show that our model achieves better performance than other state-of-the-art methods and demonstrates the effectiveness of each component.

## 2 RELATED WORK

There has been a large body of work focusing on epidemic forecasting in literature, including statistical models [8, 13, 18], compartment models [4, 19], and swarm intelligence models [3]. In recent years, deep learning models have shown excellent performance in

various prediction tasks due to their powerful training and data-driven capabilities. CNNRNN-Res [20] is the first to apply deep learning for epidemiological prediction. Deng et al. [2] proposed Cola-GNN that treats regions as nodes in a graph and applies Graph Neural Networks (GNNs) to capture dependencies among regions. Jung et al. [10] proposed SAIFlu-Net that combines Long Short-Term Memory and self-attention to capture inter-dependencies between regions. Jin et al. [9] developed ACTs based on inter-series attention for COVID-19 forecasting. Cui et al. [1] designed a multi-range encoder-decoder framework for COVID-19 prediction. Nowadays, improving epidemic prediction is an open research problem to help the world mitigate the crisis that threatens public health.

## 3 THE PROPOSED METHOD

### 3.1 Problem Formulation

We formulate the epidemic prediction problem as a time series forecasting task. We have a total of  $N$  regions, and each region is associated with a time series input for a window  $T$ , where  $T$  is the length of historical observation data. Furthermore, we denote the epidemiology profiles  $\mathbf{X} = [\mathbf{x}_{t-T+1}, \dots, \mathbf{x}_t] \in \mathbb{R}^{N \times T}$  at time point  $t$ . An instance for region  $i$  is represented by  $\mathbf{x}_i = [x_{i,t-T+1}, \dots, x_{i,t}] \in \mathbb{R}^T$ . The goal of this task is to predict the epidemiology profile of the future time point  $t+h$ , where  $h$  is the horizon also called lead time. The proposed model SEFNet is shown in Figure 2. In the following sections, we introduce the building blocks of SEFNet in detail.

### 3.2 Intra-Series Embedding Module

The first module is Intra-Series Embedding module, which uses the historical information of time series to focus on the autocorrelation also called temporal dependencies of a single time series. In this work, we apply the Long Short-Term Memory (LSTM) [5] to capture temporal sequential dependency. The Recurrent Neural Networks (RNNs) are shown effective in sequence modeling and LSTM is a variant of RNNs, which can solve the vanishing gradients and exploding gradients problems in traditional RNNs [15]. Let  $D$  be

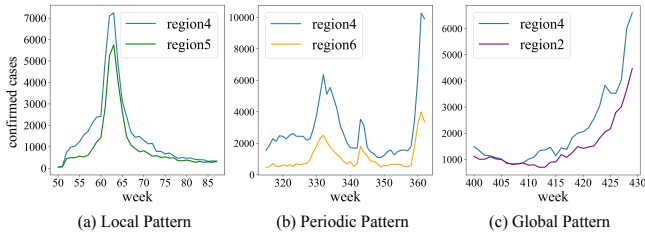
the dimension of the hidden state of LSTM, we use the original version of LSTM and formulate it as:

$$\mathbf{h}_{i,t} = \text{LSTM}(x_{i,t}, \mathbf{h}_{i,t-1}), \quad (1)$$

where  $\mathbf{h}_{i,t}$  is the output representation of region  $i$  at time point  $t$ . For each region, we use the last output of LSTM as region's intra-series embedding  $\mathbf{h}_i^{\text{intra}} \in \mathbb{R}^D$ .

### 3.3 Inter-Series Embedding Module

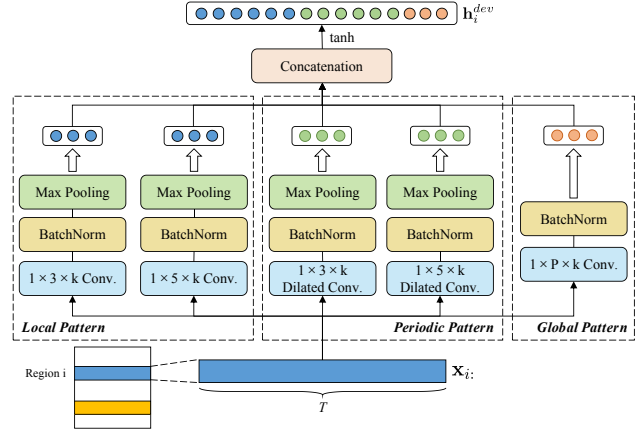
The second module is Inter-Series Embedding module which focuses on dependencies between time series. First, we obtain temporal patterns through the proposed Region-Aware Convolution (RAConv), which is a multi-scale unified convolution component. Next, we feed the output of RAConv into an attention layer to generate embeddings of dynamic dependencies between regions.



**Figure 3: The region 4 (blue line) has a dynamic pattern correlations with different regions within different time periods.**

The correlation distribution is calculated based on the feature similarity between nodes [12]. The more accurate feature describes, the better performance of the attention layer can be improved. In epidemic prediction task, there are many similar progression patterns shared among regions, such as local patterns, periodic patterns, and global patterns. Figure 3 shows different temporal patterns of influenza case trends in different Health and Human Services (HHS) regions in the United States. Inspired by the Inception [16] in computer vision, we propose a multi-scale unified component called Region-Aware Convolution (RAConv) that can extract local, periodic, and global patterns simultaneously. The structure of RAConv is shown in Figure 4. RAConv consists of three branches that apply convolution blocks with different scales or different types, thus is capable of capturing multi-scale and more complex feature patterns. Each convolution block has  $K$  filters. The local pattern branch applies standard convolution with some small kernel sizes to extract local patterns in the time series through local mapping. The periodic pattern branch inspired by skip-RNN in LSTM [11] applies dilated convolution that enables a large receptive field via dilation factor to capture the periodic pattern. Formally, the dilated convolution is a standard convolution applied to input with defined gaps. The global pattern branch applies standard convolution with the same size as  $T$  to extract time-invariant patterns of all time steps for regions [6] (e.g., time series uptrend in Figure 3c). We denote convolution filter in RAConv as  $f_{1 \times s, d}$  where  $s$  is kernel size and  $d$  is dilated factor. We empirically choose the kernel size  $s$  to  $\{3, 5, T\}$ , and dilated factor  $d$  to  $\{1, 2\}$ . We can get the local, periodic, and global features of region  $i$  by following equations:

$$\mathbf{h}_i^l = [\text{Pool}(\text{BN}(x_i \star f_{1 \times 3, 1})); \text{Pool}(\text{BN}(x_i \star f_{1 \times 5, 1}))], \quad (2)$$



**Figure 4: The structure of Region-Aware Convolution, which consists of three pattern branches.**

$$\mathbf{h}_i^p = [\text{Pool}(\text{BN}(x_i \star f_{1 \times 3, 2})); \text{Pool}(\text{BN}(x_i \star f_{1 \times 5, 2}))], \quad (3)$$

$$\mathbf{h}_i^g = \text{BN}(x_i \star f_{1 \times T, 1}), \quad (4)$$

where  $\star$  is convolution operator, and  $[\cdot]$  is concatenation operation.  $\text{Pool}(\cdot)$  is the Adaptive Max Pooling layer that can not only capture the most representative features, but also effectively reduce the amount of parameters. Adaptive Max Pooling is able to control the output size same as parameters  $P$ .  $\text{BN}(\cdot)$  is the Batch Normalization [7] layer that normalize the data and speed up convergence. The convolution operation of  $\mathbf{x}$  with  $\mathbf{f}$  at step  $j$  is represented as:

$$\mathbf{x} \star f_{1 \times s, d}(j) = \sum_{i=1}^s f_{1 \times k}(i) \mathbf{x}(j - d \times i). \quad (5)$$

Next, we concatenate three patterns and apply an element-wise activation function (e.g., hyperbolic tangent):

$$\mathbf{h}_i^{\text{dev}} = \tanh([\mathbf{h}_i^l; \mathbf{h}_i^p; \mathbf{h}_i^g]). \quad (6)$$

For each time series, we execute the above process and get the intermediate matrix called  $\mathbf{H}^{\text{dev}} \in \mathbb{R}^{N \times (P+1)K}$ .

Due to the powerful feature extraction capability of the self-attention network, we apply a typical self-attention network inspired by the Transformer [17] to capture the dependencies among regions. Let  $A$  be the dimension of inter-series embedding. We can calculate attention distribution  $\mathbf{A}$  and inter-series embedding matrix  $\mathbf{H}^{\text{inter}} \in \mathbb{R}^{N \times A}$  by following equations:

$$\mathbf{A} = \text{softmax}((\mathbf{H}^{\text{dev}} \mathbf{W}^Q)(\mathbf{H}^{\text{dev}} \mathbf{W}^K)^T), \quad (7)$$

$$\mathbf{H}^{\text{inter}} = \mathbf{A} \mathbf{H}^{\text{dev}} \mathbf{W}^V, \quad (8)$$

where  $\mathbf{W}^Q$ ,  $\mathbf{W}^K$ , and  $\mathbf{W}^V \in \mathbb{R}^{(P+1)K \times A}$  are the weight matrices that linearly map the  $\mathbf{H}^{\text{dev}}$  to query, key, and value matrices.

### 3.4 Fusion

Directly concatenating or summing inter-series embedding and intra-series embedding will have the following problems: (1) **Inconsistent scale**. Since two feature embeddings come from different neural network modules, the structural differences of each module (e.g., activation function) will lead to inconsistent scales of feature embeddings; (2) **Different importance**. Two feature embeddings describe different feature information of time series so the

importance of two feature embeddings is very different in the process of epidemiology forecasting (e.g. temporal dependency is more significant for a region with periodic recurrence of an epidemic, although there may be similar development patterns to others). Therefore, to address these problems, we adopt parametric-matrix fusion [22] to adaptively control the flow of inter-series embedding and intra-series embedding and fuse them together:

$$\mathbf{H}^{fus} = [\mathbf{W}^{inter} \circ \mathbf{H}^{inter}; \mathbf{W}^{intra} \circ \mathbf{H}^{intra}], \quad (9)$$

where  $\mathbf{H}^{fus} \in \mathbb{R}^{N \times (D+A)}$  is the output of fusion operation.  $\circ$  is element-wise multiplication.  $\mathbf{W}^{inter}$  and  $\mathbf{W}^{intra}$  are the learnable parameters that adjust the degrees affected by inter-series embedding and intra-series embedding respectively.

### 3.5 Prediction

Due to the nonlinear characteristics of Convolutional, Recurrent and self-attention components, the scale of neural network output is not sensitive to input. Meanwhile, the historical infection cases of each region are not purely nonlinear, which cannot be fully handled well by neural networks. To address these drawbacks, we retain the advantages of traditional linear models and neural networks by combining a linear part to design a more accurate and robust prediction framework inspired by [11, 14]. Specifically, we adopt the classical AutoRegressive (AR) model as the linear component in a parallel manner. Denote the forecasting result of AR component as  $\hat{y}_{t+h}^l \in \mathbb{R}^N$  that can be calculated by following equation:

$$\hat{y}_{i,t+h}^l = \sum_{m=0}^{q-1} \mathbf{W}_m^{ar} x_{i,t-m} + b^{ar}, \quad (10)$$

where  $\mathbf{W}^{ar}$  is the weight matrix and  $b^{ar}$  is the bias.  $q$  is the look-back window of AR that need be less than or equal to input window size  $T$ . Then, we feed the output after fusion operation to a dense layer to get the nonlinear part of the final prediction:

$$\hat{y}_{t+h}^n = \mathbf{H}^{fus} \mathbf{W}^n + \mathbf{b}^n. \quad (11)$$

The final prediction of model is then obtained by summing the nonlinear part and the linear part got by AR component:

$$\hat{y}_{t+h} = \hat{y}_{t+h}^n + \hat{y}_{t+h}^l. \quad (12)$$

In the training process, we adopt the Mean Square Error as the loss function that defined as:

$$\underset{\theta}{\text{minimize}} \|\mathbf{y}_{t+h} - \hat{\mathbf{y}}_{t+h}\|_2^2, \quad (13)$$

where  $\mathbf{y}_{t+h} = [y_{1,t+h}, \dots, y_{N,t+h}] \in \mathbb{R}^N$  is the true value at time point  $t+h$ , and  $\theta$  are all learnable parameters in the model.

**Table 1: Dataset statistics: min, max, mean, and standard deviation (SD) of patient counts; dataset size means the number of regions multiplied by the number of samples.**

Datasets	Size	Min	Max	Mean	SD
Japan-Prefectures	47×348	0	26635	655	1711
US-Regions	10×785	0	16526	1009	1351
US-States	49×360	0	9716	223	428
Canada-Covid	13×717	0	127199	3082	8473

## 4 EXPERIMENTS

### 4.1 Datasets and Metrics

We prepare four real-world epidemic-related datasets as follows, and their data statistics are shown in Table 1.

- **Japan-Prefectures.** This dataset is collected from the Infectious Diseases Weekly Report (IDWR) in Japan,<sup>3</sup> which contains weekly influenza-like-illness statistics from 47 prefectures, ranging from August 2012 to March 2019.
- **US-Regions.** This dataset is the ILINet portion of the US-HHS dataset,<sup>4</sup> consisting of weekly influenza activity levels for 10 HHS regions of the U.S. mainland for the period of 2002 to 2017. Each HHS region represents some collection of associated states.
- **US-States.** This dataset is collected from the Center for Disease Control (CDC).<sup>4</sup> It contains the count of patient visits for ILI (positive cases) for each week and each state in the United States from 2010 to 2017. After removing a state with missing data we keep 49 states remaining in this dataset.
- **Canada-Covid.** This dataset is publicly available at JHU-CSSE.<sup>5</sup> We collect daily COVID-19 cases from January 25, 2020 to January 10, 2022 in Canada (including 10 provinces and 3 territories).

We adopt two metrics for evaluation that are widely used in epidemic forecasting, including the **Root Mean Squared Error (RMSE)** and the **Pearson’s Correlation (PCC)**. RMSE measures the difference between predicted and true value after projecting the normalized values into the real range. PCC is a measure of the linear dependence between time series. For RMSE lower value is better, while for PCC higher value is better.

### 4.2 Methods for Comparison

We compared the proposed model with the following methods.

- **AR** The most classic statistical methods in time series analysis.
- **LRidge** The vector autoregression (VAR) with L2-regularization.
- **LSTNet [11]** A deep learning model that combines CNN and RNN to extract short- and long-term patterns.
- **TPA-LSTM [14]** An attention based LSTM network that employs CNN for pattern representations;
- **CNNRNN-Res [20]** A deep learning model that combines CNN, RNN, and residual links for epidemiological prediction.
- **SAIFlu-Net [10]** A self-attention based deep learning model for regional influenza prediction.
- **Cola-GNN [2]** A deep learning model that combines CNN, RNN and GNN for epidemiological forecasting.

### 4.3 Experimental Details

All programs are implemented using Python 3.8.5 and PyTorch 1.9.1 with CUDA 11.1 (1.9.1 cu111) in an Ubuntu server with an Nvidia Tesla K80 GPU. Our source codes are publicly available.<sup>6</sup>

**Experimental setting.** All datasets have been split into training set(50%), validation set(20%) and test set(30%). The batch size is set to 128. We use Min-Max normalization to convert data to [0,1] scale and after prediction, we denormalize the prediction value and use it

<sup>3</sup><https://tinyurl.com/y5dt7stm>

<sup>4</sup><https://tinyurl.com/y39tog3h>

<sup>5</sup><https://github.com/CSSEGISandData/COVID-19>

<sup>6</sup><https://github.com/Xiefeng69/SEFNet>

**Table 2: RMSE and PCC performance of different methods on four datasets with horizon = 3, 5, 10. Bold face indicates the best result of each column and underlined the second-best. For RMSE lower value is better, while for PCC higher value is better.**

Dataset		Japan-Prefectures			US-Regions			US-States			Canada-Covid		
		Horizon			Horizon			Horizon			Horizon		
Methods	Metrics	3	5	10	3	5	10	3	5	10	3	5	10
AR	RMSE	1705	2013	2107	757	997	1330	204	251	306	3488	4545	<u>7154</u>
	PCC	0.579	0.310	0.238	0.878	0.792	0.612	0.909	0.863	0.773	0.973	0.955	0.869
LRidge	RMSE	1711	2025	1942	870	1059	1270	276	295	324	3326	4372	7179
	PCC	0.308	0.429	0.238	0.878	0.792	0.612	0.909	0.863	0.773	0.975	0.957	0.868
LSTNet	RMSE	1459	1883	1811	801	998	1157	249	299	292	3270	6789	9561
	PCC	0.728	0.432	0.518	0.868	0.746	0.609	0.850	0.759	0.760	0.967	0.847	0.645
TPA-LSTM	RMSE	1142	1192	1677	761	950	1388	203	247	<u>236</u>	<u>2731</u>	<u>3905</u>	7671
	PCC	0.879	0.868	0.644	0.847	0.814	0.675	0.892	0.833	<u>0.849</u>	0.980	0.956	0.767
CNNRNN-Res	RMSE	1550	1942	1865	738	936	1233	239	267	260	6175	8644	9755
	PCC	0.673	0.380	0.438	0.862	0.782	0.552	0.860	0.822	0.820	0.659	0.589	0.475
SAIFlu-Net	RMSE	1356	1430	1527	661	871	1158	<u>167</u>	238	<u>236</u>	4409	7128	8514
	PCC	0.765	0.654	0.592	<u>0.903</u>	0.800	0.674	0.927	0.842	0.845	0.745	0.775	0.596
Cola-GNN	RMSE	<u>1051</u>	<b>1117</b>	<u>1372</u>	636	855	<u>1134</u>	167	202	241	2954	4036	7336
	PCC	<u>0.901</u>	<u>0.890</u>	<u>0.813</u>	<b>0.909</b>	<u>0.835</u>	<u>0.717</u>	<u>0.933</u>	<u>0.897</u>	0.822	<u>0.986</u>	<u>0.975</u>	<u>0.882</u>
SEFNet	RMSE	(↓) <b>1020</b>	<u>1123</u>	<b>1319</b>	<b>618</b>	<b>821</b>	<b>1036</b>	<b>162</b>	<b>196</b>	<b>232</b>	<b>2157</b>	<b>3339</b>	<b>7079</b>
	PCC	(↑) <b>0.904</b>	<b>0.893</b>	<b>0.826</b>	<b>0.909</b>	<b>0.842</b>	<b>0.725</b>	<b>0.935</b>	<b>0.900</b>	<b>0.833</b>	<b>0.990</b>	<b>0.978</b>	<b>0.895</b>

for evaluation. The input window size  $T$  is set to 20, and the horizon  $h$  is set to {3,5,10} in turn. All the parameters of models are trained using the Adam optimizer with weight decay  $5e-4$ , and the dropout rate is set to 0.2. We performed early stopping according to the loss on the validation set to avoid overfitting. The learning rate is chosen from {0.01,0.005,0.001}.

**Hyperparameters setting.** The hidden dimension of LSTM  $D$  and attention layer  $A$  is chosen from {16,32,64}. The number of LSTM layers  $L$  is chosen from {1,2}. The number of kernels  $K$  is chosen from {4,8,12,16}. The output dimension of Adaptive Max Pooling  $P$  is chosen from {1,3,5}. The look-back window of AR component  $q$  is chosen from {0,10,20}.

#### 4.4 Main Results

We evaluate our model in short-term (horizon = 3) and long-term (horizon = 5,10) settings. Table 2 summarizes the results of all methods. The large difference in RMSE values across different datasets is due to the scale and variance of the datasets, i.e., the scale of the Japan-Prefectures and Canada-Covid is greater than the US-Regions and US-States datasets, which is closely related to the prevalence of epidemics and population density. There is an overall trend that the prediction accuracy drops as the prediction horizon increases because the larger the horizon, the harder the problem.

We observe that the proposed SEFNet achieves the state-of-the-art results on most of the tasks. Traditional statistics methods (AR and LRidge) do not perform well in influenza-related datasets. The main reason is that they are based on oversimplified assumptions and only rely on historical records, they cannot model the strong seasonal effects in influenza datasets. For deep learning-based methods, the performance is improved since they make efforts to deal with nonlinear characteristics and complex patterns behind time series. However, some deep learning-based models work well on some datasets, while not well on others.

- (1) **Performance.** The methods mainly focused on dependencies between regions/time series (Cola-GNN and TPA-LSTM) have better performance than the method mainly focused on dependencies between time points (LSTNet), which can point out that inter-series dependencies are quite valuable information. Our proposed model takes inter-series correlations between different regions and temporal relationships in a single region into consideration and carefully fuses them for prediction to achieve better performance.
- (2) **Stability.** For Japan-Prefectures and Canada-Covid datasets, the prediction performance of many compared methods greatly decreases when the horizon increases. Because as the variance of the dataset increases, the fluctuation within time series and dependencies between regions are more intricate. SEFNet can make full use of dependencies information, so its predict error rises smoothly and slowly within the growing prediction horizon. Therefore, SEFNet has a better stability.

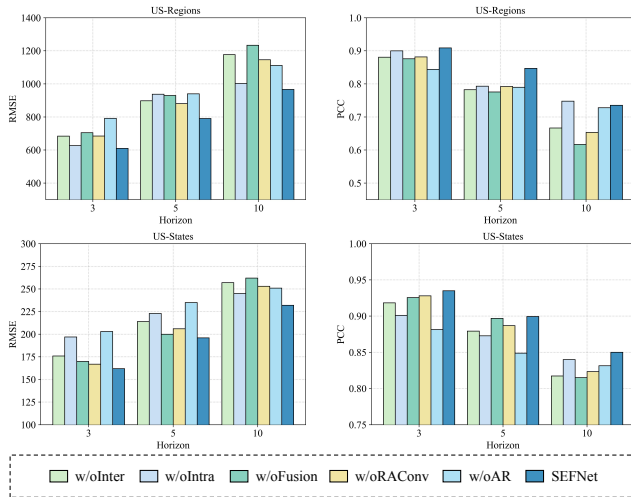
Through this experiment, it can be concluded that SEFNet has better performance and stability in epidemic forecasting, especially in long-term prediction.

#### 4.5 Ablation Study

In order to clearly verify that the above improvement comes from each added component, we conduct an ablation study on the US-Regions and US-States datasets. Specifically, we remove each component one at a time in SEFNet. We name the model without different components as follows:

- **w/oInter** The model without Inter-Series Embedding module.
- **w/oIntra** The model without Intra-Series Embedding module.
- **w/oAR** The model without the AR component.
- **w/oRACConv** The model uses  $1 \times 3$  convolution blocks only instead of Region-Aware Convolution.





**Figure 5: Results of the ablation studies on the US-Region (top) and US-States (bottom) datasets.**

- **w/oFusion** The model concatenates Inter-Series Embedding and Intra-Series Embedding directly instead of using Parametric Matrix Fusion.

The ablation results are shown in Figure 5. We highlight several observations from these results:

- (1) The full SEFNet model achieves almost the best results.
- (2) Directly concatenating two feature embeddings will bring performance drops, while the fusion method used in SEFNet can bring performance gains, especially in long-term prediction (horizon=5,10). Because when the horizon increases, the difficulty of prediction will increase accordingly, and the complex dependencies among regions are more difficult to detect. Parametric-matrix fusion adaptively learns the importance of inter- and intra-series embeddings, which facilitates capturing complex and potential relationships in long-term prediction.
- (3) Compared with single-scale convolutions, Region-Aware Convolution brings performance improvement by capturing and aggregating multi-scale features, which proves its strong feature representation power.
- (4) Removing the AR component from the full model caused significant performance drops, showing the crucial role of the AR component in general.

This ablation study concludes that our model design is the most robust across all baselines, especially with large horizons.

## 5 CONCLUSION

In this paper, we propose an Inter- and Intra-Series Embeddings Fusion Network (SEFNet) for epidemic forecasting. We first extract inter- and intra-series embeddings from two parallel modules. Specifically, in inter-series embedding module, we design a Region-Aware Convolution component that is better to extract feature representations of time series and capture the dynamic dependencies among regions. Then, we fuse two embeddings through parametric-matrix fusion for prediction. To further enhance the robustness, we apply an AutoRegressive component as the linear part. Experiments

on four real-world epidemic-related datasets show the proposed model outperforms the state-of-the-art baselines in terms of performance and stability. In future work, we plan to delve into the dynamic dependencies and mutual influences among regions.

## ACKNOWLEDGMENTS

This work is supported by the Key R&D Program of Guangdong Province No.2019B010136003 and the National Natural Science Foundation of China No. 62172428, 61732004, 61732022.

## REFERENCES

- [1] Yue Cui, Chen Zhu, Guanyu Ye, Ziwei Wang, and Kai Zheng. 2021. Into the Unobservables: A Multi-range Encoder-decoder Framework for COVID-19 Prediction. In *Proc. of CIKM*.
- [2] Songgaojun Deng, Shusen Wang, Huzefa Rangwala, Lijing Wang, and Yue Ning. 2020. Cola-gnn: Cross-location attention based graph neural networks for long-term ili prediction. In *Proc. of CIKM*.
- [3] Shuhui Guo, Fan Fang, Tao Zhou, Wei Zhang, Qiang Guo, Rui Zeng, Xiaohong Chen, Jianguo Liu, and Xin Lu. 2021. Improving Google flu trends for COVID-19 estimates using Weibo posts. *Data Science and Management* (2021).
- [4] Tiberiu Harko, Francisco SN Lobo, and MK3197716 Mak. 2014. Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates. *Appl. Math. Comput.* (2014).
- [5] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* (1997).
- [6] Siteng Huang, Donglin Wang, Xuehan Wu, and Ao Tang. 2019. DSANet: Dual self-attention network for multivariate time series forecasting. In *Proc. of CIKM*.
- [7] Sergey Ioffe and Christian Szegedy. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proc. of ICML*.
- [8] Jayson S Jia, Xin Lu, Yun Yuan, Ge Xu, Jianmin Jia, and Nicholas A Christakis. 2020. Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature* (2020).
- [9] Xiaoyong Jin, Yu-Xiang Wang, and Xifeng Yan. 2021. Inter-series attention model for covid-19 forecasting. In *Proc. of SDM*.
- [10] Seungwon Jung, Jaeuk Moon, Sungwoo Park, and Eenjun Hwang. 2021. Self-attention-based Deep Learning Network for Regional Influenza Forecasting. *IEEE Journal of Biomedical and Health Informatics* (2021).
- [11] Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. 2018. Modeling long- and short-term temporal patterns with deep neural networks. In *Proc. of SIGIR*.
- [12] Yue Liu, Wenxuan Tu, Sihang Zhou, Xinwang Liu, Linxuan Song, Xihong Yang, and En Zhu. 2022. Deep Graph Clustering via Dual Correlation Reduction. In *Proc. of AAAI*.
- [13] Edson Zangiacomini Martinez, Elisângela Aparecida Soares da Silva, and Amaury Lelis Dal Fabbro. 2011. A SARIMA forecasting model to predict the number of cases of dengue in Campinas, State of São Paulo, Brazil. *Revista da Sociedade Brasileira de Medicina Tropical* (2011).
- [14] Shun-Yao Shih, Fan-Keng Sun, and Hung-yi Lee. 2019. Temporal pattern attention for multivariate time series forecasting. *Machine Learning* (2019).
- [15] Sima Siami-Namini, Neda Tavakoli, and Akbar Siami Namin. 2019. The performance of LSTM and BiLSTM in forecasting time series. In *2019 IEEE International Conference on Big Data (Big Data)*.
- [16] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. In *Proc. of CVPR*.
- [17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Proc. of NeurIPS* (2017).
- [18] Zheng Wang, Prithwish Chakraborty, Sumiko R Mekaru, John S Brownstein, Jieping Ye, and Naren Ramakrishnan. 2015. Dynamic poisson autoregression for influenza-like-illness case count prediction. In *Proc. of KDD*.
- [19] Miguel Won, Manuel Marques-Pita, Carlota Louro, and Joana Gonçalves-Sá. 2017. Early and real-time detection of seasonal influenza onset. *PLoS computational biology* (2017).
- [20] Yuexin Wu, Yiming Yang, Hiroshi Nishiura, and Masaya Saitoh. 2018. Deep learning for epidemiological predictions. In *Proc. of SIGIR*.
- [21] Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, Xiaojun Chang, and Chengqi Zhang. 2020. Connecting the dots: Multivariate time series forecasting with graph neural networks. In *Proc. of KDD*.
- [22] Junbo Zhang, Yu Zheng, and Dekang Qi. 2017. Deep spatio-temporal residual networks for citywide crowd flows prediction. In *Proc. of AAAI*.