# A Deep Residual Network Integrating Spatial-temporal Properties to Predict Influenza Trends at an Intra-urban Scale

Guikai Xi
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
Shenzhen, China
University of Chinese Academy of Sciences
Beijing, China
gk.xi@siat.ac.cn

Ling Yin*
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
Shenzhen, China
yinling@siat.ac.cn

Ye Li
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
Shenzhen, China
ye.li@siat.ac.cn

Shujiang Mei
Shenzhen Center for Disease Control and Prevention
Shenzhen, China
sjmei@szcdc.net

## ABSTRACT

Influenza is one of the most common causes of human illness and death; thus, accurate and timely predictions for influenza trends are critical tasks for public health. Many studies have attempted to conduct influenza prediction at or beyond the city scale; however, larger spatial scales are too coarse to help analyze influenza epidemics or allow offering precise interventions inside a city. Moreover, the existing prediction models often ignore the spatial correlations of influenza activity between neighbouring regions although such correlations are potentially helpful in influenza prediction. To address the above issues, this study proposes an influenza prediction model based on a deep residual network that predicts influenza trends by integrating the spatial-temporal properties of influenza at an intra-urban scale. Using a real dataset of influenza in Shenzhen City, China, we tested our prediction model on 10 districts within the city. Our results show that our proposed deep residual model outperforms four baseline models, including linear regression (LR), artificial neural network (ANN), long short-term memory (LSTM) and spatiotemporal LSTM (ST-LSTM) models, thus demonstrating the effectiveness of the proposed prediction model. To our best knowledge, although deep-learning-based approaches have been shown to be useful in many fields in recent years, there has been no attempt to apply such approaches to influenza prediction. Therefore, this study is an initial attempt to introduce a deep learning model into influenza prediction. The proposed deep residual network is able to incorporate the spatial correlations of influenza, and it has obvious potential for making influenza predictions at finer spatial scales within a city, which can offer critical support for preciser public health interventions.

## CCS CONCEPTS

• **Information systems** → *Geographic information systems*; • **Computing methodologies** → *Supervised learning by regression*;

## KEYWORDS

Influenza prediction, deep learning, convolutional neural network, spatial-temporal properties

**ACM Reference Format:**
Guikai Xi, Ling Yin, Ye Li, and Shujiang Mei. 2018. A Deep Residual Network Integrating Spatial-temporal Properties to Predict Influenza Trends at an Intra-urban Scale. In *2nd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery (GeoAI'18), November 6, 2018, Seattle, WA, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3281548.3281558

## 1 INTRODUCTION

Influenza is caused by a virus that is both highly infectious and highly transmissible, allowing it to quickly spread worldwide. In the past 10 years, up to 650,000 deaths have been caused annually by respiratory diseases resulting from seasonal influenza[17]. In 2009, the pandemic influenza strain H1N1 spread across more than 214 countries, overseas territories, and communities. More than 18,000 deaths were reported during the first year this virus strain was in circulation[16].

Thus, establishing an effective influenza prediction framework is an urgent need and is critical for analyzing influenza trends and protecting public health. For instance, an efficient influenza prediction framework would allow health officials to deploy various preventive measures, interventions and countermeasures, assist administrators at medical institutions in making timely and optimal staffing and stocking decisions[2], and help individuals protect themselves from influenza by taking early precautions.

*The corresponding author.


*GeoAI'18, November 6, 2018, Seattle, WA, USA*

ACM ISBN 978-1-4503-6036-4/18/11...$15.00
https://doi.org/10.1145/3281548.3281558

Many regions have set up outpatient illness and virologic surveillance systems to conduct influenza surveillance[7]. Such surveillance systems can provide substantial amounts of information that are useful for influenza prediction. Although virologic surveillance systems can provide accurate influenza-confirmed illness data that can be used in influenza analysis, such systems usually have a long time lag—several weeks or even several months; thus, such systems are hysteretic and cannot be applied to timely influenza analysis. Nevertheless, prediction approaches that can provide closer to real-time influenza estimates are essential. Patient data for influenza-like illnesses (ILI) can be quickly and easily collected by clinical doctors through outpatient illness surveillance systems. This approach is more efficient, easier and allows health institutions to conduct influenza activity analyses faster and earlier. ILI percentages are computed weekly based on the totals of ILI and outpatient cases. These percentages are often used to assess influenza activities and estimate influenza trends.

In recent years, many attempts have been made to provide better estimations of ILI activity and to improve influenza prediction. Some studies developed linear models or other statistical models to assess influenza activities using single-source Internet-based data such as Yahoo[18], Google[9], Twitter[4] and Wikipedia[14]. Others have used multiple data sources to develop influenza prediction models. For example, Achrekar et al. devised auto-regression models with exogenous inputs that used historical CDC data and Twitter data to predict the ILI activity level in a population[1].

Machine learning has been successfully applied in many fields, including influenza prediction. Santillana et al. utilized historical ILI percentages reported by the CDC in conjunction with several statistics from Athenahealth that were available at least one week ahead of the reported ILI percentages to propose an autoregressive electronic health record support vector machine (SVM) model (ARES) to estimate ILI percentages at national and regional levels[19]. Xu et al. combined ILI data, Google search data and meteorological data to develop an ANN model and also considered a fusion model (a Bayesian model) of averaging several different models to improve the final performance[23].

However, most existing studies did not achieve fully satisfactory weekly influenza prediction results; moreover, they function at global[6], national [1, 6, 14, 24], US regional[9, 18, 19], or city levels[3, 23]. Such spatial scales are too coarse to support fine analysis of influenza epidemics or offer precise interventions within a city. Therefore, more accurate flu estimates at a finer spatial resolution are needed. However, to our knowledge, no study has yet been conducted at a finer spatial resolution, such as at the subregion level within a city.

Unlike temperate countries, which typically have one sharp winter peak and exhibit a clear seasonal pattern, tropical and subtropical regions have less distinct seasonal patterns, which makes influenza prediction more challenging[22]. Thus far, only a few studies have attempted to conduct influenza prediction in these regions—and the existing studies also focus on coarse spatial resolutions[3, 8, 23].

Some studies have also tested their proposed approaches in multiple neighboring regions respectively. Despite the fact that we know neighboring regions may have highly positive correlations of influenza estimates, almost all the previous studies have regarded these regions as independent, ignoring the spatial correlations between them and then trained respective models [9, 18, 19]. However, in reality, spatial correlations of influenza activity are potentially helpful for influenza predictions within multiple neighboring regions.

Deep learning has been demonstrated to be successful and competitive in many fields[5, 12, 13, 21]. In recent years, some attempts have integrated spatiotemporal properties into deep learning for specific domain applications. For example, Zhang et al. developed deep spatiotemporal networks to make citywide crowd flows prediction[25]. Thus far, however, no deep-learning-based approach that integrates the spatiotemporal properties of influenza activity has been applied to influenza prediction.

In this study, we propose a prediction model based on a deep residual network to predict influenza epidemics by integrating the spatiotemporal properties of influenza activity. We then test the proposed prediction model in Shenzhen City at a fine spatial resolution (i.e., the district level inside the city). Shenzhen is located in a subtropical region near Hong Kong and is one of the largest migratory metropolitan cities in southern China. Therefore, a refined approach for accurate influenza prediction in Shenzhen should benefit influenza prediction not only in subtropical regions but also in other areas, regardless of whether they exhibit clear seasonal patterns.

Due to the coarse spatial resolution and restricted-access policies of geotagged influenza-related data from the Internet, such as social media data, it is difficult to obtain such data for the subregions in Shenzhen. Therefore, this study uses only historical ILI activity data from a database of Shenzhen electronic medical records to integrate spatial-temporal influenza properties into the prediction model.

We summarize our contributions as follows. To the best of our knowledge, this study is the first attempt to develop a deep-learning-based model to predict influenza trends of multiple neighboring districts within an urban area. The effectiveness of the proposed model is demonstrated by the fact that it outperforms two commonly used models and two potential deep learning models. In addition, the proposed model's ability to integrate the spatial-temporal properties of influenza activities allows effective influenza predictions at finer scales within urban areas.

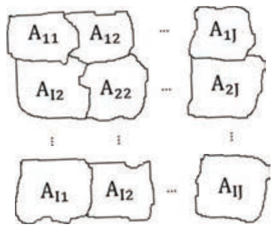
The rest of the paper is organized as follows. In Section 2, we define some basic concepts and formulate the prediction problem. We describe our approach for influenza prediction in Section 3. Section 4 presents the details of the experiments and their results. Finally, we conclude the paper with a discussion of future work in Section 5.

## 2 PROBLEM FORMULATION

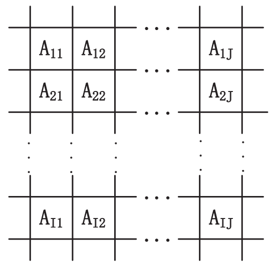
In this section, we first introduce several basic concepts that will be used throughout this work and formally define the problem that we aim to solve.

### 2.1 Basic Concepts

*Definition 2.1.* (Region). We consider a city as a region consisting of many spatial units called subregions. These subregions can be districts, blocks or defined by other divisions. In this study, we assume that the region is the area shown in Figure 1a. We can




**(a) A region with multiple irregular subregions**




**(b) The grid map corresponding to the region**

**Figure 1: A region with multiple subregions**

partition a region into an $I$x$J$ grid map as shown in Figure 1b based on the spatial positions and spatial topologies between the subregions. Here, $I$ denotes the number of rows and $J$ denotes the number of columns in the grid map, and $I$x$J$ is equal to the sum of the subregions. Each square of the grid map forms a spatial unit. Then, it is easy to construct some 2-dimensional information matrixes corresponding to the grid map.

*Definition 2.2.* (Time series). A time series is a series of data points indexed by the time sequence in which they occurred. For grid($i$, $j$), which lies at the $i$-th row and the $j$-th column of the grid map, we can define a time series as follows:

$$X_{i,j}^{t-2,s} = \left\{X_{i,j}^{t-2}, X_{i,j}^{t-3}, \ldots, X_{i,j}^{t-s-1}\right\}$$

where $X_{i,j}^{t}$ represents the ILI percentage of grid($i$, $j$) at week $t$, and $s$ is the temporal lag. Here, $X_{i,j}^{t-2,s}$ represents the ILI percentages of grid($i$,$j$) for the $s$ weeks preceding week *t-2*.

*Definition 2.3.* (Spatial-temporal Data) The spatial matrix of week $t$ can be denoted as $X^t \in \mathbb{R}^{I \times J}$, which records the ILI percentages of every spatial unit of the $I$x$J$ grid map at week $t$. The spatial-temporal (ST) data, also called the ST matrix, is considered as a group of many spatial matrixes at different timestamps and can be denoted as

$$X_{ST}^{t-2,s} = \left\{X^{t-2}, X^{t-3}, \ldots, X^{t-s-1}\right\}$$

The spatial-temporal matrix $X_{ST}^{t-2,s}$ can also be regarded as a matrix consisting of many historical time series related to multiple spatial units.

## 2.2 Problem Statement

Because there is a time lag of at least one week for the reported ILI percentages, given the time series $X_{i,j}^{t-2,s}$, the goal is to predict $X_{i,j}^{t+m}$, $m \in (0, 1)$. Similarly, to make a prediction for every spatial unit in the grid map, given the spatial-temporal matrix $X_{ST}^{t-2,s}$, the goal is to predict $X^{t+m}$, $m \in (0, 1)$ (see Figure 2). This approach provides a one-week-ahead influenza prediction (nowcast) for the current week when $m$ equals 0 and a two-week-ahead influenza prediction (forecast) for the following week when $m$ equals 1.

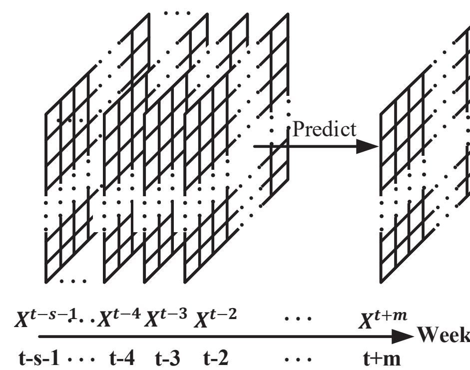



**Figure 2: Spatial-temporal matrix for future spatial matrix predictions**

# 3 METHODOLOGY

## 3.1 Spatial-temporal Properties

In time series prediction, there is no doubt that historical and non-random data have considerable autocorrelations and will contribute to future predictions. As we all know, influenza is caused by influenza viruses and does not occur arbitrarily. Thus, for a region or a subregion, the time series consisting of the ILI percentages of several previous weeks is a non-stationary series that harbors important information about future influenza activities. Also, in a time series, data with two adjacent timestamps are likely to have a higher similarity than data with more distant timestamps, which means that influenza activities from adjacent timestamps provide more timely information for the target prediction.

Influenza is infectious and also easily influenced by climate. Adjacent regions usually have similar epidemic features; consequently, influenza activity in one subregion provides important information for the analysis of influenza activities in neighboring subregions. At the spatial level, for example, it implies that the ILI percentage of a given subregion is either increasing or is likely to increase in future weeks when the ILI percentages in its neighboring subregions are increasing. Moreover, the current ILI percentages of the neighboring subregions can also be reflected by their own historical data. Consequently, the current ILI percentage of a subregion can be reflected by the historical ILI percentages of its neighboring subregions. These relationships involve spatial dimensions.

To briefly summarize, the ILI percentages of a subregion in a current or future week are reflected by the historical data of that subregion at the temporal level and by the historical data of neighboring subregions at the spatial-temporal level. We can easily construct a 3-dimensional matrix as defined in Definition 2.3 (the first two dimensions are a 2-dimensional space and the third is the time

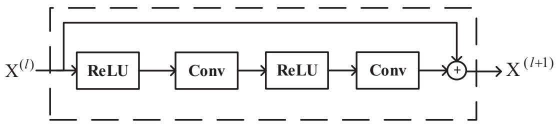



**Figure 3: Residual unit**

dimension) from the collected data corresponding to the grid map and use it for influenza modeling and prediction.

## 3.2 Convolution and Residual Units

In this study, we apply one type of deep learning model, a convolutional neural network (CNN), to influenza prediction. A convolutional layer is an important part of a CNN that can capture the spatial dependencies in neighboring subregions. Adding more convolutional layers can further capture the spatial dependencies of more distant subregions. Generally, deep models can be beneficial for difficult tasks. Sometimes, however, adding additional layers to a deep model leads to higher errors during training because of the degradation problem. When deeper networks are able to start converging, their accuracy can become saturated, but residual learning can address this problem. Using identity mappings obtained via shortcuts can be effective in training deep networks[10]. It has been proved that residual units, which are the most important component of deep residual networks using pre-activation, are better than the units using post-activation[11].

## 3.3 Structure of the Deep Model

Rectified linear units (ReLU) $f(x) = max(0, x)$ have nonsaturating nonlinearity. Training deep convolutional neural networks that use ReLUs can be several times faster than training equivalent models that use tanh units[13, 15]. Small convolutional filters also result in a significant performance improvement for deep convolutional neural networks[20].

In this study, we stack $L$ residual units using pre-activation as shown in Figure 3 to build our deep residual network. Each residual unit is defined as follows:

$$X^{(l+1)} = X^l + F\left(X^l\right), l = 1, 2, 3, \ldots, L,$$

where $F$ denotes the residual function. Our model is shown in Figure 4. Assuming that the grid map is $I$x$J$, the input of the model is $X_{ST}^{t-2,s} \in \mathbb{R}^{I \times J \times s}$ and the output should be $X^{t+m} \in \mathbb{R}^{I \times J \times 1}$. Other details are listed in Table 1. All the convolutional layers including Conv1 and Conv2 use 3x3 convolution filters, and the number of filters in every convolutional layer is 64. The first convolutional layer, Conv1, is designed to increase the number of input feature maps for the first residual unit, because $X^{(l)}$ and $X^{(l+1)}$ should have the same dimensions to allow element-wise addition. The last convolutional layer, Conv2, decreases the number of output feature maps for the last residual unit from 64 to 1 for the prediction target because we want to predict the target at only a specific time. All the residual units will attempt to learn the nonlinear features from the input data.

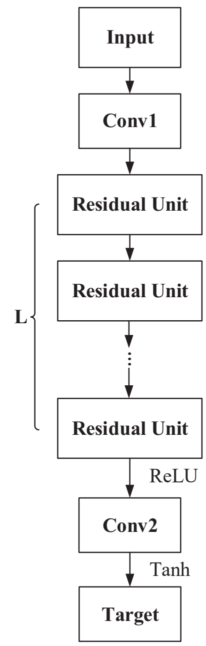



**Figure 4: Deep residual network**

**Table 1: Architectures of our model**

| layer name | filter param | output size |
|---|---|---|
| Conv1 | 3x3,64 | |
| ResUnit1 | 3x3,64<br>3x3,64 | |
| ResUnit2 | 3x3,64<br>3x3,64 | IxJx64 |
| ... | ... | |
| ResUnitL | 3x3,64<br>3x3,64 | |
| Conv2 | 3x3,1 | IxJx1 |

# 4 EXPERIMENT

We evaluated the proposed model using a real dataset. In this section, we first introduce the dataset, the baselines and the performance measurements. Then, we present the results and discuss them.

## 4.1 Dataset

The dataset used in this experiment was extracted from an electronic medical records database provided by Shenzhen Medical Information Center—a unit directly responsible to the Health and Family Planning Commission of Shenzhen Municipality. The original dataset consists of every outpatient case, including every ILI case, from Shenzhen hospitals and community health service centers (CHSCs) from January 1, 2015 to August 30, 2017. From the original dataset, we obtained the daily number of both ILI cases

Table 2: Dataset details

| Item | Comments |
|---|---|
| Location | Shenzhen |
| Date span | 1/1/2015 - 8/30/2017 |
| Number of days | 973 |
| Number of weeks | 139 |
| Number of recorded hospitals or CHSCs | 716 |

and outpatients from all the hospitals and CHSCs. Based on the locations of the hospitals and the CHSCs, we were able to calculate the ILI percentages for all ten districts of Shenzhen by week, as shown in Figure 5. All the details of the dataset are shown in Table 2.

## 4.2 Baselines

Thus far, no experiments have been conducted using this dataset. Therefore, in this experiment, we compare the proposed model with four baseline machine learning models, including two commonly-used models (linear regression (LR) and the artificial neural network (ANN)) and two other potential deep learning models (long short-term memory (LSTM) and spatiotemporal LSTM (ST-LSTM)) regarding their accuracies on influenza prediction tasks.

- **LR**: Linear regression is a classical and well-known model for time series modeling.
- **ANN**: The artificial neural network utilizes the temporal features of each spatial unit input into the network to forecast a predicted target for each individual spatial unit.
- **LSTM**: Long short-term memory model is a variant of recurrent neural network (RNN). Like the ANN, the LSTM also utilizes the temporal features of each spatial unit input into the network. However, unlike an ANN, the LSTM treats the input features as a time series and takes temporal correlations into consideration when making influenza predictions.
- **ST-LSTM**: In contrast to the LSTM, the ST-LSTM model not only aggregates the temporal features of each spatial unit input into the network but also the temporal features of several nearby districts.

## 4.3 Preprocessing

Before the final output of the deep model, which is shown in Figure 4, we use a tanh layer whose range is between -1 and 1 as our final activation layer. Thus, we use the min-max normalization method to scale all the data into the range [-1, 1].

Also, to build the spatial matrix of Shenzhen, whose corresponding map is shown in Figure 6, considering the spatial positions of the districts and the spatial topology between the districts, we divide the city of Shenzhen into a district matrix with two rows and five columns. The details of how this spatial matrix corresponds to Shenzhen are shown in Table 3.

## 4.4 Performance Measurements

Measuring the prediction performance using only mean absolute error is not a reasonable approach to compare the performance differences of the various models because such metrics focus only on the degree of deviation, which is not a fair approach for those that achieve lower true values. Therefore, we use both mean absolute error and mean absolute percentage error to evaluate and compare the performance of the different models:

- **MAE**: Mean absolute error is defined as follows:

$$MAE = \frac{1}{z}\sum_{i}|v_i - \widehat{v}_i|$$

- **MAPE**: Mean absolute percentage error is defined as

$$MAPE = \frac{1}{z}\sum_{i}\frac{|v_i - \widehat{v}_i|}{v_i} \times 100\%$$

where $v_i$ and $\widehat{v}_i$ represent the true value and the predicted value, respectively, and $z$ is the total number of predicted values.

As a measure of fit, lower MAE and MAPE scores imply better forecasting performance. The MAE shows the degree of deviation directly, but MAPE plays a more significant role in measuring prediction performance than does MAE in this experiment.

## 4.5 Evaluation Results

In this section, we perform some experiments with the proposed model and then discuss the advantages of our model based on the results. We divide the dataset into two parts: training data and testing data. Due to the peculiarities of the influenza virus, the testing data consists of the last 52 weeks of the year from September 1, 2016 to August 30, 2017. This approach is fairer and more accurate for evaluating the prediction model throughout the year, regardless of whether the ILI percentage level is high.

First, we calculate the correlation coefficients for the time series formed by the ILI percentages between every pair of districts in the training data. The correlation matrix in Table 4 shows that all the neighboring subregions have strong spatial correlations. These results reflect that the spatial correlation of influenza that occurs in subregions at a certain spatial resolution is quite strong.

After calculating the autocorrelation coefficients for all the time series consisting of the ILI percentages from 10 districts of Shenzhen in the training data, we find that when the temporal lag $s$ is greater than 8, almost half the autocorrelation coefficients are close to 0.1. Thus, in the subsequent experiments, we empirically set the temporal lag $s$ as 8 uniformly. The autocorrelation coefficients are shown in Table 5.

Our model with ResNet attempts to predict the ILI percentages of all the districts through a unified model. We trained the model on a GPU and averaged the results after executing the experiments 5 times to avoid bias introduced by the decimal calculation of MAPE.

Using the residual units improves the prediction performance of the model. The more residual units in the model, the better performance is. However, as the number of the residual units increases, eventually, adding residual units results in minimal performance improvements but leads to a continuous increase in the amount of model computation. Therefore, we investigated to determine the best number of residual units that both improve the performance

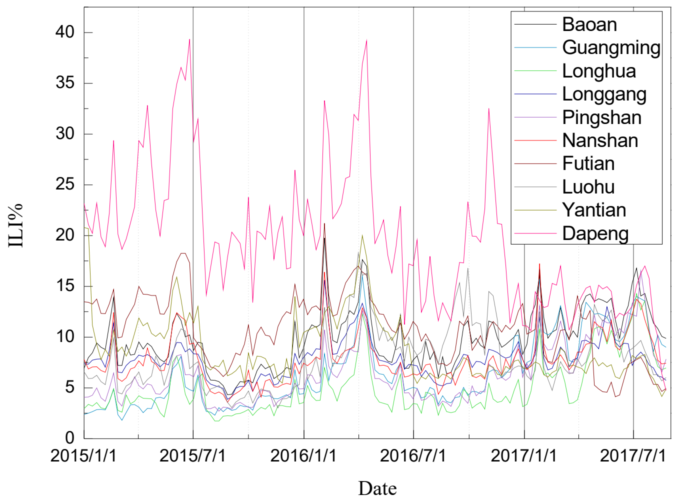



**Figure 5: All the ILI data from ten districts of Shenzhen**


**Figure 6: Shenzhen City in China**

**Table 3: The spatial matrix of Shenzhen**

| | Col. 1 | Col. 2 | Col. 3 | Col. 4 | Col. 5 |
|---|---|---|---|---|---|
| Row 1 | Baoan | Guangming | Longhua | Longgang | Pingshan |
| Row 2 | Nanshan | Futian | Luohu | Yantian | Dapeng |

and also limit the computation insofar as possible. We compared the performances with different numbers of residual units to determine the smallest $L$ that results in substantially better performance. As shown in Table 6, the average performance for 10 districts improves as $L$ increases until $L$ equals 10. Consequently, we set L to 10 in the subsequent experiments when making one-week-ahead and two-week-ahead predictions.

We trained the LR, ANN, LSTM and ST-LSTM models independently for different districts to maximize the prediction performance in all ten districts, as shown in Figure 7. We trained all the baseline models on a CPU using multicore parallel programming techniques.

Both the proposed model and the baseline models were dynamically retrained every week to capture new influenza activities and achieve better predictions. In addition, we performed data augmentation by moving the start of the unnatural week, resulting in a nearly sixfold expansion of the training data. All the results are shown in Tables 7 and 8. For space reasons, Figure 8 shows the results of only the one-week-ahead predictions for all ten Shenzhen districts.

As shown in Table 7, LR is good at capturing the temporal correlation when it can gain sufficient data from some adjacent timestamps. Although the LR model is simpler than the other four models, it

**Table 4: The correlation matrix**

| | Baoan | Guangming | Longhua | Longgang | Pingshan | Nanshan | Futian | Luohu | Yantian | Dapeng |
|---|---|---|---|---|---|---|---|---|---|---|
| Baoan | 1 | 0.76 | 0.86 | 0.95 | 0.91 | 0.89 | 0.78 | 0.71 | 0.66 | 0.7 |
| Guangming | 0.76 | 1 | 0.87 | 0.74 | 0.89 | 0.66 | 0.55 | 0.83 | 0.61 | 0.55 |
| Longhua | 0.86 | 0.87 | 1 | 0.79 | 0.89 | 0.77 | 0.64 | 0.72 | 0.68 | 0.7 |
| Longgang | 0.95 | 0.74 | 0.79 | 1 | 0.91 | 0.9 | 0.84 | 0.74 | 0.72 | 0.73 |
| Pingshan | 0.91 | 0.89 | 0.89 | 0.91 | 1 | 0.85 | 0.76 | 0.77 | 0.73 | 0.76 |
| Nanshan | 0.89 | 0.66 | 0.77 | 0.9 | 0.85 | 1 | 0.84 | 0.7 | 0.69 | 0.79 |
| Futian | 0.78 | 0.55 | 0.64 | 0.84 | 0.76 | 0.84 | 1 | 0.54 | 0.69 | 0.78 |
| Luohu | 0.71 | 0.83 | 0.72 | 0.74 | 0.77 | 0.7 | 0.54 | 1 | 0.56 | 0.43 |
| Yantian | 0.66 | 0.61 | 0.68 | 0.72 | 0.73 | 0.69 | 0.69 | 0.56 | 1 | 0.69 |
| Dapeng | 0.7 | 0.55 | 0.7 | 0.73 | 0.76 | 0.79 | 0.78 | 0.43 | 0.69 | 1 |

**Table 5: The autocorrelation coefficients for all ten districts of Shenzhen**

| Week lag | Baoan | Guangming | Longhua | Longgang | Pingshan | Nanshan | Futian | Luohu | Yantian | Dapeng |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.74 | 0.85 | 0.72 | 0.83 | 0.79 | 0.77 | 0.82 | 0.85 | 0.83 | 0.73 |
| 2 | 0.57 | 0.72 | 0.4 | 0.7 | 0.62 | 0.59 | 0.69 | 0.78 | 0.74 | 0.55 |
| 3 | 0.47 | 0.59 | 0.22 | 0.59 | 0.52 | 0.41 | 0.55 | 0.67 | 0.6 | 0.42 |
| 4 | 0.42 | 0.52 | 0.18 | 0.53 | 0.49 | 0.33 | 0.43 | 0.57 | 0.52 | 0.3 |
| 5 | 0.32 | 0.44 | 0.13 | 0.46 | 0.38 | 0.26 | 0.33 | 0.53 | 0.44 | 0.16 |
| 6 | 0.29 | 0.39 | 0.14 | 0.4 | 0.28 | 0.19 | 0.26 | 0.49 | 0.34 | 0.11 |
| 7 | 0.29 | 0.35 | 0.15 | 0.35 | 0.22 | 0.14 | 0.24 | 0.4 | 0.27 | 0.14 |
| 8 | 0.28 | 0.29 | 0.14 | 0.31 | 0.22 | 0.14 | 0.17 | 0.4 | 0.22 | 0.15 |
| 9 | 0.26 | 0.27 | 0.09 | 0.27 | 0.23 | 0.11 | 0.11 | 0.34 | 0.11 | 0.09 |
| 10 | 0.16 | 0.18 | -0.02 | 0.18 | 0.12 | 0 | 0.05 | 0.32 | -0.02 | 0.03 |

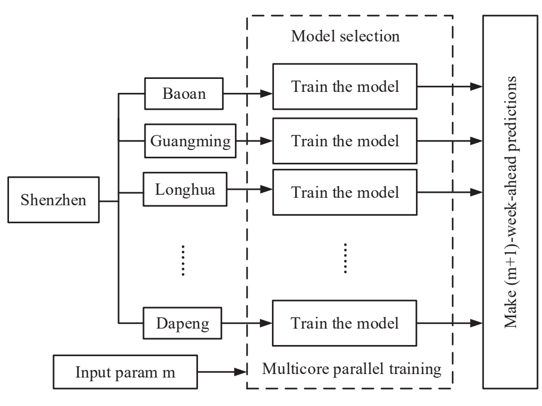



**Figure 7: Training procedure of the baselines**

does not result in the worst average MAE and MAPE scores when making one-week-ahead predictions. However, due to the lack of timely data from nearby timestamps, LR resulted the worst average performance when making two-week-ahead predictions.

**Table 6: Performance comparison as the number of residual units varies**

| Number of Residual Units | One-week-ahead | | Two-week-ahead | |
|---|---|---|---|---|
| | MAE | MAPE | MAE | MAPE |
| 4 | 0.0174 | 19.02% | 0.0184 | 20.30% |
| 8 | 0.0167 | 18.11% | 0.0175 | 19.30% |
| 10 | 0.0163 | 17.78% | 0.0172 | 18.94% |
| 12 | 0.0164 | 17.84% | 0.0172 | 19.01% |

By comparing the results of ANN with LR, we can see that the ANN's prediction performance is poor in terms of both MAE and MAPE when making one-week-ahead predictions. In this case, the ANN's performance in most of the districts is worse than that of LR. And it is the same for its average performance. However, when making two-week-ahead predictions, the ANN is not only better than LR in most of the districts but also at the average level. This result occurs because an ANN is good at capturing the high-dimensional nonlinear features even when insufficient timely data is available.

Compared with the ANN, when making one-week-ahead and two-week-ahead predictions, the LSTM, which considers the temporal correlations in the data, achieves a slightly better average MAPE—although there is no significant difference between the average MAE between the two models. The performance of the LSTM is better than that of the ANN in half of the districts. Thus, taking the temporal correlations into consideration results in a valid but nonobvious effect.

With the help of the ILI data from the nearby districts, the ST-LSTM outperforms the LSTM, achieving a better average performance in all the districts and a better performance for most of the individual districts in Shenzhen. The advantage of ST-LSTM is more obvious when making one-week-ahead predictions.

The proposed ResNet utilizes the convolutional layers to extract the deep spatial correlation features. When the deep CNN has more convolutions, the proposed ResNet can capture distant spatial dependencies—even the city-wide dependencies. As shown in Tables 7 and 8, ResNet achieves a better performance than the other four baseline models on both average MAE and average MAPE. Moreover, ResNet achieves an improved performance in half the individual districts.

## 5 CONCLUSIONS AND FUTURE WORK

In this study, we propose a model based on a deep residual network to predict influenza trends for multiple neighboring districts within an urban area by integrating the spatiotemporal properties of influenza activities. We conducted experiments using a real dataset of ILI activities in Shenzhen City, China. Our results show that the proposed model performs better for one-week-ahead and two-week-ahead predictions than the other four baseline models. We showed that strong spatial correlations exist between the 10 districts of Shenzhen City. Moreover, influenza's spatiotemporal properties do indeed help to improve the performance of our proposed model compared with the performances of the other baselines.

In the future, we will consider applying our model to a finer spatial resolution, such as traffic analysis zones or communities, which can further enhance the spatial precision of influenza prediction, thus better supporting precise intervention. Moreover, the proposed prediction approach only uses influenza data itself as model input. Since existing influenza models have already revealed that there are many factors that contribute to the spread of influenza, such as weather and population movements, we can integrate those factors in our prediction model to improve the prediciton performance.

## ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (Grant No.: 41771441, 61433012 and U1435215) and the Basic Research Program of Shenzhen (Grant No. JCYJ20170307164104491). And it is also supported by Joint Engineering Research Center for Health Big Data Intelligent Analysis Technology.

## REFERENCES

[1] Harshavardhan Achrekar, Avinash Gandhe, Ross Lazarus, Ssu-Hsin Yu, and Benyuan Liu. 2011. Predicting flu trends using twitter data. In *Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on*. IEEE, 702–707.

[2] John S Brownstein and Kenneth D Mandl. 2006. Reengineering real time outbreak detection systems for influenza epidemic monitoring. In *AMIA Annual Symposium Proceedings*, Vol. 2006. American Medical Informatics Association, 866.

[3] Pei-Hua Cao, Xin Wang, Shi-Song Fang, Xiao-Wen Cheng, King-Pan Chan, Xi-Ling Wang, Xing Lu, Chun-Li Wu, Xiu-Juan Tang, Ren-Li Zhang, et al. 2014. Forecasting influenza epidemics from multi-stream surveillance data in a subtropical city of China. *PloS one* 9, 3 (2014), e92945.

[4] Aron Culotta. 2010. Towards detecting influenza epidemics by analyzing Twitter messages. In *Proceedings of the first workshop on social media analytics*. ACM, 115–122.

[5] George E Dahl, Dong Yu, Li Deng, and Alex Acero. 2012. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *IEEE Transactions on audio, speech, and language processing* 20, 1 (2012), 30–42.

[6] Gunther Eysenbach. 2006. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. In *AMIA Annual Symposium Proceedings*, Vol. 2006. American Medical Informatics Association, 244.

[7] US Centers for Disease Control and Prevention. 2017. Overview of Influenza Surveillance in the United States. Retrieved September 1, 2018 from https://www.cdc.gov/flu/weekly/overview.htm

[8] Nicholas Generous, Geoffrey Fairchild, Alina Deshpande, Sara Y Del Valle, and Reid Priedhorsky. 2014. Global disease monitoring and forecasting with Wikipedia. *PLoS computational biology* 10, 11 (2014), e1003892.

[9] Jeremy Ginsberg, Matthew H Mohebbi, Rajan S Patel, Lynnette Brammer, Mark S Smolinski, and Larry Brilliant. 2009. Detecting influenza epidemics using search engine query data. *Nature* 457, 7232 (2009), 1012.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Identity mappings in deep residual networks. In *European Conference on Computer Vision*. Springer, 630–645.

[12] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. 2014. Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 1725–1732.

[13] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*. 1097–1105.

[14] David J McIver and John S Brownstein. 2014. Wikipedia usage estimates prevalence of influenza-like illness in the United States in near real-time. *PLoS computational biology* 10, 4 (2014), e1003581.

[15] Vinod Nair and Geoffrey E Hinton. 2010. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*. 807–814.

[16] World Health Organization. 2010. Pandemic (H1N1) 2009 - update 100. Retrieved August 8, 2018 from http://www.who.int/csr/don/2010_05_14/en/

[17] World Health Organization. 2018. Up to 650 000 people die of respiratory diseases linked to seasonal flu each year. Retrieved August 8, 2018 from http://www.who.int/news-room/headlines/14-12-2017-up-to-650-000-people-die-of-respiratory-diseases-linked-to-seasonal-flu-each-year

[18] Philip M Polgreen, Yiling Chen, David M Pennock, Forrest D Nelson, and Robert A Weinstein. 2008. Using internet searches for influenza surveillance. *Clinical infectious diseases* 47, 11 (2008), 1443–1448.

[19] Mauricio Santillana, AT Nguyen, Tamara Louie, Anna Zink, Josh Gray, Iyue Sung, and John S Brownstein. 2016. Cloud-based electronic health records for real-time, region-specific influenza surveillance. *Scientific reports* 6 (2016), 25732.

[20] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[21] Yi Sun, Xiaogang Wang, and Xiaoou Tang. 2014. Deep learning face representation from predicting 10,000 classes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1891–1898.

[22] Cécile Viboud, Wladimir J Alonso, and Lone Simonsen. 2006. Influenza in tropical regions. *PLoS medicine* 3, 4 (2006), e89.

[23] Qinneng Xu, Yulia R Gel, L Leticia Ramirez Ramirez, Kusha Nezafati, Qingpeng Zhang, and Kwok-Leung Tsui. 2017. Forecasting influenza in Hong Kong with Google search queries and statistical model fusion. *PloS one* 12, 5 (2017), e0176690.

[24] Qingyu Yuan, Elaine O Nsoesie, Benfu Lv, Geng Peng, Rumi Chunara, and John S Brownstein. 2013. Monitoring influenza epidemics in china with search query from baidu. *PloS one* 8, 5 (2013), e64323.

[25] Junbo Zhang, Yu Zheng, and Dekang Qi. 2017. Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction.. In *AAAI*. 1655–1661.


**Figure 8: One-week-ahead prediction results**

**Table 7: Comparison of different models in ten Shenzhen districts (one-week-ahead predictions)**

**(a) MAE**

| Districts | LR | ANN | LSTM | ST-LSTM | ResNet |
|---|---|---|---|---|---|
| Baoan | 0.0167 | 0.0177 | 0.0175 | **0.0146** | 0.0155 |
| Guangming | 0.0170 | 0.0190 | 0.0185 | 0.0164 | **0.0133** |
| Longhua | 0.0187 | 0.0181 | 0.0203 | 0.0181 | **0.0162** |
| Longgang | **0.0104** | 0.0107 | 0.0106 | 0.0121 | 0.0110 |
| Pingshan | 0.0171 | 0.0171 | 0.0185 | 0.0176 | **0.0150** |
| Nanshan | 0.0150 | 0.0157 | 0.0148 | 0.0146 | **0.0135** |
| Futian | 0.0188 | 0.0192 | 0.0200 | 0.0195 | **0.0167** |
| Luohu | **0.0192** | 0.0215 | 0.0201 | 0.0226 | 0.0214 |
| Yantian | 0.0093 | 0.0083 | **0.0081** | 0.0090 | 0.0094 |
| Dapeng | 0.0328 | 0.0352 | 0.0353 | **0.0305** | 0.0315 |
| Average | 0.0175 | 0.0183 | 0.0184 | 0.0175 | **0.0163** |

**(b) MAPE**

| Districts | LR | ANN | LSTM | ST-LSTM | ResNet |
|---|---|---|---|---|---|
| Baoan | 14.06% | 14.86% | 14.25% | **12.47%** | 13.45% |
| Guangming | 17.40% | 19.45% | 18.38% | 18.23% | **14.64%** |
| Longhua | 27.91% | 26.32% | 27.92% | 27.00% | **23.14%** |
| Longgang | **12.41%** | 12.80% | 12.52% | 13.94% | 13.06% |
| Pingshan | **19.69%** | 19.73% | 20.93% | 20.38% | 17.55% |
| Nanshan | 16.07% | 16.28% | 15.64% | 15.20% | **14.07%** |
| Futian | 26.55% | 27.43% | 27.28% | 24.17% | **21.48%** |
| Luohu | **22.23%** | 22.82% | 22.47% | 26.41% | 23.38% |
| Yantian | 14.51% | 13.37% | **12.58%** | 14.09% | 14.47% |
| Dapeng | 26.32% | 27.58% | 27.87% | 24.72% | **22.29%** |
| Average | 19.72% | 20.06% | 19.98% | 19.66% | **17.78%** |

**Table 8: Comparison of different models in ten Shenzhen districts (two-week-ahead predictions)**

**(a) MAE**

| Districts | LR | ANN | LSTM | ST-LSTM | ResNet |
|---|---|---|---|---|---|
| Baoan | 0.0184 | 0.0173 | 0.0178 | 0.0162 | **0.0155** |
| Guangming | 0.0201 | 0.019 | 0.0201 | 0.0176 | **0.0143** |
| Longhua | 0.0217 | 0.0192 | 0.0227 | 0.0241 | **0.0186** |
| Longgang | 0.0109 | 0.0116 | **0.0103** | 0.0123 | 0.0114 |
| Pingshan | 0.0197 | 0.0179 | 0.0203 | 0.0192 | **0.0156** |
| Nanshan | 0.0160 | 0.0151 | 0.0140 | 0.0150 | **0.0133** |
| Futian | 0.0216 | 0.0226 | 0.0226 | 0.0206 | **0.0180** |
| Luohu | 0.0229 | 0.023 | **0.0217** | 0.0278 | 0.0225 |
| Yantian | 0.0114 | 0.0094 | **0.0084** | 0.0099 | 0.0103 |
| Dapeng | 0.0406 | 0.0376 | 0.0376 | **0.0329** | 0.0330 |
| Average | 0.0203 | 0.0193 | 0.0196 | 0.0196 | **0.0172** |

**(b) MAPE**

| Districts | LR | ANN | LSTM | ST-LSTM | ResNet |
|---|---|---|---|---|---|
| Baoan | 15.20% | 14.88% | 14.60% | 13.51% | **13.31%** |
| Guangming | 19.73% | 19.43% | 20.28% | 18.95% | **15.47%** |
| Longhua | 30.41% | 26.27% | 30.05% | 31.04% | **26.13%** |
| Longgang | 12.93% | 13.33% | **12.05%** | 14.27% | 13.66% |
| Pingshan | 22.50% | 20.47% | 22.32% | 22.29% | **17.57%** |
| Nanshan | 16.76% | 15.79% | 14.08% | 15.62% | **14.06%** |
| Futian | 31.96% | 33.54% | 32.57% | 26.51% | **24.23%** |
| Luohu | 26.24% | 25.19% | **24.36%** | 28.92% | 25.47% |
| Yantian | 18.11% | 15.02% | **12.97%** | 15.53% | 15.80% |
| Dapeng | 33.46% | 31.24% | 30.53% | 26.06% | **23.75%** |
| Average | 22.73% | 21.52% | 21.38% | 21.27% | **18.95%** |