
Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Fundamental Research 2 (2022) 311–320

Contents lists available at ScienceDirect

Fundamental Research

journal homepage: http://www.keaipublishing.com/en/journals/fundamental-research/

Article

# Adaptively temporal graph convolution model for epidemic prediction of multiple age groups

Yuejiao Wang[a,b], Dajun Daniel Zeng[a,b,c], Qingpeng Zhang[d], Pengfei Zhao[a], Xiaoli Wang[e], Quanyi Wang[e], Yin Luo[a,b,c], Zhidong Cao[a,b,c,*]

[a] The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China
[b] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China
[c] Shenzhen Artificial Intelligence and Data Science Institute (Longhua), Shenzhen, China
[d] School of Data Science, City University of Hong Kong-Hong Kong SAR, China
[e] Institute for Infectious Disease and Endemic Disease Control, Beijing Center for Disease Prevention and Control, Beijing, China

ARTICLE INFO

*Article history:*
Received 28 April 2021
Received in revised form 27 June 2021
Accepted 21 July 2021
Available online 8 August 2021

*Keywords:*
Graph convolution model
Infectious disease prediction
Multiple age group
Multivariate time series
Public health

ABSTRACT

*Introduction:* Multivariate time series prediction of infectious diseases is significant to public health, and the deep learning method has attracted increasing attention in this research field.
*Material and methods:* An adaptively temporal graph convolution (ATGCN) model, which learns the contact patterns of multiple age groups in a graph-based approach, was proposed for COVID- 19 and influenza prediction. We compared ATGCN with autoregressive models, deep sequence learning models, and experience- based ATGCN models in short-term and long-term prediction tasks.
*Results:* Results showed that the ATGCN model performed better than the autoregressive models and the deep sequence learning models on two datasets in both short-term (12.5% and 10% improvements on RMSE) and long-term (12.4% and 5% improvements on RMSE) prediction tasks. And the RMSE of ATGCN predictions fluctuated least in different age groups of COVID- 19 (0.029 ± 0.003) and influenza (0.059±0.008). Compared with the Ones-ATGCN model or the Pre-ATGCN model, the ATGCN model was more robust in performance, with RMSE of 0.0293 and 0.06 on two datasets when horizon is one.
*Discussion:* Our research indicates a broad application prospect of deep learning in the field of infectious disease prediction. Transmission characteristics and domain knowledge of infectious diseases should be further applied to the design of deep learning models and feature selection.
*Conclusion:* The ATGCN model addressed the multivariate time series forecasting in a graph-based deep learning approach and achieved robust prediction on the confirmed cases of multiple age groups, indicating its great potentials for exploring the implicit interactions of multivariate variables.

## 1. Introduction

The struggle between humans and infectious diseases always exists: influenza [1] causes 290,000–650,000 deaths worldwide each year [2]; COVID- 19 [3] is prone to spread widely due to unknown causes [4] and lack of effective treatment options [5]. Multivariate time series prediction [6] of infectious diseases is essential for emergency management of public health: onset time and outbreak size prediction determine the vaccination time [7] and the reserve of public health resources [8]; when facing emerging infectious diseases, predicting the epidemic trend is conducive to formulating prevention and control measures [9].

With the rise of big data, the patients' age, gender, and geographic location can be used for prediction. Many studies have indicated that the epidemic prediction of multiple age groups has important research value. Firstly, prediction by the age group is conducive to accurate prevention and control measures, because people of different age groups show immune heterogeneity to the pathogen [10,11], and their clinical manifestations and mortality rates are also various [33,50]. For example, children 0 to 14 years of age are less susceptible to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection than adults 15 to 64 years of age, whereas individuals more than 65 years of age are more susceptible to infection [33]; Influenza A subtype H3N2 causes more cases and deaths [12] in older adults than influenza A subtype H1N1 in America; and children aged 0-2 years are more susceptible to hand, foot and mouth disease (HFMD) than children aged 3–5 years in China [13].

* Corresponding author.
*E-mail address:* zhidong.cao@ia.ac.cn (Z. Cao).

https://doi.org/10.1016/j.fmre.2021.07.007

More importantly, the contact patterns of multiple age groups can help to improve the epidemic predictions, because human social contacts largely determine the transmission mode of infectious diseases and the time series of cases in multiple age groups have implicit associations. For example, the coresidence patterns of elderly persons in Africa increase their vulnerability to deaths of COVID-19 [14]; bring of students together at the start of the school year can produce annual cycles in disease transmission [45]; and gathering for holidays can provide opportunities for transmission of pneumococcus from children to older relatives [46].

Popular mathematical methods for multivariate time series prediction of infectious disease mainly include time-series autoregression models and compartmental disease transmission dynamic models. The classical autoregression models contain the autoregressive (AR) model [15] and global autoregressive (GAR) model [17], which uses observations from previous time steps as input to a regression equation to predict the value at the next time step. Such simple idea makes them fail to exploit latent interdependencies among variables. The vector autoregressive (VAR) model [16] can only consider the linear relationship between variables, which overlooks the complex nonlinear interdependencies among variables. Classical compartmental dynamic models such as SIR (Susceptible, Infectious, or Recovered) and SEIR (Susceptible, Exposed, Infectious, or Recovered) cannot tackle the age influence on the disease spread. Age-stratified compartmental dynamic models [47–49] are a group of models that are widely used to predict disease transmission dynamics under different intervention scenarios by constructing differential equations for each age group independently, which also neglect the interdependences between different age groups.

Recurrent neural networks (RNNs) are widely applied for multivariate time series prediction due to the ability to learn data representations and handle the uncertainty and non-linear problems. LSTM model or bidirectional LSTM model learned from the assembled data (e.g., incidence data, google trend, air pollution data, and lockdown measures) to forecast the epidemic trend [18,19]. Combined with the CNN module, the CNNRNN-Res model [20] and convolutional LSTM-based spatiotemporal model [21] extracted geographic interactions of influenza or COVID- 19 to improve prediction accuracy. However, the LSTM-based model projects interactions among variables into a global hidden state, which is implicit and uninterpretable. And the CNN module is only applicable to multivariate time series with geographical adjacency and is not suitable for learning the interaction of abstract variables, such as multiple age groups.

On the other hand, graph neural networks (GNNs) are increasingly used in feature extraction and prediction of spatial-temporal time series: MRes-RGNN [22], ASTGCN [23] and STGCN [24] capture graph-based spatial-temporal dependencies jointly for traffic flow prediction. There are also several studies that use GNN to integrate population migration data for infectious disease prediction [25–28]. Different from RNNs, multivariate time series are processed in a graph structure in GNN, so the independence of time series can be defined by interaction weights and be learned in a data-driven way [29].

To this end, inspired by GNN and the epidemic knowledge of age groups, we innovatively propose an adaptively temporal graph convolution (ATGCN) model. ATGCN mainly overcomes two limitations of existing methods: (1) different from RNNs, ATGCN models the interactions of time series into a directed graph, which is explicit and interpretable; (2) ATGCN extends the concept of the geographic adjacency matrix in CNN and GNN modules into the interaction among abstract variables, e.g., the contact patterns of age groups.

Taking COVID- 19 and influenza as prediction objects, this study aims to compare the performance of the proposed ATGCN with the autoregressive models and deep sequence learning models (models with RNNs as the core module) in short-term and long-term prediction tasks, and to explore the consistency of their performance in different age groups. Our major contributions are as follows:

(1) We propose an adaptively temporal graph convolution model, which is a novel approach for epidemic prediction;
(2) This paper addresses the multivariate time series prediction in a graph-based approach to exploit interdependencies among variables;
(3) For the first time, this paper incorporates the domain knowledge of epidemic on multiple age groups to guide the design of the deep learning model.

## 2. Material and methods

### 2.1. Datasets for case studies

Two multivariate time series datasets, the COVID-19 dataset of Maryland, USA and the influenza dataset of Beijing, China are used to compare the effectiveness of models in infectious diseases prediction.

**(1) The COVID-19 dataset of Maryland, USA**. The dataset consists of time series of daily new confirmed cases of COVID-19 of multiple age groups, and it is updated every week by the state of Maryland. We downloaded it from the "Dataset Freshness Dashboard" (https://opendata.maryland.gov/stories/s/xdqw-5b5w) on January 1, 2021. Specifically, the confirmed cases are divided into nine age groups according to their ages: 0-9, 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, and over 80 years old. Because of the delay in data aggregation, the effect of working days, and other uncertain factors [30], there are sawtooth fluctuations in the time series of daily new confirmed cases. So this study carries out seven-day average on the original dataset. And the final dataset for prediction (from April 5, 2020 to January 1, 2021) is obtained, as shown in Fig. 1a.

**(2) The influenza dataset of Beijing, China**. As shown in Fig. 1b, the dataset consists of time series of the weekly number of influenza cases of five age groups, which are 0–4, 5–14, 15–24, 25–59 and over 60 years old. It spans 263 weeks from September 30, 2013 to October 14, 2018. The influenza dataset is extracted from the system of influenza-like-illness (ILI) surveillance, which is conducted by outpatient and emergency departments in all general hospitals in Beijing, reporting for the weekly number of ILI cases by age groups. The number of influenza cases is calculated using the reported number of ILI cases from the ILI surveillance system, multiplied by the proportion of ILI cases that are positive for influenza in each week [31], and the positive rate for influenza is provided by 24 sentinel hospitals through the virological test.

There are two reasons for choosing these two datasets for case studies:

(1) The first reason is that the datasets can reflect two different contact patterns of multiple age groups: due to the regulation from the government policies, the COVID-19 dataset represents a situation that the average work and travel frequency of residents decrease. Meanwhile, on the contrary, their average social distance increase [50]. While the influenza dataset represents a social situation without non-pharmaceutical interventions, and the contact pattern of multiple age groups is more complex [33].
(2) The second reason is that the two datasets have different temporal characteristics: The COVID-19 dataset has no obvious periodicity but with an upward trend (Fig. 1). There are three peak outbreaks in the dataset, and the scale of the third round of outbreaks is much higher than the previous two rounds, indicating that the epidemic in Maryland was still getting worse at that time. On the contrary, the influenza dataset has obvious annual periodicity, even though the scale of the outbreak varies from year to year. These two different temporal characteristics could challenge the prediction performance of a model. Because some models, e.g., the autoregressive model, are suitable for prediction tasks with dataset that has strong autocorrelation [34], while other models, such as RNN model, are adept at periodic time series predictions, we want to use these two types of datasets to verify the practicability of our proposed model.

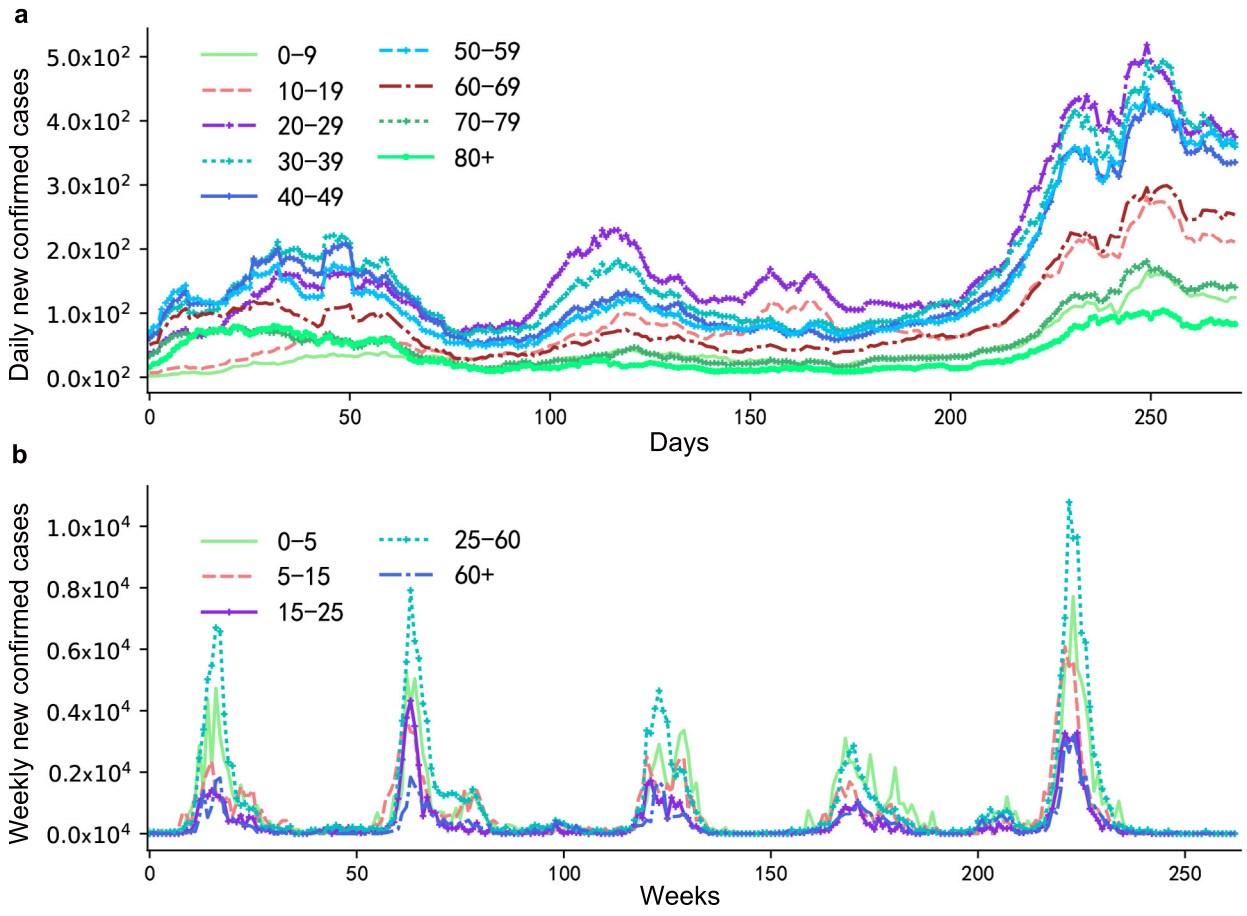


**Fig. 1. New confirmed cases of COVID-19 and influenza.** (a) Daily new confirmed cases of COVID-19 in Maryland, USA; (b) weekly new confirmed cases of influenza in Beijing, China.

### 2.2. Task definition and basic assumptions

Our study defines the epidemic prediction problem as a multivariate time series prediction task. $\boldsymbol{x}_t \in \mathbb{R}^{\boldsymbol{m}}$ is the multivariate time series profile at time $t$, whose elements are the new confirmed cases from $m$ age groups. Further, we denote $\mathbf{X} = [\boldsymbol{x}_{t+1}, \boldsymbol{x}_{t+2}, \cdots, \boldsymbol{x}_{t+w}]$ as the available input data in a time-span of size $w$ for model training. And the prediction target is the time series profile $\boldsymbol{Y} = \boldsymbol{x}_{t+w+h} \in \mathbb{R}^{\boldsymbol{m}}$, where $h$ refers to the horizon of prediction.

Since the deep learning model learns the dependences of variables from historical data and make prediction of future trend based on training data, our study assumes that the transmissibility of SARS- CoV-2 variants and influenza strains as well as human immunity to the pathogen is similar over the study period, and the human contact patterns of multiple age groups do not significantly change over this period.

### 2.3. Framework of ATGCN model

To discover the contact patterns among age groups, the ATGCN model treats their contact relationship as a directed graph $\boldsymbol{G} = (V, E)$, where $V$ is the set of nodes and the $E$ is the set of edges. $v_i \in V$ denotes a node, which represents an age group in our study. And $e_{j,i} \in E$ denotes a weighted edge pointing from $v_i$ to $v_j$, which means the contact weight from the age group $i$ to the age group $j$. The adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ is a mathematical representation of $G$, with $\boldsymbol{A}_{j,i} = e_{j,i}$. Because of the heterogeneity of immunity to the virus in different age groups [35], we assume that $e_{j,i} \neq e_{i,j}$ in epidemic prediction tasks.

The diagram of the ATGCN model proposed in our study is shown in Fig. 2. The model is composed of an adaptive graph learning module, a LSTM module, a two-layer GCN module and two linear transformation layers as output module. By calculating the similarity among multivariate time series, the adaptive graph learning module computes a graph adjacency matrix $\mathbf{A}$, which is used later as an input to the GCN module. At the same time, the LSTM module extracts the long-distance periodic feature $\boldsymbol{H}$ from the input $\mathbf{X}$ in the time domain. Then, combined with $\boldsymbol{A}$, $\boldsymbol{H}$ is converted from a Euclidean structure to a directed graph structure $\boldsymbol{G}$, which works as the input of the two-layer GCN module. In addition, the trend of the time series itself cannot be ignored: the linear prediction ($\hat{\mathbf{X}}_1$) of $\mathbf{X}$ is calculated through the linear transformation module. Finally, $\hat{\mathbf{X}}_1$ and the output of the GCN module $\hat{\mathbf{X}}_2$ are summed up to obtain the final prediction result $\hat{\boldsymbol{Y}}$.

Python 3.7.3 and the deep learning platform PyTorch 1.7.1 are used to build the ATGCN model, and its key components are illustrated in detail in the following:

(1) *The adaptive graph learning module.* Constructing the graph adjacency matrix $\boldsymbol{A}$ is the premise of GCN operation. There are explicit or latent dependencies among multiple time series. For example, in the traffic flow prediction, there is a fixed topological relationship among the roads [34]. In the prediction of confirmed cases in multiple age groups, the contact pattern of multiple age groups is simultaneously affected by contact frequency, immune heterogeneity and other factors, so there are no explicit dependencies among age groups. If the model can automatically extract the contact pattern among multiple age groups, the latent

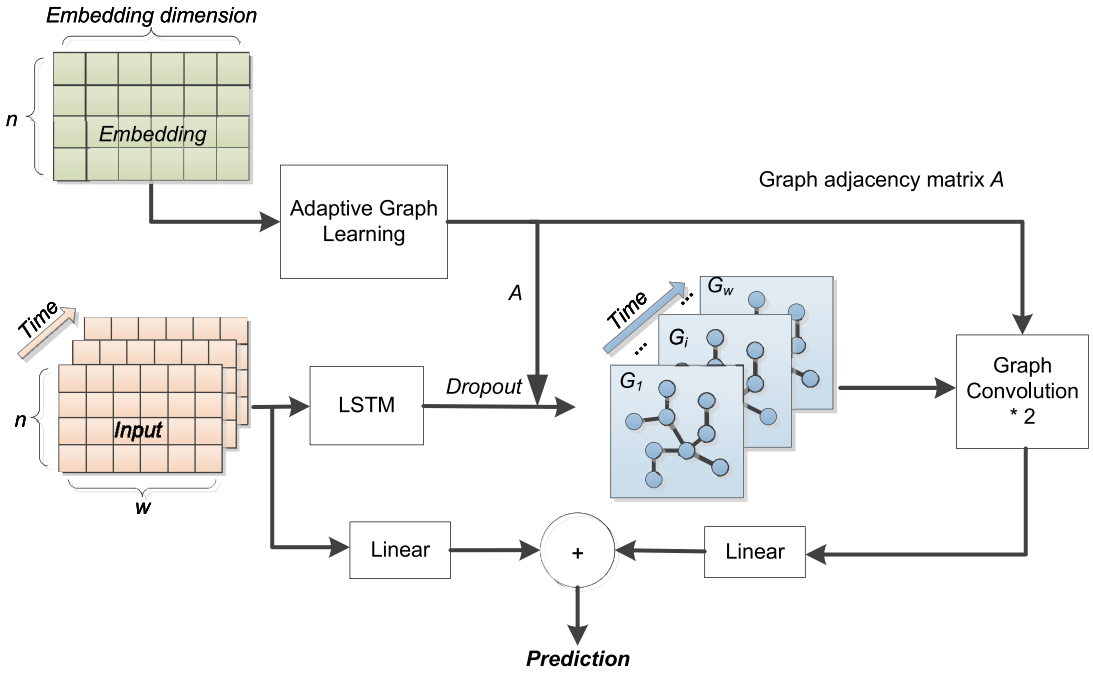


**Fig. 2. The structure of the adaptively temporal graph convolution model.**

dependencies of multivariate time series can be combined to improve the prediction. Our study adopts the adaptive graph learning module [29], which is illustrated as follows:

$$\boldsymbol{M}_1 = tanh\left(\alpha \boldsymbol{E}_1 \boldsymbol{\Theta}_1\right),$$

$$\boldsymbol{M}_2 = tanh\left(\alpha \boldsymbol{E}_2 \boldsymbol{\Theta}_2\right),$$

$$\boldsymbol{A} = \mathrm{ReLU}\left(\tanh\left(\alpha\left(\boldsymbol{M}_1 \boldsymbol{M}_2^T - \boldsymbol{M}_2 \boldsymbol{M}_1^T\right)\right)\right),$$

where $\boldsymbol{E}_1$, $\boldsymbol{E}_2$ represent randomly initialized node embeddings, which are learnable during training; $\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2$ are model parameters; $\alpha$ is a hyperparameter for controlling the saturation rate of the activation function. The adaptive graph learning module measures the similarity between the embeddings of each node and computes the matrix $\boldsymbol{A}$ of the directed graph $\boldsymbol{G}$.

(2) *The LSTM module.* The LSTM module aims to extract the long-distance periodic feature $\boldsymbol{H}$ from the input $\mathbf{X}$ in the time domain. For each element in the input sequence, each layer computes the following function:

$$\boldsymbol{i}_t = \sigma\left(\boldsymbol{W}_{ii}\boldsymbol{x}_t + b_{ii} + \boldsymbol{W}_{hi}\boldsymbol{h}_{t-1} + b_{hi}\right),$$

$$\boldsymbol{f}_t = \sigma\left(\boldsymbol{W}_{if}\boldsymbol{x}_t + b_{if} + \boldsymbol{W}_{hf}\boldsymbol{h}_{t-1} + b_{hf}\right),$$

$$\boldsymbol{g}_t = tanh\left(\boldsymbol{W}_{ig}\boldsymbol{x}_t + b_{ig} + \boldsymbol{W}_{hg}\boldsymbol{h}_{t-1} + b_{hg}\right),$$

$$\boldsymbol{o}_t = \sigma\left(\boldsymbol{W}_{io}\boldsymbol{x}_t + b_{io} + \boldsymbol{W}_{ho}\boldsymbol{h}_{t-1} + b_{ho}\right),$$

$$\boldsymbol{c}_t = \boldsymbol{f}_t \odot \boldsymbol{c}_{t-1} + \boldsymbol{i}_t \odot \boldsymbol{g}_t,$$

$$\boldsymbol{h}_t = \boldsymbol{o}_t \odot tanh\left(\boldsymbol{c}_t\right),$$

where $\boldsymbol{h}_t$ is the hidden state at time $t$, $\boldsymbol{c}_t$ is the cell state at time $t$, $\boldsymbol{x}_t$ is the input at time $t$, $\boldsymbol{h}_{t-1}$ is the hidden state of the layer at time $t$-1 or the initial hidden state at time 0, and $\boldsymbol{i}_t$, $\boldsymbol{f}_t$, $\boldsymbol{g}_t$, $\boldsymbol{o}_t$ are the input, forget, cell, and output gates, respectively. $\sigma$ is the sigmoid function, and $\odot$ is the hadamard product.

(3) *The GCN module.* The two-layer graph convolution module can effectively convolve the second-order neighborhoods of the nodes to handle spatial dependencies in a graph. The graph convolutional operator is illustrated as follows:

$$\boldsymbol{H}' = \hat{\boldsymbol{D}}^{-1/2}\hat{\boldsymbol{A}}\hat{\boldsymbol{D}}^{-1/2}\boldsymbol{H}\boldsymbol{\Theta},$$

where $\hat{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{I}$ denotes the adjacency matrix with inserted self-loops and $\hat{\boldsymbol{D}}_{ii} = \sum_{j=0} \hat{\boldsymbol{A}}_{ij}$ its diagonal degree matrix.

Its node-wise formulation is given by:

$$\boldsymbol{h}_i' = \boldsymbol{\Theta} \sum_{j \in \mathcal{N}(v) \cup \{i\}} \frac{e_{j,i}}{\sqrt{\hat{d}_j \hat{d}_i}} \boldsymbol{h}_j,$$

with $\hat{d}_i = 1 + \sum_{j \in \mathcal{N}(i)} e_{j,i}$, where $e_{j,i}$ denotes the edge weight from source node $j$ to target node $i$.

(4) *The output module.* The LSTM module and GRU module learn to capture the long-term dependence and the latent interaction among variables, but the historical information of the input sequence cannot be ignored. So, one linear transformation layer in the output module is applied to the input $\mathbf{X}$:

$$\hat{\boldsymbol{X}}_1 = \boldsymbol{X}\boldsymbol{C}^T + \boldsymbol{b},$$

where $\hat{\boldsymbol{X}}_1$ denotes the linear prediction, $\boldsymbol{C}^T$ is the learnable weights, and $\boldsymbol{b}$ is the learnable bias of the linear transformation layer. Similarly, the output $\boldsymbol{H}'$ of the GRU module is dimensionally transformed by another linear transformation layer to obtain $\hat{\mathbf{X}}_2$:

$$\hat{\boldsymbol{X}}_2 = \boldsymbol{H}'\boldsymbol{C}_2^T + \boldsymbol{b}_2,$$

The final prediction result of the ATGCN model is the sum of $\hat{\boldsymbol{X}}_1$ and $\hat{\boldsymbol{X}}_2$:

$$\hat{\boldsymbol{Y}} = \hat{\boldsymbol{X}}_1 + \hat{\boldsymbol{X}}_2.$$

### 2.4. Baseline models

The performance of the ATGCN model is evaluated by comparing with baseline models, which are summarized as follows:

#### 2.4.1. Autoregressive models

The AR model, VAR model, and GAR model have been the most popular models in time series prediction. Their core idea is to predict the future state using the linear combination of historical states, so the autoregressive models are suitable for predicting time series with obvious trends, e.g., the COVID-19 dataset of Maryland, USA.

The AR model treats the new confirmed cases of different age groups independently, i.e., it assumes different age groups are not correlated. The AR model is formalized as follows:

$$\hat{x}_{t+h}^{(i)} = \sum_{p=0}^{w-1} \alpha_p^{(i)} x_{t-p}^{(i)} + \varepsilon_{t+h} + c^{(i)},$$

where $p$ is the order of AR, w is the length of input window, $x_t^{(i)}$ is the input signal from the $i$-th age group, and $\alpha_p^{(i)}$ is the weight parameter. $\varepsilon_{t+h}$ denotes the random noise at time $t+h$, and $c^{(i)}$ is the intercept term.

VAR can model the dependencies across different age groups, and it is more complex and expressive. VAR is formalized as follows:

$$\hat{\boldsymbol{x}}_{t+h} = \sum_{p=0}^{w-1} W_p \boldsymbol{x}_{t-p} + \boldsymbol{\varepsilon}_{t+h} + \boldsymbol{c}$$

where $\boldsymbol{x}_t$ is the input vector containing the historical sequences of all age groups, and $W_p$ is the parameter matrix to capture the correlation among age groups.

The GAR model is a simplification of the AR model. GAR only uses one set of $\alpha_p$ and $c$ to predict different age groups, and it is suitable for scenarios where the training data is limited and the multivariate time series exhibit similar patterns. GAR is formalized as follows:

$$\hat{\boldsymbol{x}}_{t+h} = \sum_{p=0}^{w-1} \alpha_p \boldsymbol{x}_{t-p} + \varepsilon_{t+h} + c.$$

#### 2.4.2. Deep sequence learning models

We apply the recurrent neural network (RNN) model, convolutional RNN (CNNRNN) model, and residual CNNRNN (CNNRNN-Res) model [20] to learn the long-term dependencies among age groups. These three models all rely on RNN as the core module for sequence learning and prediction, which are suitable for predicting time series with obvious periodicity, e.g., the influenza dataset of Beijing, China.

Our study utilizes the gated recurrent unit (GRU) in our RNN model. Unlike the LSTM module described above, GRU contains fewer parameters, and its iterative speed is faster:

$$\boldsymbol{r}_t = \sigma\left(\boldsymbol{W}_{ir}\boldsymbol{x}_t + b_{ir} + \boldsymbol{W}_{hr}\boldsymbol{h}_{t-1} + b_{hr}\right),$$

$$\boldsymbol{z}_t = \sigma\left(\boldsymbol{W}_{iz}\boldsymbol{x}_t + b_{iz} + \boldsymbol{W}_{hz}\boldsymbol{h}_{t-1} + b_{hz}\right),$$

$$\boldsymbol{n}_t = tanh\left(\boldsymbol{W}_{in}\boldsymbol{x}_t + b_{in} + \boldsymbol{r}_t \odot \left(\boldsymbol{W}_{hn}\boldsymbol{h}_{t-1} + b_{hn}\right)\right),$$

$$\boldsymbol{h}_t = \left(1 - \boldsymbol{z}_t\right) \odot \boldsymbol{n}_t + \boldsymbol{z}_t \odot \boldsymbol{h}_{t-1},$$

where $\boldsymbol{r}_t$, $\boldsymbol{z}_t$, and $\boldsymbol{n}_t$ are the reset, update, and new gates, respectively.

The CNNRNN model adds a CNN layer before the GRU. CNN layer performs a two-dimensional convolution operation [36] on the input $\boldsymbol{x}_t$ to fuse multivariate time series. We can distinguish the difference between the CNN and GCN module from the perspective of multiple age groups: CNN processes data in the Euclidean structure, so it can only integrate the contact information between adjacent age groups. While GCN processes data in a graph structure, it can integrate the contact information among any two age groups, e.g., patients aging between 60–69 and 20–29, as long as there are interaction weights between them.

To avoid overfitting, the CNNRNN-Res model adopts residual links [37] on the basis of the CNNRNN model. The residual links bypass some of the intermediate modules, which can effectively mitigate the gradient vanishing phenomenon during training. For more details of the three deep sequence learning models, please refer to the research paper by CMU [20].

#### 2.4.3. Experience-driven ATGCN models

In order to verify the effectiveness of the adaptive graph learning module in the ATGCN model, we replace the graph learning module with two experience-driven adjacency matrices and input them into the ATGCN model. The two experience-driven matrices are:

(1) A square matrix $A_1 \in \mathbb{R}^{m \times m}$ whose elements are one. The setting of $A_1$ is based on the simplest assumption that the contact patterns between all age groups are the same. And the ATGCN model based on $A_1$ is abbreviated as Ones-ATGCN.

(2) A matrix that records the contact frequency of multiple age groups in China or USA, which is obtained through social contact surveys. For the prediction of the COVID-19 dataset of Maryland, we use the matrix (predefined-$A_2 \in \mathbb{R}^{m \times m}$) of the contact frequency of people in multiple age groups in Maryland, calculated based on human mobility surveillance system during the COVID- 19 epidemic and previous social survey results [32] (please refer to supplementary material 1 for the detailed calculation process of $A_2$). As for the prediction of the influenza dataset of Beijing, we adopt a contact matrix (predefined-$A_3 \in \mathbb{R}^{m \times m}$) of multiple age groups in Shanghai during the normal period [33], assuming that the population contact frequency in Shanghai and Beijing is similar (in supplementary material 2). And the ATGCN model based on these two predefined matrices is abbreviated as Pre-ATGCN.

### 2.5. Experimental setting

All datasets are split into three sets: training set (60%), validation set (20%), and test set (20%) in chronological order. We tune the input window size for all models from the set {4, 8, 16, 32}. For three deep sequence learning models, we tune the hidden dimension of GRU from {5, 10, 20, 40}. The number of residual links is searched from set {4, 8, 16}. The parameter optimization algorithm is Adam, and its learning rate is searched from {0.01, 0.015, 0.02}. All models iterate 300 times on the training set, with prediction horizon= {1, 2, 4, 6, 8}, respectively. Finally, we select the model that performs best on the validation set to make predictions on the test set. In this paper, the predictive period ranges from one day or one week (short-term prediction) to 8 days or 8 weeks (long-term prediction). The long-term period (8 days or 8 weeks) is relative to short-term period (1 day or 1 week). Thus, the ATGCN model is accurate in both within-year and inter-annual prediction.

We adopt three metrics for comparison: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R- squares (R2). The calculation of them is defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2},$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|,$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}_i\right)^2},$$

where n denotes the number of data points in the test set, $y_i$ denotes the true value, and $\hat{y}_i$ denotes the predicted value. The best possible score of R2 is 1.0, and it can be negative (because the model can be arbitrarily worse). And if an R2 score is close to 0, then the model always predicts an expected value of y, disregarding the input features.

## 3. Results

Table 1 provides the experimental results of ATGCN model and baseline models. ATGCN model achieves state-of-the-art results on most of the tasks, whether it is the COVID-19 dataset with obvious trends or the influenza dataset with periodicity. In the following, we will discuss the experimental results of the two datasets, respectively.

### 3.1. Predictions of the COVID-19

From Table 1, we observe that the ATGCN model has the highest overall prediction accuracy for the COVID-19 dataset (12.5% and 12.4%

**Table 1**
**Baseline comparison for the COVID-19 dataset prediction and the influenza dataset prediction.**

| Dataset | | COVID-19 | | | | | Influenza | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Models | Metrics | Horizon (day) | | | | | Horizon (week) | | | | |
| | | 1 | 2 | 4 | 6 | 8 | 1 | 2 | 4 | 6 | 8 |
| AR | RMSE | 0.0378 | 0.0615 | 0.1379 | 0.1914 | 0.2388 | 0.0845 | 0.1321 | 0.2114 | 0.2401 | 0.2465 |
| | MAE | 0.0295 | 0.0489 | 0.1159 | 0.1615 | 0.2040 | 0.0431 | 0.0663 | 0.1091 | 0.1316 | 0.1413 |
| | $R^2$ | 0.9452 | 0.8546 | 0.2696 | -0.408 | -1.190 | 0.8804 | 0.7076 | 0.2516 | 0.0352 | -0.017 |
| VAR | RMSE | 0.0616 | 0.1042 | 0.2483 | 0.2958 | 0.3307 | 0.0863 | 0.1129 | 0.2030 | 0.2350 | 0.2436 |
| | MAE | 0.0473 | 0.0863 | 0.2093 | 0.2550 | 0.2970 | 0.0455 | 0.0636 | 0.1298 | 0.1383 | 0.1497 |
| | $R^2$ | 0.8542 | 0.5827 | -1.368 | -2.360 | -3.201 | 0.8754 | 0.7866 | 0.3101 | 0.0753 | 0.0068 |
| GAR | RMSE | 0.0335 | 0.0532 | 0.0910 | 0.1465 | 0.1798 | 0.0838 | 0.1260 | 0.2086 | 0.2368 | 0.2461 |
| | MAE | 0.0253 | 0.0415 | 0.0738 | 0.1207 | 0.1501 | 0.0417 | 0.0628 | 0.1062 | 0.1296 | 0.1366 |
| | $R^2$ | 0.9569 | 0.8914 | 0.6819 | 0.1755 | -0.242 | 0.8825 | 0.7342 | 0.2713 | 0.0615 | -0.014 |
| RNN | RMSE | 0.1096 | 0.1736 | 0.2023 | 0.2657 | 0.3097 | 0.0890 | 0.1095 | 0.1554 | 0.2415 | 0.2829 |
| | MAE | 0.0872 | 0.1416 | 0.1723 | 0.2382 | 0.2868 | 0.0389 | 0.0527 | **0.0825** | 0.1456 | 0.1538 |
| | $R^2$ | 0.5386 | -0.158 | -0.572 | -1.711 | -2.683 | 0.8675 | 0.7994 | 0.5957 | 0.0236 | -0.340 |
| CNNRNN | RMSE | 0.2100 | 0.2013 | 0.1949 | 0.3108 | 0.3055 | 0.0745 | 0.1017 | 0.1592 | 0.2155 | 0.2779 |
| | MAE | 0.1708 | 0.1630 | 0.1662 | 0.2439 | 0.2494 | 0.0336 | 0.0481 | 0.0857 | 0.1244 | 0.1764 |
| | $R^2$ | -0.694 | -0.557 | -0.460 | -2.709 | -2.585 | 0.9071 | 0.8269 | 0.5756 | 0.2227 | -0.293 |
| CNNRNN-Res | RMSE | 0.1047 | 0.1094 | 0.1292 | 0.1947 | 0.2667 | 0.0669 | 0.0937 | 0.1603 | 0.2252 | 0.2363 |
| | MAE | 0.0821 | 0.0856 | 0.1078 | 0.1652 | 0.2329 | **0.0313** | 0.0466 | 0.0868 | 0.1349 | 0.1388 |
| | $R^2$ | 0.5791 | 0.5407 | 0.3588 | -0.456 | -1.731 | 0.9251 | 0.8529 | 0.5699 | 0.1509 | 0.0655 |
| Ones-ATGCN | RMSE | 0.0311 | 0.0488 | 0.0907 | 0.1362 | 0.1784 | 0.0743 | 0.1120 | 0.1802 | 0.2247 | 0.2431 |
| | MAE | 0.0241 | 0.0394 | 0.0740 | 0.1114 | 0.1471 | 0.0380 | 0.0555 | 0.0868 | 0.1394 | 0.1493 |
| | $R^2$ | 0.9627 | 0.9087 | 0.6842 | 0.2878 | -0.223 | 0.9075 | 0.7899 | 0.4561 | 0.1544 | 0.0105 |
| Pre-ATGCN | RMSE | 0.0300 | 0.0488 | 0.0900 | 0.1327 | 0.1722 | 0.0796 | 0.1168 | 0.1760 | 0.2201 | 0.2473 |
| | MAE | 0.0234 | 0.0396 | 0.0754 | 0.1088 | 0.1418 | 0.0413 | 0.0601 | 0.0936 | **0.1152** | 0.1426 |
| | $R^2$ | 0.9655 | 0.9084 | 0.6889 | 0.3232 | -0.139 | 0.8939 | 0.7717 | 0.4811 | 0.1889 | -0.024 |
| ATGCN | RMSE | **0.0293** | **0.0461** | **0.0809** | **0.1253** | **0.1575** | **0.0600** | **0.0911** | **0.1511** | **0.2143** | **0.2242** |
| | MAE | **0.0226** | **0.0374** | **0.0671** | **0.1035** | **0.1357** | 0.0332 | **0.0461** | 0.0840 | 0.1273 | **0.1342** |
| | $R^2$ | **0.9670** | **0.9184** | **0.7487** | **0.3971** | **0.0469** | **0.9397** | **0.8611** | **0.6176** | **0.2308** | **0.1585** |

*Notes:* The bold font represents the best prediction under a horizon, and the underline font represents the sub-optimal prediction under a horizon.

improvement compared with GAR on RMSE when h=1 and 8), followed by the experience-driven ATGCN models and autoregressive models, while the deep sequence learning models have the worst prediction performance.

Among the three autoregressive models, the GAR model performs best, especially in the short-term prediction (h=1, 2, 4). When h= 6 or 8, R2 of AR (-0.408 and - 1.19) and VAR (-2.36 and -3.201) show that the predictions of them are significantly lagged. Although the VAR model integrates information from other age groups, its accuracy is the worst. This may be because the strong consistency of the time series is beneficial to the GAR model, while the VAR model only linearly combines the features of each age group, and its modeling ability is insufficient.

As for the three deep sequence learning models, CNNRNN-Res has the highest accuracy, but it is still much lower than the GAR model and the ATGCN model. The negative $R^2$ indicates that the predictions of RNN and CNNRNN models have poor correlation with the test dataset in both short-term and long- term prediction. This may be because the COVID-19 data set has no obvious periodicity, and the distribution of the training set and the test set are quite different, which is not conducive to the training of the deep learning model with RNN as the core module.

The ATGCN model and the experience-driven ATGCN models perform similarly in the prediction of the COVID-19 data set. However, since the two static adjacency matrices based on expert experiences cannot be adjusted adaptively, the prediction performance of Ones-ATGCN and Pre-ATGCN is not as robust as the ATGCN model. Although the LSTM modules in these three models learn the periodicity of time series, similar to the function of the RNN module discussed above, the performances of the three ATGCN-based models are not greatly affected, because they have linear transformation layers directly connected from the input to the output module.

### 3.2. Predictions of the influenza

As shown in Table 1, in the prediction of the influenza dataset which is periodic, the ATGCN model still achieves the best accuracy most of the time. (10% and 5% improvement compared with CNNRNN-Res on RMSE when h=1 and 8). The prediction results of the ATGCN model are well correlated with the test set, and its $R^2$ is much higher than the suboptimal result. When h=1, 4, or 6, the MAEs (0.0332, 0.084, 0.1273) of ATGCN are slightly higher than the corresponding deep sequence learning models (0.0313 of CNNRNN-Res, 0.0825 of RNN, and 0.1244 of CNNRNN).

Different from the results of the COVID-19 dataset, the performance of the deep sequence learning models this time is more accurate than that of the autoregressive models. The reason for this performance changes is that the CNNRNN-Res model and the ATGCN model are more suitable for forecasting seasonal infectious diseases. However, the GCN module in the ATGCN models the multiple age groups as a directed graph, which can learn the contact propagation between any two age groups; while the CNN module in the CNNRNN-Res can only learn the contact relationship between the adjacent age groups. Obviously, the ATGCN model is more in line with the actual transmission mode of infectious diseases.

### 3.3. Predictions on multiple age groups

Because the models predict the number of confirmed cases for multiple age groups simultaneously, in the following, whether their prediction effect for each age group is consistent is compared. We select the GAR, CNNRNN-Res, and ATGCN model as representatives for analysis, because these three models perform best in their respective categories.

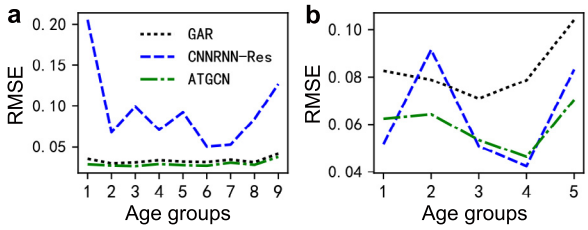


**Fig. 3. The prediction error (RMSE) of GAR, CNNRNN-Res and ATGCN for each age group when horizon is 1**. (a) RMSE for the COVID-19 dataset; (b) RMSE for the influenza dataset.

For the COVID-19 dataset, as shown in Fig. 3a, the RMSE of the GAR model and the ATGCN model of the nine age groups remains within the range of 0.033±0.003 and 0.029±0.003, while the RMSE of the CNNRNN-Res model is much higher and unstable (0.094±0.045). Fig. 4 shows the prediction results of the three models on the test set in detail. The ATGCN model and GAR model closely fit the real curves. However, the CNNRNN-Res model has lagged predictions on the onset time and lower predicted values of the outbreak size for age groups of 0–9 and 20–29, or larger predicted outbreak size for age groups of 30–39, 40–49, and above 80.

As for their performances on the influenza dataset (Fig. 3b), the ATGCN model has the most robust performances on the five age groups, and its RMSE is in the range of 0.059±0.008. Consistent with the results in Table 1, the effect of the GAR model on influenza decreases, in the range of 0.083±0.011 (RMSE). Though the prediction performance of CNNRNN-Res improves, with RMSE in the range of 0.064±0.020, its prediction accuracy for each age group still fluctuates the most. In Fig. 5, it can be found that ATGCN has the most accurate predictions on the onset time, while the lagged predictions of the GAR model are the most obvious, especially in age groups of 0–5 and 25–59. As for the prediction of peak size, the ATGCN model achieves good accuracy for each age group, except for the age group of 15–24.

The above results show that the prediction accuracy of CNNRNN-Res for different age groups fluctuates greatly. It cannot effectively integrate the interactive information of multiple age groups maybe for the reason that its CNN module only extracts the time series features of adjacent age groups. The RMSE of the GAR model on the two datasets fluctuates little because it uses the same parameter modeling for each age group, and the time series trends of each age group are similar. The ATGCN model performs most robustly in age groups because it uses the GCN module to effectively integrate the interactive information of different age groups.

### *3.4. Contact patterns of multiple age groups*

Graph adjacency matrix A is produced by the adaptive graph learning module of the ATGCN model. A represents the contact patterns of multiple age groups, which are related not only to the frequency of contacts among age groups, but also to the immunity of age groups to the virus. So, the cell of $\boldsymbol{A}$, $A_{ij}$, represents the risk of transmission from the $j$th age group to the $i$th age group.

Fig. 6 shows the graph adjacency matrices of the two datasets with h=1 or 8. For the influenza dataset Fig. 6a, b, both matrices show consistent contact patterns. In the five age groups, adolescents aged 0–15 years are more susceptible to infections from people aged 15 and older. On the contrary, children aged 0–15 years present a low risk of transmission to the higher age group, and their impact is obvious only to the adjacent age group (i.e., 0.99 and 1 in the first row of Fig. 6a). This may be related to the low immunity of the younger age group to influenza virus and the relatively strong immunity of the older age group.

Since there are 9 age groups in the COVID- 19 dataset, it is not easy to observe obvious patterns in Fig. 6a, b. So, we sum the rows or columns of the matrix: the sum of the rows represents the total risk of infection of one age group received from the others; the sum of the columns represents the total risks of transmission of one age group to other age groups. When h=1, there is no obvious difference between the influence of one age group on other age groups, which are evenly distributed between 3 and 4. This is due to the ATGCN's short-term forecast for the COVID19 dataset being closer to the autocorrelation forecast. When h=8, the influence of incubation period is more obvious among different age groups: older people over 50 with low immunity are more likely to be infected (the sums of the first four rows are 4, 4.7, 4.2, and 5 in Fig. 6d), while young and middle-aged people under 40 are more likely to infect others (the sum of the four three columns are 3.08, 5, 4, and 4 in Fig. 6d), because young people have more active social contacts.

## 4. Discussion

With COVID-19 and influenza as the case study objects, the ATGCN model is more accurate than the autoregressive models and the deep sequence learning models in both short-term and long-term prediction tasks. ATGCN accurately predicts the onset time and outbreak size of both datasets, and its accuracies fluctuate least in different age groups. Compared with the Ones-ATGCN model or Pre-ATGCN model which are based on expert experiences, the ATGCN model with the adaptive graph learning module is more adaptable to scene changes and more robust in performance.

Based on the above research results, two issues need to be further discussed:

First, a suitable model should be selected when predicting infectious diseases with various transmission characteristics. In this study, COVID-19 is an emerging infectious disease. Though the pathogenic virus SARS-CoV-2 has the behavior of a seasonal respiratory virus [38], the population is generally susceptible to the virus [39], so at this stage, the temporal characteristics of COVID-19 are more affected by the government intervention measures [40]. While seasonal influenza occurs in spring and winter every year, and the number of onset cases has obvious periodicity. The experimental results show that the autoregressive models are more suitable for predicting the COVID-19 dataset, while the deep sequence learning models have better performance on the influenza dataset. Because the ATGCN combines the characteristics of two sequence modules (the LSTM module and a linear transformation layer), its performance on the two datasets is both optimal. However, in the existing papers studying the prediction of infectious diseases, few of them have discussed the difference in the temporal characteristics of time series caused by the various rules of disease transmission. Accordingly, in practical application, more attention should be paid in selecting the appropriate forecasting models according to the temporal characteristics of infectious disease. Or a fusion model which can dynamically combine the advantages and adjust the weight of basic models [41,42] should be developed.

Secondly, domain knowledge of the infectious diseases should be further applied to the design of deep learning models and the feature selection, for the fusion of domain knowledge can improve the accuracy and interpretability of a model. As observed from this study, since the contact pattern of multiple age groups is an important factor for prediction, we chose the GCN module and the adaptive graph learning module as the core functional modules to learn the interaction relationship between multiple age groups. The experimental results also prove our choice that the graph neural network is more suitable for the learning contact pattern than the recurrent neural network. And with the expansion of the available infectious disease data, such as age, gender, and population movement trajectory, more deep learning models with epidemiological characteristics can be designed, based on the immune heterogeneity, age distribution, family residence pattern and other features of the susceptible population.

The limitations of the ATGCN model still exist: first, the adaptive graph learning module only updates the adjacency matrix $\boldsymbol{A}$ within the

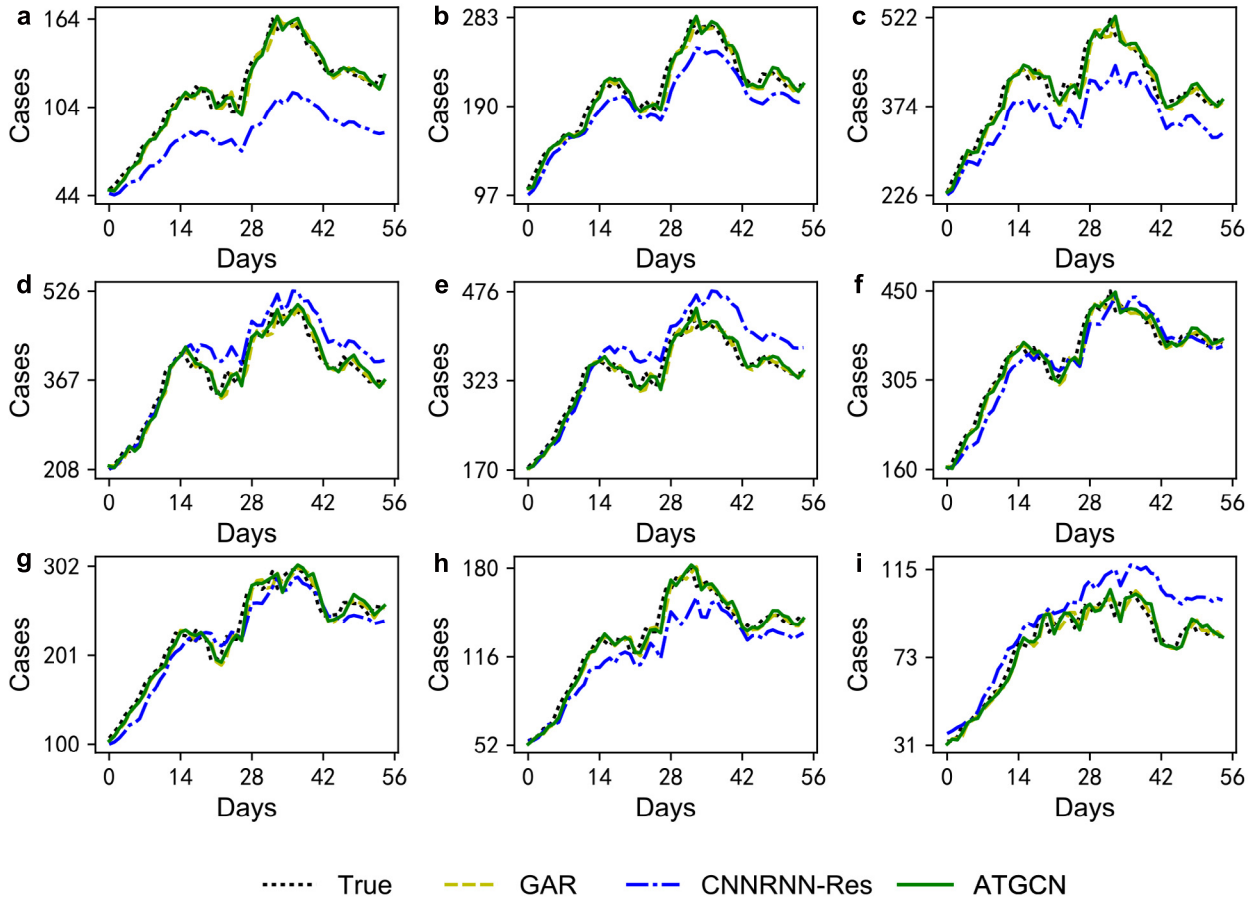


**Fig. 4. Prediction results of GAR model, CNNRNN-Res model and ATGCN model on the COVID-19 dataset (horizon=1 day).** (a) Age group of 0–9; (b) age group of 10–19; (c) age group of 20–29; (d) age group of 30–39; (e) age group of 40–49; (f) age group of 50–59; (g) age group of 60–69; (h) age group of 70–79; (i) age group of above 80.

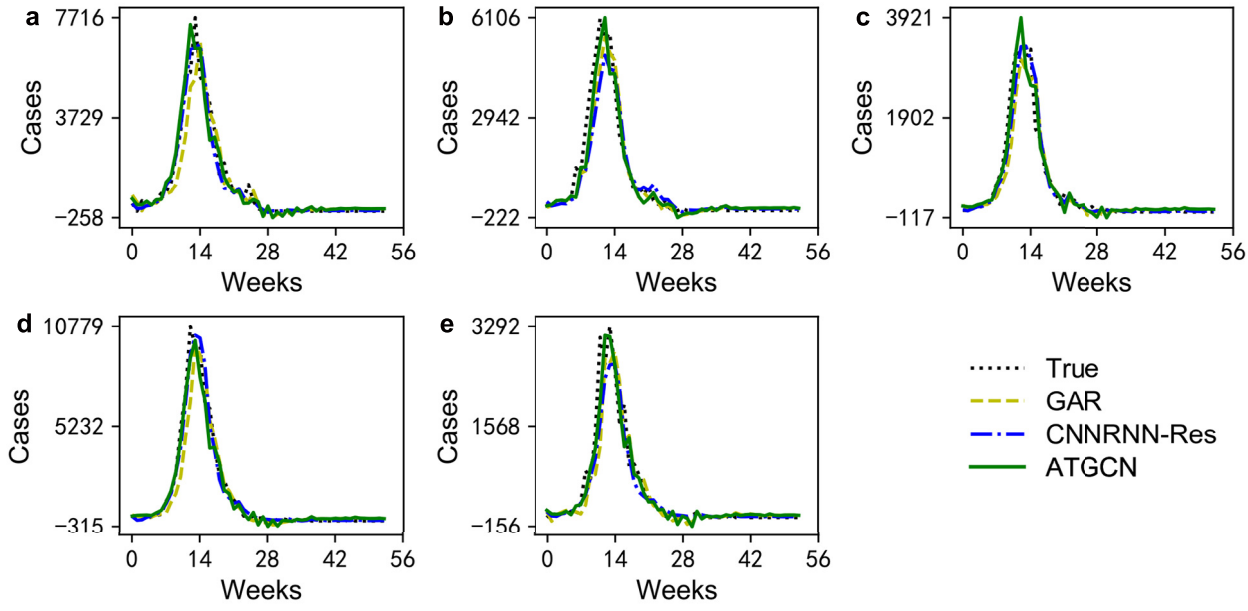


**Fig. 5. Prediction results of GAR model, CNNRNN-Res model and ATGCN model on the influenza dataset (horizon=1 week).** (a) Age group of 0–5; (b) age group of 5–14; (c) age group of 15–24; (d) age group of 25–59; (e) age group of above 60.

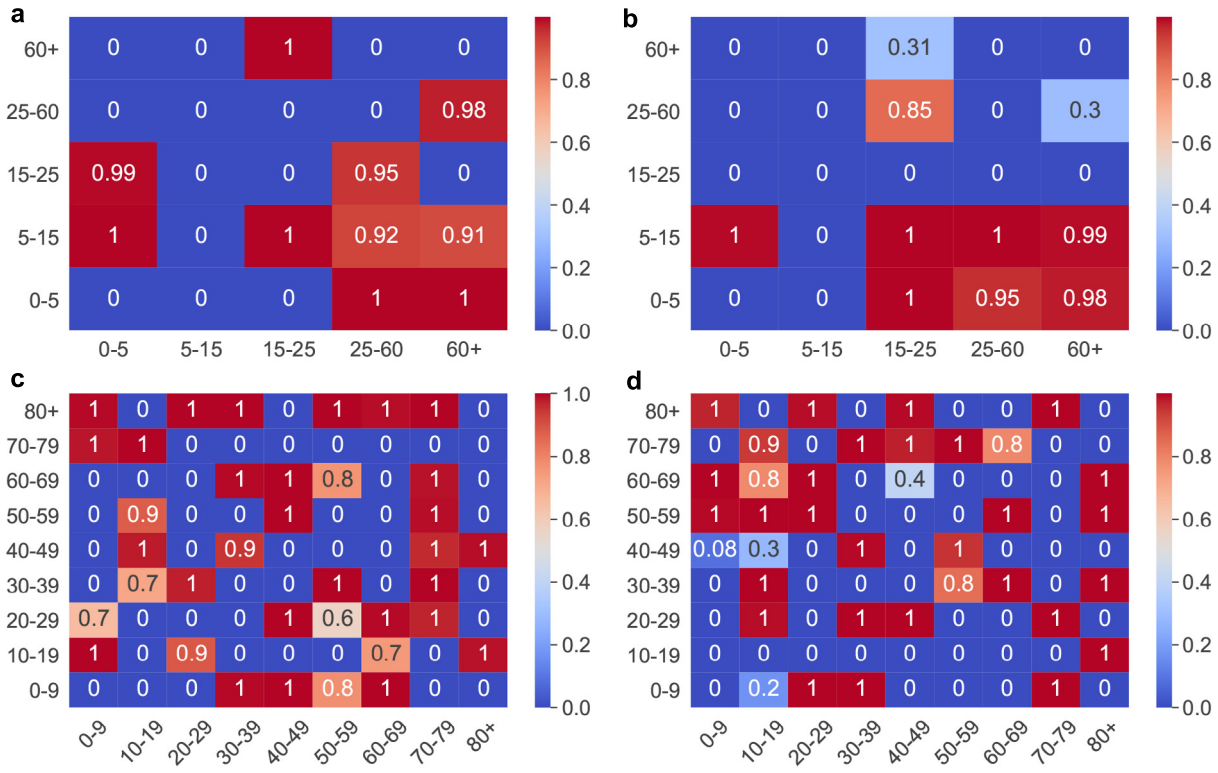


**Fig. 6. Graph adjacency matrices of multiple age groups from ATGCN**. (a) the graph adjacency matrix of the influenza dataset with horizon=1; (b) the graph adjacency matrix of the influenza dataset with horizon=8; (c) the graph adjacency matrix of the COVID-19 dataset with horizon=1; (d) the graph adjacency matrix of the COVID-19 dataset with horizon=8.

time range of the training dataset, and when predicting the test dataset, the parameters of the ATGCN model are fixed and no longer updated. Therefore, the adjacency matrix $\boldsymbol{A}$ used for test dataset prediction can only represent the historical multiple age group contact patterns, which will lead to a decrease in the accuracy of the ATGCN model. This is also the limitation of the basic assumption of our study: the contact pattern of the crowd is actually dynamic, rather than stable. Secondly, the ATGCN model cannot give the uncertainty of the prediction results and this is a common problem with deep learning models. However, uncertainty is important in public health management [43,44], because decision makers need to estimate the credibility of the results based on the upper and lower bounds, so as to plan the redundancy of public health resources. In addition, the limitation of data is a common problem in epidemic prediction. Neither the public available COVID- 19 dataset nor the influenza dataset obtained through public health surveillance can truly reflect the scales of outbreak and the age structure of cases. The adjustment of diagnostic criteria, inadequate diagnosis capacity and asymptomatic infections can influence the patients' healthcare-seeking behavior, leading to misdiagnosed or left untreated, especially in the early stage of an emerging epidemic, when the symptoms of the disease are poorly understood.

## 5. Conclusion

In this paper, we introduce a novel framework for time series prediction of multiple age groups, integrating graph convolution and adaptive graph learning module to combine the contact pattern of people into epidemic prediction. It addresses the multivariate time series forecasting in a graph-based deep learning approach and achieves robust prediction on the confirmed cases of multiple age groups. Experiments show that our model outperforms the state-of-the-art methods on two epidemic datasets, indicating its great potentials in exploring the implicit interactions of abstract multivariate variables. And our method demonstrates that integrating the domain knowledge of infectious diseases into the design of the deep learning framework can effectively improve the accuracy and applicability of the model.

## Ethics statement

All methods were carried out in accordance with the principles of the Declaration of Helsinki. The experimental protocols were approved by the ethical committee of Beijing Center for Diseases Prevention and Control. And informed consents were obtained from all participants and their legal guardians. All records were anonymized and no individual information can be identified.

## Declaration of Competing Interest

The authors declare that they have no conflicts of interest in this work.

## Acknowledgments

This work was supported in part by grants from the National Natural Science Foundation of China (Grants No. 72025404 and 71621002), Beijing Natural Science Foundation (Grant No. L192012), and Beijing Nova Program (Grant No. Z201100006820085).

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.fmre.2021.07.007.

## References

[1] K. Thursky, S.P. Cordova, D. Smith, et al. , Working towards a simple case definition for influenza surveillance, J. Clin. Virol. 27 (2) (2003) 170–179.
[2] A.D. Iuliano, K.M. Roguski, H.H. Chang, et al. , Estimates of global seasonal influenza-associated respiratory mortality: a modelling study, Lancet 391 (10127) (2018) 1285–1300.

[3] T.P. Velavan, CG. Meyer , The COVID-19 epidemic, Trop. Med. Int. Health 25 (3) (2020) 278.
[4] S.P. Adhikari, S. Meng, Y.J. Wu, et al. , Epidemiology, causes, clinical manifestation and diagnosis, prevention and control of coronavirus disease (COVID-19) during the early outbreak period: a scoping review, Infect. Dis. Poverty 9 (1) (2020) 1–12.
[5] S. Felsenstein, J.A. Herbert, P.S. McNamara, et al. , COVID-19: immunology and treatment options, Clin. Immunol. (2020) 108448.
[6] M. Paul, L. Held, A.M. Toschke , Multivariate modelling of infectious disease surveillance data, Stat. Med. 27 (29) (2008) 6250–6267.
[7] Q. Li, M. Li, L. Lv, et al. , A new prediction model of infectious diseases with vaccination strategies based on evolutionary game theory, Chaos Solitons Fractals 104 (2017) 51–60.
[8] Y. Ji, Z. Ma, M.P. Peppelenbosch, et al. , Potential association between COVID-19 mortality and health-care resource availability, Lancet Glob. Health 8 (4) (2020) e480.
[9] G. Zhang, X. Liu , Prediction and control of COVID-19 spreading based on a hybrid intelligent model, PloS One 16 (2) (2021) e0246360.
[10] K.T. Thai, H. Nishiura, P.L. Hoang, N.T.T. Tran, G.T. Phan, H.Q. Le, et al. , Age-specificity of clinical dengue during primary and secondary infections, PLoS Negl. Trop. Dis. 5 (6) (2011) e1180.
[11] M.G. Guzmán, G. Kouri, J. Bravo, et al. , Effect of age on outcome of secondary dengue 2 infections, Int. J. Infect. Dis. 6 (2) (2002) 118–124.
[12] K.M. Gostic, R. Bridge, S. Brady, et al. , Childhood immune imprinting to influenza A shapes birth year-specific risk during seasonal H1N1 and H3N2 epidemics, PLoS Pathog. 15 (12) (2019) e1008109.
[13] J. Zhao, F. Jiang, L. Zhong, et al. , Age patterns and transmission characteristics of hand, foot and mouth disease in China, BMC Infect. Dis. 16 (1) (2016) 1–12.
[14] A. Esteve, I. Permanyer, D. Boertien, et al. , National age and coresidence patterns shape COVID-19 vulnerability, Proc. Natl. Acad. Sci. U. S. A. 117 (28) (2020) 16118–16120.
[15] S. Nassar, K.P. Schwarz, A. Noureldin , Modeling inertial sensor errors using autoregressive (AR) models, J. Inst. Navig. 51 (4) (2004) 259–268.
[16] A. Hatemi-j , A new method to choose optimal lag order in stable and unstable VAR models, Appl. Econ. Lett. 10 (3) (2003) 135–137.
[17] J. Xiao, Y. Li, L. Xie, et al. , A hybrid model based on selective ensemble for energy consumption forecasting in China, Energy 159 (2018) 534–546.
[18] L. Liu, M. Han, Y. Zhou, et al. , LSTM recurrent neural networks for influenza trends prediction, in: F. Zhang, Z. Cai, P. Skums, S. Zhang (Eds.), Bioinformatics Research and Applications. ISBRA 2018. Lecture Notes in Computer Science, 10847, Springer, Cham, 2018, pp. 259–264.
[19] Said A.B., Erradi A., Aly H., et al. Predicting COVID-19 cases using bidirectional LSTM on multivariate time series. arXiv preprint arXiv:200912325. 2020. https://link.springer.com/article/10.1007/s11356-021-14286-7
[20] , Deep learning for epidemiological predictions, in: Y. Wu, Y. Yang, H. Nishiura, M. Saitoh (Eds.), Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, 2018.
[21] S.K. Paul, S. Jana, P. Bhaumik , A multivariate spatiotemporal model of COVID-19 epidemic using ensemble of ConvLSTM networks, J. Inst. Eng. (India) Ser. B (2020) 1–6.
[22] , Gated residual recurrent graph neural networks for traffic prediction, in: C. Chen, K. Li, S.G. Teo, X. Zou, K. Wang, J. Wang, Z. Zeng (Eds.), Proceedings of the AAAI Conference on Artificial Intelligence, 2019.
[23] , Attention based spatial-temporal graph convolutional networks for traffic flow forecasting, in: S. Guo, Y. Lin, N. Feng, C. Song, H. Wan (Eds.), Proceedings of the AAAI Conference on Artificial Intelligence, 2019.
[24] Yu B., Yin H., Zhu Z. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. arXiv preprint arXiv:170904875. 2017.
[25] Z. Li, X. Luo, B. Wang, et al. , A Study on graph-structured recurrent neural networks and sparsification with application to epidemic forecasting, in: H. Le Thi, H. Le, T. Pham Dinh (Eds.), Optimization of Complex Systems: Theory, Models, Algorithms and Applications. WCGO 2019. Advances in Intelligent Systems and Computing, 991, Springer, Cham, 2020, pp. 730–739. 2020.
[26] J. Gao, R. Sharma, C. Qian, et al. , STAN: spatio-temporal attention network for pandemic prediction using real-world evidence, J. Am. Med. Inform. Assoc. 28 (4) (2021) 733–743.
[27] V. La Gatta, V. Moscato, M. Postiglione, et al. , An epidemiological neural network exploiting dynamic graph structured data applied to the COVID-19 outbreak, IEEE Trans. Big Data (2020).
[28] Kapoor A., Ben X., Liu L., et al. Examining covid-19 forecasting using spatio-temporal graph neural networks. arXiv preprint arXiv:200703113. 2020.
[29] Z. Wu, S. Pan, G. Long, et al. , Connecting the dots: multivariate time series forecasting with graph neural networks, in: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2020.
[30] G.J. Griffith, T.T. Morris, M.J. Tudball, et al. , Collider bias undermines our understanding of COVID-19 disease risk and severity, Nat. Commun. 11 (1) (2020) 1–12.
[31] Y. Zhang, Z. Cao, V. Costantino, et al. , Influenza illness averted by influenza vaccination among school year children in Beijing, 2013–2016, Influenza Other Respir. Viruses 12 (6) (2018) 687–694.
[32] K. Prem, A.R. Cook, M. Jit , Projecting social contact matrices in 152 countries using contact surveys and demographic data, PLoS Comput. Biol. 13 (9) (2017) e1005697.
[33] J. Zhang, M. Litvinova, Y. Liang, et al. , Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China, Science 368 (6498) (2020) 1481–1486.
[34] T. Cheng, J. Haworth, J. Wang , Spatio-temporal autocorrelation of road network data, J. Geogr. Syst. 14 (4) (2012) 389–413.
[35] M. O'Driscoll, G.R. Dos Santos, L. Wang, et al. , Age-specific mortality and immunity patterns of SARS-CoV-2, Nature 590 (7844) (2021) 140–145.
[36] S. Albawi, T.A. Mohammed, S. Al-Zawi (Eds.), Understanding of a convolutional neural network. Proceedings of the International Conference on Engineering and Technology (ICET), IEEE, 2017.
[37] B. Yue, J. Fu, J. Liang , Residual recurrent neural networks for learning sequential representations, Information 9 (3) (2018) 56.
[38] M.M. Sajadi, P. Habibzadeh, A. Vintzileos, et al. , Temperature, humidity, and latitude analysis to estimate potential spread and seasonality of coronavirus disease 2019 (COVID-19), JAMA Netw. Open 3 (6) (2020) e2011834 -e.
[39] X. Huang, F. Wei, L. Hu, et al. , Epidemiology and clinical characteristics of COVID-19, Arch. Iran. Med. 23 (4) (2020) 268–271.
[40] C.J. Carlson, A.C. Gomez, S. Bansal, et al. , Misconceptions about weather and seasonality must not misguide COVID-19 response, Nat. Commun. 11 (1) (2020) 1–4.
[41] K. Su, L. Xu, G. Li, et al. , Forecasting influenza activity using self-adaptive AI model and multi-source data in Chongqing, China, EBioMedicine. 47 (2019) 284–292.
[42] M. Soliman, V. Lyubchich, Y.R. Gel , Complementing the power of deep learning with statistical model fusion: probabilistic forecasting of influenza in Dallas County, Texas, USA, Epidemics 28 (2019) 100345.
[43] H. Rutter, M. Wolpert, T. Greenhalgh , Managing uncertainty in the covid-19 era, BMJ 370 (2020).
[44] Ghoshal B., Tucker A. Estimating uncertainty and interpretability in deep learning for coronavirus (COVID-19) detection. arXiv preprint arXiv:200310769. 2020.
[45] R.M. Anderson, R.M. May, Infectious Diseases of Humans: Dynamics and Control, Oxford University Press, 1992.
[46] S.F. Dowell, C.G. Whitney, C. Wright, et al., Seasonal patterns of invasive pneumococcal disease, Emerg. Infect. Dis. 9 (5) (2003) 574.
[47] F. Balabdaoui, D. Mohr, Age-stratified discrete compartment model of the COVID-19 epidemic with application to Switzerland, Sci. Rep. 10 (1) (2020) 1–12.
[48] V.P. Bongolan, J.M.A. Minoza, R. de Castro, et al., Age-stratified infection probabilities combined with a quarantine-modified model for COVID-19 needs assessments: model development study, J. Med. Internet Res. 23 (5) (2021) e19544.
[49] M.S. Islama, K.A. Kabirb, J.I. Irac, et al., Isolation effect on age-stratified compartmental model of the COVID-19, Commun. Nonlinear Anal. 8 (1) (2021) 1–12.
[50] N.G. Davies, P. Klepac, Y. Liu, et al., Age-dependent effects in the transmission and control of COVID-19 epidemics, Nat. Med. 26 (8) (2020) 1205–1211.

**Yuejiao Wang** is currently working toward the master's degree in the Institute of Automation, Chinese Academy of Sciences, China. She received the B.E. degree in electronic information engineering from the Harbin Institute of Technology at Weihai, China. Her research interests include time series prediction, social computing and public health management.

**Zhidong Cao** received the Ph.D. degree in geographic information science from the Institute of Geographic Sciences and Nature Resources Research, Chinese Academy of Sciences, Beijing, China, in 2008. He is currently a Professor at the Institute of Automation, Chinese Academy of Sciences. His current research interests include social computing and big data, public health emergency management, and spatial statistical analysis.