Fundamental Research 4 (2024) 527–539

Contents lists available at ScienceDirect

Fundamental Research

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

Article

# High-resolution short-term prediction of the COVID-19 epidemic based on spatial-temporal model modified by historical meteorological data

Bin Chen[a,b,*], Ruming Chen[a,b], Lin Zhao[a], Yuxiang Ren[a], Li Zhang[a], Yingjie Zhao[a], Xinbo Lian[a], Wei Yan[a], Shuoyuan Gao[a]

[a] College of Atmospheric Sciences, Lanzhou University, Lanzhou 730000, China
[b] Collaborative Innovation Center of Western Ecological Security, Lanzhou University, Lanzhou 730000, China

ARTICLE INFO

*Article history:*
Received 17 January 2023
Received in revised form 7 February 2024
Accepted 19 February 2024
Available online 5 March 2024

*Keywords:*
COVID-19
Prediction
ConvLSTM
Refined prediction
Meteorological factors
Spatial-temporal analysis

ABSTRACT

In the global challenge of Coronavirus disease 2019 (COVID-19) pandemic, accurate prediction of daily new cases is crucial for epidemic prevention and socioeconomic planning. In contrast to traditional local, one-dimensional time-series data-based infection models, the study introduces an innovative approach by formulating the short-term prediction problem of new cases in a region as multidimensional, gridded time series for both input and prediction targets. A spatial-temporal depth prediction model for COVID-19 (ConvLSTM) is presented, and further ConvLSTM by integrating historical meteorological factors (Meteor-ConvLSTM) is refined, considering the influence of meteorological factors on the propagation of COVID-19. The correlation between 10 meteorological factors and the dynamic progression of COVID-19 was evaluated, employing spatial analysis techniques (spatial autocorrelation analysis, trend surface analysis, etc.) to describe the spatial and temporal characteristics of the epidemic. Leveraging the original ConvLSTM, an artificial neural network layer is introduced to learn how meteorological factors impact the infection spread, providing a 5-day forecast at a $0.01^\circ \times 0.01^\circ$ pixel resolution. Simulation results using real dataset from the 3.15 outbreak in Shanghai demonstrate the efficacy of Meteor-ConvLSTM, with reduced RMSE of 0.110 and increased $R^2$ of 0.125 (original ConvLSTM: RMSE = 0.702, $R^2$ = 0.567; Meteor-ConvLSTM: RMSE = 0.592, $R^2$ = 0.692), showcasing its utility for investigating the epidemiological characteristics, transmission dynamics, and epidemic development.

## 1. Introduction

COVID-19 has triggered a major global crisis affecting life, health security, economic and social order crises. The healthcare system is facing enormous challenges. The severe impact on human health is evident through various clinical symptoms, primarily fever, dyspnea, and weakness, with severe cases leading to fatalities [1,2]. There are a total of 62 diseases that are significantly associated with COVID-19 [3]. To limit the spread of the virus, countries implement countermeasures and formulate prevention strategies. Under the epidemic prevention policy dominated by containment, the global supply chain has been adversely affected [4], leading to changes in trade structures [5], which directly resulted in a decrease in public social and economic activities, and an increase in the unemployment rate. The outbreak has also seriously affected education and food security in low-income countries [6,7]. Some studies have quantitatively analyzed the impact of national control measures in the early stage of the epidemic. The results indicate that the national emergency responses can effectively slow the growth and limit the scale of epidemics [8]. The construction of appropriate models to predict the number of daily new cases can be used to study and predict the risk of the outbreak [9,10]. By leveraging the experience and practices gained during the pandemic, these models provide scientific support for human society to respond to major emerging diseases.

Researchers have proposed various prediction methods for the number of cases, deaths, and incidence [11–15]. Studies have integrated statistical models with dynamic Susceptible-Infected-Recovered (SIR) models to explore the periodicity and variability of epidemic, and investigate the mechanisms of epidemic spread [16]. In another study, an Susceptible Exposed-Infected-Recovered (SEIR) dynamic prediction model with seasonal signals was constructed to quantify the impact of seasonal signals and assess the effect of seasonal amplitude on the total number of infections, emphasizing the need to consider seasonal factors when formulating prevention policies. In some studies, models for the spread of epidemics triggered by crowd gathering events to simulate the impact of crowd gathering events on the overall epidemics dynamics [17]. The extreme gradient boosting model, developed based on the spatial-temporal characteristics of interpersonal interactions, can predict county-level COVID-19 new cases weekly [18,19]. An Artificial

* Corresponding author.
*E-mail address:* chenbin@lzu.edu.cn (B. Chen).

https://doi.org/10.1016/j.fmre.2024.02.006

Neural Network (ANN) model can be used to analyze the relationship between new and total confirmed cases, yielding more accurate case prediction results [20]. Some study used a variety of machine learning models and historical COVID-19 data to predict future disease transmission in near real time [21]. ARIMA (Autoregressive Moving Average) and LSTM (Long Short-Term Memory) models based on time series data have also been applied to predict the number of COVID-19 cases [22,23].

Effective outbreak control depends on the timely availability of crucial information, and the utilization of variables to establish accurate links to COVID-19 facilitates the level of prediction reliability [24,25]. Climatic disasters can lead to an increased incidence and the transmission of more than 100 vector-borne diseases [26]. The climatic environment prior to COVID-19 provided favorable conditions for the emergence of an epidemic. SARS-CoV-2 belongs to the beta genus coronavirus, with a diameter ranging from 60 nm to 140 nm, existing either single or in aggregates. It forms aerosols by adsorbing to droplets, particles, or other substances in the environment. An aggregated SARS-CoV-2 aerosol exhibit strong and sustained transmission capacity as a fine particles in the air, capable of both short-range and long-range transport. The results suggest that higher viral loads and viral infectivity increase the risk of aerosol transmission of the SARS-CoV-2 Omicron variant. Studies have also confirmed the important role of sewer transmission due to fecal aerosols in community outbreaks through case-by-case analysis [27]. Since polymeric aerosols undergo a decay process differing from the general composition of pollution gases [28], research simulated the diffusion transport process of solid-liquid droplets in an outdoor environment by CFD to demonstrate that droplet transmission of virus-carrying droplets can lead to the outdoor spread of respiratory infections [29]. Under the influence of boundary layer turbulence characteristics and meteorological processes, their concentrations of droplets exhibit complex spatial and temporal distributions. Unevenly distributed aerosols are inhaled into the lungs by susceptible individuals, leading to lung infection with a certain probability, thus affecting the spatial and temporal distributions of cases. Therefore, it is necessary to consider the influence of meteorological conditions on the formation and spread of SARS-CoV-2 aerosols when constructing prediction models.

However, to date, few studies have constructed high-resolution spatial and temporal prediction models of epidemics, considering the spatial and temporal characteristics of meteorological factors and historical case numbers. Studies have used data on population density, population age, temperature, humidity, and economic levels to establish links with COVID-19 through LSTM models for short-term and long-term predictions of infection cases. Since air quality is the most important factor affecting lung and respiratory diseases, ANN models between air pollutant concentrations and infection cases have also been developed. A common limitation of simulation processes is model's dependence on a set of parametric assumptions reflecting real-world complexity (e.g., individual behaviors and conditions, public policies, and institutions). It is crucial to maximize the explanatory and predictive ability of the model by incorporating meteorological factors that influence virus spread and mechanisms in a holistic and detailed manner. This confirms whether high-precision spatial-temporal prediction of the epidemic can be achieved and whether meteorological factors can enhance spatial-temporal depth models.

In addition to the meteorological factors used in this study, various elements such as population migration, contact patterns, and virus variations play crucial roles in the dynamic process of the COVID-19 pandemic. To specifically explore the individual impact of meteorological factors on the epidemic, research collected the Migration Scale Index (MSI) for 30 cities, controlling for population migration [30]. They employed nonlinear regression analysis, controlling for a 3-day moving average of MSI, to independently assess the influence of meteorological factors on the number of COVID-19 cases. Population migration and contact patterns directly affect the virus transmission pathways and speed. Population migration leads to population movement, exacerbating the spread of the epidemic, and areas with high population flow, such as transportation hubs, city centers, and markets, may be more prone to virus transmission. The aerosol transmission of COVID-19 is of significant concern, emphasizing the need for early monitoring in densely populated areas to prevent outbreaks [31]. Different social and contact networks impact the virus transmission pathways within communities, thus influencing the spread rate of cluster outbreaks. A crucial study, based on statistical methods and an improved epidemiological model, developed the Global Pandemic Early Warning System (GPEP-2), successfully predicting the global and regional spread of COVID-19 by introducing parameterized schemes for natural and anthropogenic factors [32]. Additionally, policies and intervention measures also influence virus transmission patterns. Research employed a dose-response relationship to assess the relationship between infection risk and exposure dose [33]. In future studies on cluster outbreaks, we will comprehensively consider influencing factors to conduct a comprehensive predictive analysis of epidemic development.

In summary, this study uses the 3.15 outbreak in Shanghai, China, as an example analysis to describe the spatial-temporal profile of the outbreak. Historical meteorological factors are taken into account in the depth model for case number prediction, establishing a suitable and robust spatial-temporal depth model. This model provides accurate time-forward predictions of the epidemic development, offering a solution to the COVID-19 case number prediction problem with high spatial resolution considering historical meteorological conditions.

## 2. Materials and methods

### 2.1. Research region

Shanghai (120°52′ W-122°12′ W, 30°40′ N-31°53′ N), with a land area of 6340 $km^2$, a watershed area of 121.85 $km^2$, a resident population of approximately 24,894,300, and a population density of 3927 persons per $km^2$, experienced a rapid outbreak of the SARS-CoV-2 Omicron subvariant, BA.2, in early March 2022. By May 31, the social surface was cleared, with a total of over 600,000 cases. During this period, the Shanghai Municipal Health Commission released high-quality data on cases, clear information on the residence of cases, and timely updates of new data. Based on the above reasons, Shanghai was chosen as the study target region to comprehensively assess and predict the spatial-temporal characteristics of the epidemic.

### 2.2. Definition of the infected

Unlike SARS, COVID-19 infected patients are also infectious during their incubation period, and the presence of many asymptomatic infected patients poses a great challenge for the implementation of epidemic prevention efforts. A confirmed infected person is defined as those who presents with clinical symptoms of a new and positive viral antigens on examination of respiratory epithelial cells. An asymptomatic infected person is defined as those who has no clinical symptoms but tests positive for nucleic acid. Asymptomatic infected patients become confirmed infected patients if they test positive for nucleic acid and developed symptoms (or have lesions on the lungs, as revealed by a CT scan). Both types of patients are infectious vectors, so in this study, asymptomatic and confirmed infected patients were collectively referred to as the infected patients, and their daily additions were analyzed and predicted in terms of spatial and temporal characteristics.

### 2.3. Data sources and collection

A combination of manual downloading and scripting was used to collect daily summary tables of positive infection data between March 6, 2022 and May 30, 2022, from the Shanghai Municipal Health Commission (https://wsjkw.sh.gov.cn). The data included the official date of positive reporting, the district where the positive cases were located,

**Table 1**
**Data description and selection of model hyperparameter.**

(a) Detailed explanation of meteorological data used in this study.

| Abbreviation | Meaning | Temporal resolution | Spatial resolution |
|---|---|---|---|
| 2m_dt | 2m dew point temperature | daily | 0.25° × 0.25° |
| 2m_t | 2m temperature | daily | 0.25° × 0.25° |
| blh | Boundary layer height | daily | 0.25° × 0.25° |
| sk_t | Skin temperature | daily | 0.25° × 0.25° |
| sp | Surface pressure | daily | 0.25° × 0.25° |
| u10 | 10m east-west wind field | daily | 0.25° × 0.25° |
| u100 | 100m east-west wind field | daily | 0.25° × 0.25° |
| v10 | 10m south-east wind field | daily | 0.25° × 0.25° |
| v100 | 100m south-east wind field | daily | 0.25° × 0.25° |

(b) Effects of network width on model performance.

| Model | Trainable Parameters | Cross Entropy |
|---|---|---|
| ConvLSTM (3 × 3) (32, 32, 32, 1) | 188897 | 0.0659 |
| ConvLSTM (3 × 3) (64, 64, 64, 1) | 746433 | 0.0652 |
| ConvLSTM (3 × 3) (128, 64, 64, 1) | 2129473 | 0.0689 |
| ConvLSTM (3 × 3) (128, 128, 128, 1) | 2967425 | 0.0658 |

(c) List of model hyperparameter.

| Hyperparameter | Value |
|---|---|
| Optimizer | Adam |
| Error function | Cross entropy |
| Learning rate | 0.005 |
| epoch | 20 |
| Batch size | 5 |
| Network convolution layers | 3 |
| Number of convolution kernels every layer | (5, 5) (3, 3) (1, 1) (3, 3, 3) |

detailed residence information (accurate to the street), the type of infection, etc. Meteorological data were obtained from the daily reanalysis of ERA5 for 2022. ERA5 is the fifth generation of the ECMWF atmospheric reanalysis of the global climate, assimilating a large amount of satellite information and conventional observations such as ground and upper air observations. Its accuracy and resolution are sufficiently high, with a spatial resolution of 0.25° × 0.25°

Epidemics exhibit distinct transmission characteristics in varied spatial and temporal contexts. The spread of COVID-19, characterized by spatial heterogeneity, is significantly influenced by meteorological factors [34]. Changes in meteorological factors (temperature, humidity, wind, air pressure, and diffusion conditions, etc.) are important indicators of changes in climatic, environmental, and ecological factors on which the COVID-19 virus depends. The temperature and humidity affect the availability of virus liquid carriers and the abnormal multiplication of pathogens. The boundary layer height determines the potential spatial extent of contaminant diffusion dilution. The spatial distribution of cases under fixed diffusion conditions is best reflected by the physical properties of the wind field, and specific flow field patterns affect the unidirectional accumulation of virus aerosols. The ambient air pressure and body temperature affect the oxygen supply capacity of the organism, serving as crucial meteorological triggers for respiratory diseases. Therefore, the 2 m temperature, 2 m dew point temperature, boundary layer height, surface pressure, skin temperature, u and v wind fields at 10 m and 100 m above ground Table 1a were selected from the ECMWF's fifth generation atmospheric reanalysis of global climate (ERA5) dataset as the meteorological features affecting epidemic development.

### 2.4. Methods

The spatial analysis technique was used to interpret the changes in the current Shanghai epidemic in an overview manner. The following sections describe the data preprocessing process, the methodology used in the study, and the rationale behind constructing the ConvLSTM model.

#### 2.4.1. Data preprocessing

The target model used in this study, the ConvLSTM model, possessed a five-dimensional tensor with the following input dimensions: the number of samples, sample timesteps, matrix width, matrix height, and the number of variable channels (samples, timesteps, rows, columns, channels). Because of its end-to-end structure, the output aligns with the five-dimensional tensor. To meet the input requirements of the model, the target dataset is interpolated and dimensionally transformed, including the latitude and longitude of the discrete points, and their corresponding case numbers.

To obtain the spatial distribution of cases at a specific resolution, assuming that more data can be obtained in a small area for the convolution operation, the third spline interpolation method was employed to downscale the sample data to 0.01° × 0.01°. Meteorological factors data were similarly resampled to 0.01° × 0.01°. Considering the impact of the convolution operation in the neural network on different magnitude datasets, the data were normalized to the maximum-minimum value Eq. 1:

$$S = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

where $S$ is the normalized result after the linear transformation of the original data, $X$ is the sample data, and $X_{\max}$ and $X_{\min}$ are the maximum and minimum values in the sample data, respectively. Finally, the normalized three-dimensional historical case count and meteorological factor matrix are processed into a five-dimensional tensor.

#### 2.4.2. Spatial data interpolation technology

The cubic spline function method employed in this study is a variable spline function composed of some cubic polynomials, creating a smooth curve through a series of discrete points. This approach can be used to overcome the cusp problem, improve the smoothness of the interpolation function, and ensure the interpolation accuracy compared to the Newton interpolation and Lagrange interpolation methods [35]. It can also be used to eliminate the problems of the nearest neighbor interpolation method [36] which include grayscale image discontinuity and

jaggedness. Compared to the parabolic interpolation method [37] does not require a priori knowledge, and compared to other segmental interpolation methods (Hermite interpolation, linear interpolation) [38], the disadvantage of discontinuity of higher order derivatives is overcome when using the cubic spline function method. In summary, the cubic spline function method was chosen to accomplish the task of discrete point interpolation in this study.

#### 2.4.3. Spatial autocorrelation evaluation using the Moran's I

Moran's *I*, a statistical measure of the spatial interaction and spread of geographic data, was often used to describe the overall distribution of the phenomenon, and to determine whether the research subject has aggregation properties in space. Moran's $I$ is calculated by Eq. 2:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \times \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(x_i - \bar{x}\right)\left(x_j - \bar{x}\right)}{\sum_{j=1}^{n}\left(x_i - \bar{x}\right)^2} \tag{2}$$

where n is the number of spatial objects and $w_{ij}$ is the spatial adjacency weight matrix of each spatial unit i and j ($j = 1, 2, \ldots, n$) in the study area. Apply the spatial autocorrelation algorithm to study the regional aggregation of infectious diseases, providing an effective tool for epidemic analysis.

### 2.5. Model

The deep learning algorithm for refined spatial-temporal prediction in this study is based on the ConvLSTM model [39]. The ConvLSTM network layer has a composite network structure whose key components are the convolutional structure distributed over the time axis, and the cellular information $C_{t-j}$ ($j = 1, 2, \ldots, q$) distributed across the memory.

While good predictions of temporal data can be made using the classical LSTM recurrent network, it is challenging to portray the spatial features of 3D spatial-temporal data. However, the Hadamard product in the LSTM network can be replaced with a convolution operation to learn the spatial information in the dataset. The convolution structure in the ConvLSTM network is utilized to extract the spatial distribution features of cases.

In the prediction of urban traffic accidents, ConvLSTM proves effective in capturing spatiotemporal features of traffic accidents within road networks [40]. In the field of marine ecosystem prediction, ConvLSTM demonstrates universality in predicting values, leading to higher prediction accuracy compared to traditional models [41]. In the domain of epidemic prediction, the improved SEIR model has been confirmed to accurately describe the multi-wave process, achieving high accuracy in predicting COVID-19 cases [42]. Models developed for predicting COVID-19 have also successfully simulated and predicted the monkeypox epidemic [43]. Conclusions drawn from established findings in the above fields confirm the superior performance of ConvLSTM over traditional fully connected models. In comparison to existing models in the field of epidemic prediction, our research model provides additional spatial prediction results, validating the applicability of ConvLSTM in spatiotemporal prediction.

$C_{t-j}$ as comprehensive cellular information is continuously updated and accumulated while being passed downward. Thus, the temporal features can be maintained during the information processing over a long span of time. The ability to simultaneously model temporal sequences and characterize data spatially is the reason for choosing it as the target model. It is particularly suitable for studying physical fields with strong temporal and spatial correlations.

Assuming that the input tensor of the model at moment k of the study region is $U(k)$, and considering that it will be affected by the data of the previous $k - n + 1$ sampling time periods immediately before k, the data matrix in the statistical time period ($k - n + 1$, k) is used as the input of the network model, and the convolution kernel $W$ is specified to perform the convolution operation on the data matrix Eq. 3:

$$y(k) = \sum_{u=1}^{m}\sum_{v=1}^{n} W_{uv} \cdot U(k) \tag{3}$$

where $y(k)$ represents the output result at time k, which is the predicted result of the input data $U(k)$ by the model, the m and n denote the size of the convolution kernel, specifying the size of the convolution operation, and the $W_{uv}$ represents the weight parameters of the convolution kernel.

To achieve the spatial-temporal prediction of the infected distribution, considering the dependency between each sampling time, the specific calculation procedure of the ConvLSTM layer is shown in Eqs. (4–8):

$$f_k = \sigma\left(W_{ui} \circ U(k) + W_{hi} \circ H(k-1) + W_{ci} \circ c_{k-1} + b_f\right) \tag{4}$$

$$i_k = \sigma\left(W_{ui} \circ U(k) + W_{hi} \circ H(k-1) + W_{ci} \circ c_{k-1} + b_i\right) \tag{5}$$

$$C_k = f_k \circ C_{k-1} + i_k \circ \mathrm{ReLU}\left(W_{ui} \circ U(k) + W_{hi} \circ H(k-1) + b_c\right) \tag{6}$$

$$o_k = \sigma\left(W_{ui} \circ U(k) + W_{hi} \circ H(k-1) + W_{co} \cdot c_k + b_o\right) \tag{7}$$

$$\hat{L}(k+1) = o_k \cdot \mathrm{ReLU}\left(c_k\right) \tag{8}$$

where $\circ$ denotes the convolution operation, $\bullet$ denotes the Hadamard product, $W$ is the weight matrix, $b$ is the bias, and $\sigma$ and tanh are the activation functions.

The ConvLSTM cell structure consists of a forget gate, an input gate, and an output gate. In the above equation, $f_k$ is the forgetting gate, which serves to read the hidden layer state $H(k-1)$ of the previous period and the input $C_{k-1}$ at that moment and outputs a value between 0 and 1 to the cell state $C_{k-1}$ at the $k-1$ moment, $i_k$ is the input gate, which is used to decide which information to update and obtain the new candidate cell information $C_k$ from the ReLU activation function after updating the cell After updating the state, the output gate $o_k$ is used to calculate $H(k-1)$, $C_k$ and the final output $\hat{L}(k+1)$ obtained from $C_k$.

#### 2.5.1. Model parameters determination

Reducing computational effort and maximizing training performance are the goals of network parametric tuning, which is an essential tool for optimizing model performance optimization. The parameters used to build the model structure determine the learning capability of the model. Among them, the network width, namely the number of neurons in the convolutional layers, is an important factor that affects model performance; the higher the network width, the stronger the representation capability. Too few neurons in the convolutional layers will make the model lack complexity, but overfitting occurs in complex networks with too many neurons and insufficient training data. In order to determine the optimal number of neurons each layer of the network, the following structure was designed for each convolutional layer, and Table 1b shows the training inputs and cross-entropy losses obtained by analyzing several experiments. The optimal solution ConvLSTM (3 × 3)(64, 64, 64, 1), where '3 * 3' denotes the size of convolutional kernels and (64, 64, 64, 1) denotes the number of convolutional kernels in each network layer of the four-layer network, has the minimum training loss and moderate training cost.

The batch sizes determine the model's convergence speed, and its choice affects the network training accuracy. The optimal parameters are found by plotting the loss curves with different batch sizes (Fig. 1a). The loss drop is smoother and larger when the batch size is 5, which also yields the smallest training loss and validation loss among all parameter combinations. Later, by setting different batch sizes to compare the spatial prediction of the cumulative number of infected patients (Fig. 1b), the simulation results with batch size = 5 can best reflect the true distribution of the number of cases. In summary, the best performance is

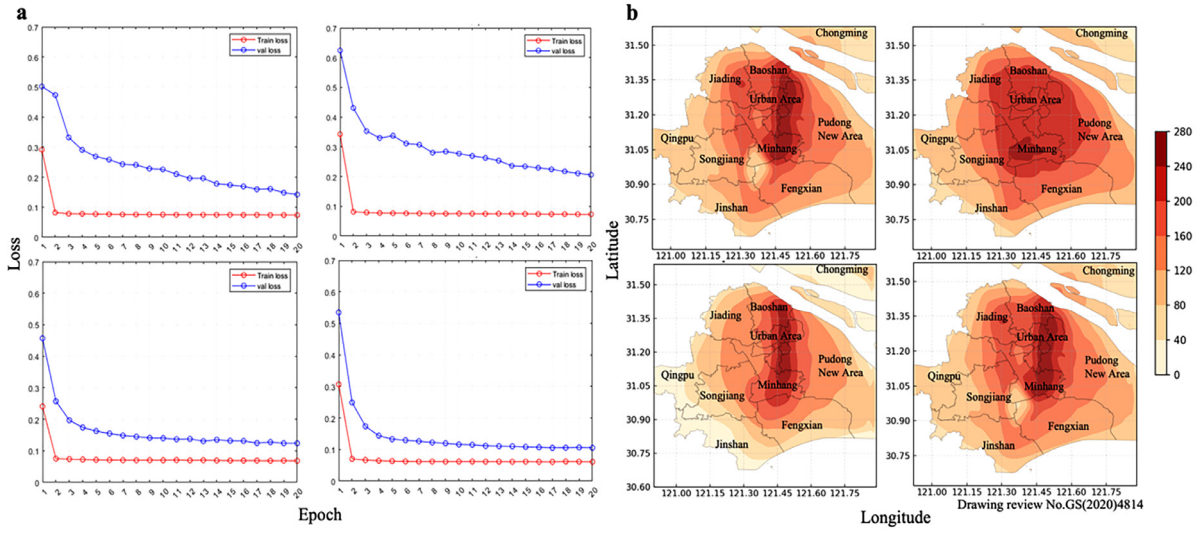


**Fig. 1. Model performance under various parameter settings**. (a) The impact of the number of epochs and batch size on the model performance. (b) Prediction of the cumulative number of infected persons using different batch sizes. Notes: in order from left to right and top to bottom, the batch size are 20, 25, 20, 5, respectively.

obtained using the model with the combination of batch size = 5 and epoch = 20.

Based on the above discussion, this model uses five consecutive time steps cases data as the input data. To balance the number of training parameters, and the sizes of the three-layer convolution kernels was set to $5 \times 5$, $3 \times 3$, and $1 \times 1$, respectively. To ensure that the output resolution of the model is consistent with the input, the padding setting was used to make the feature maps generated by the intermediate process consistent in size. Using random sampling, 80% of the sequence samples were selected as the training set, and 20% of the sequence samples were selected as the validation set.

Table 1c shows the parameter settings used in the final model. The Adam adaptive optimizer was used with the feature of setting independent learning rates by calculating the order distance of the gradient for different parameters. Cross entropy loss, which has the advantage of continuously maintaining a high gradient state and a certain convergence in many cases compared to the commonly used root mean square error, was chosen as the loss function. In this study, the cases distribution for the five days before the starting report moment was used as input, and the model outputs the prediction for the five days after the moment was output by the model.

### 2.5.2. Model construction process

To comprehensively analyze the spatial and temporal characteristics of the epidemic development, this paper further considers the mechanism of meteorological influence on disease transmission in depth based on the constructed ConvLSTM model, and multiple meteorological factors are added to the model. First, we construct a simple model, called the ConvLSTM model, in which only the number of historical cases is considered. The Meteor-ConvLSTM model, in which the influence of meteorological factors on the distribution of cases is also considered, is constructed to better predict the spatial distribution of actual cases compared to the ConvLSTM model setup, which considers the contribution of meteorological factors.

The structure of the ConvLSTM neural network for the high-precision spatial-temporal prediction of cases is shown in Fig. 2. The ConvLSTM spatial-temporal depth model is composed of an input layer (i.e., the predictor), three convolutional layers composed of ConvLSTM2D, a BN layer (batch normalization layer) after the convolutional layers, and an output The input layer (i.e., the predictor) is a combination of three convolutional layers composed of ConvLSTM2D, a BN layer after convolution, and an output layer (the predictand). The input number of historical cases is the mapping time $t-5, \ldots, t-1$ (in days), and finally the information on the distribution of cases at a future moment is output through a three-dimensional convolutional layer (i.e. Conv3D). Each layer of the model uses ReLU as the activation function to improve the nonlinear expression of the model, and the daily number of new cases at $t+1, \ldots, t+5$ is used as the variable in the output layer.

The ConvLSTM2D block diagram represents the convolution process of learning the spatial-temporal characteristics of the sample data, and its internal structure is shown in the upper structure of the block diagram, which has a convolution structure in the input-to-state and state-to-state transitions. In this model, a coded prediction structure is formed by superimposing multiple ConvLSTM2D layers, thereby building a network model of the case-count real-time-forward prediction problem. For the meteorological factors added to the input, a two-layer wide artificial neural network is used to receive this feature map to learn the driving characteristics of meteorological factors on the case number distribution. For the Meteor-ConvLSTM, fourteen time-series inputs are used; the first five time series are the historical case counts for the previous five days, the last nine time series are the meteorological factors at the current moment, and the output is the output for the next five time-series. In summary, compared to the ConvLSTM setup, a layer of an artificial neural network with the associative ability to approximate any nonlinear relationship is added to the Meteor-ConvLSTM training process to learn the meteorological factor features, and the learned features are expected to be used to optimize the output of the ConvLSTM. In this way, the end-to-end structure of the model is established, and the whole model consists of four major components: the input module, the encoder module, the forecaster module, and the output module.

## 3. Results and discussion

### 3.1. Evaluation index

In addition to the conventional assessment of the average prediction level throughout the study period, it is essential to measure the predictive capacity of the peak case distribution. To evaluate the performance

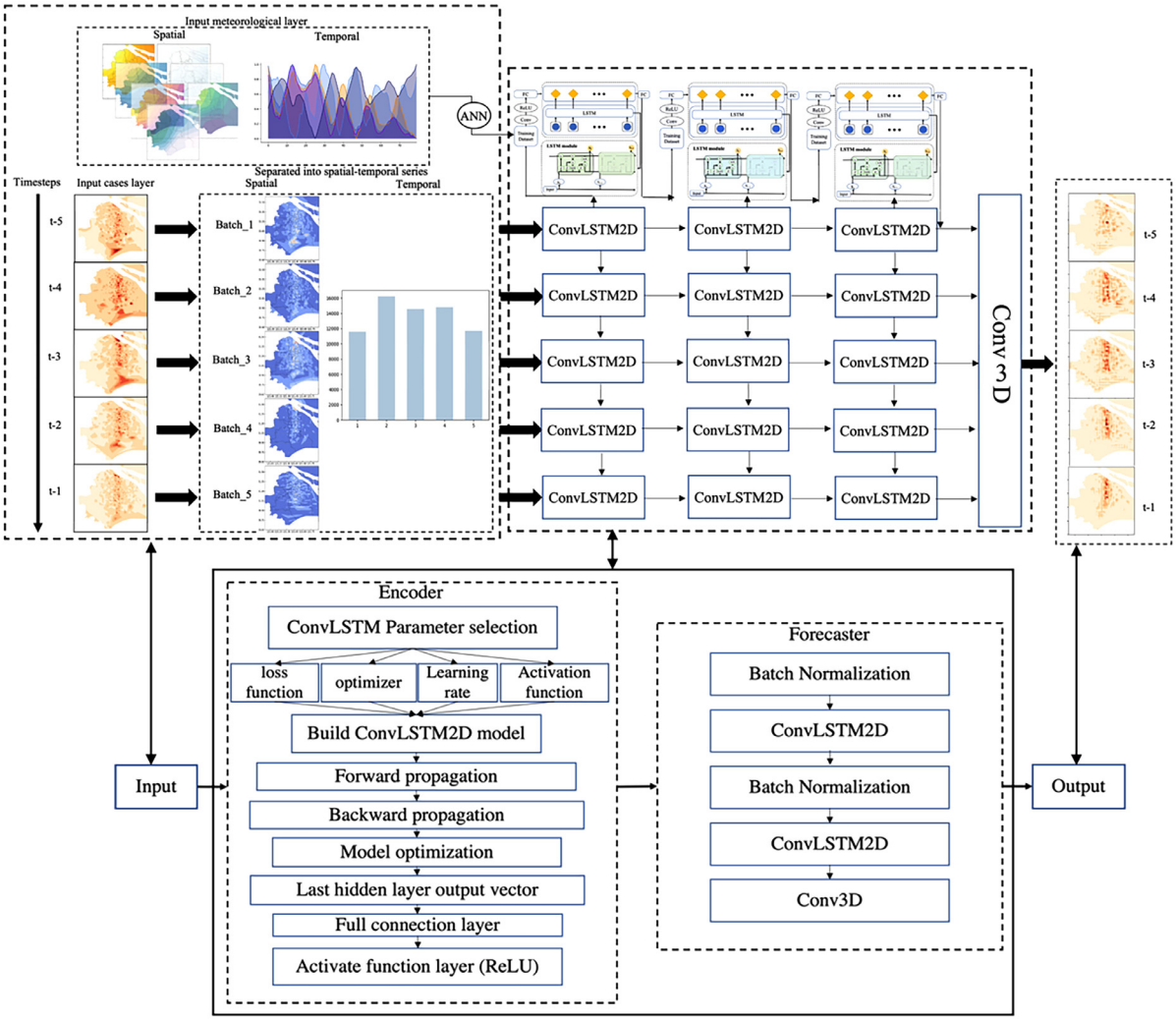


**Fig. 2. Framework of Meteor-ConvLSTM and corresponding end-to-end structure.**

of the model in terms of predicting the peak infection concentration bands and the average prediction level, we use the evaluation metrics of the Root Mean Square Error(RMSE) (Eq. 9), Mean Absolute Error (MAE) (Eq. 10), and coefficient of determination ($R^2$) (Eq. 11) are used to quantify the prediction effectiveness of the model:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum\left(\hat{y}_i - y_i\right)^2} \tag{9}$$

$$\mathrm{MAE} = \frac{1}{n}\sum\left|\hat{y}_i - y_i\right| \tag{10}$$

$$R^2 = 1 - \frac{\mathrm{SSE}}{\mathrm{SST}} = 1 - \frac{\sum\left(y_i - \hat{y}_i\right)^2}{\sum\left(y_i - \overline{y_i}\right)^2} \tag{11}$$

where $\hat{y}_i$ is the predicted and $y_i$ represents the real, the sum of squared errors (SSE) is the squared sum of the deviation between the real and the predicted, and the sum of squared total (SST) is the squared sum of the deviation between the true and the mean of the true. Among them, the RMSE reflects the average deviation between the predicted and the real; the MAE reflects the actual situation of the predicted value error; the $R^2$ reflects the goodness of fit between the predicted and the real of the model, which is an intuitive measure of model merits and demerits. The lower the RMSE and MAE, the higher the model prediction accuracy; $R^2 \in [0,1]$, and a larger $R^2$ indicates a better model fit.

### 3.2. Spatial-temporal distribution of the 3.15 epidemic in Shanghai

In the spatial distribution (Fig. 3a), the hotspots of case clustering were mainly located in the southern part of Baoshan district, a large area in Shanghai central district, and near the intersection of Minhang District and Pudong New Area, with the characteristics of being disseminated from this center to the surrounding, and similar to the distribution of population density (Fig. 3c), which means areas with high population density corresponded to cases hotspots. In the temporal distribution (Fig. 3a), The number of cases remained high from day 30 to day 50 after the first case appeared on March 1 and peaked on April 13 with more than 20,000 new cases in a single day, then slowly declined to double digits.

COVID-19 is an infectious disease where human-to-human transmission occurs based on the spread of pathogenic microorganisms through respiratory droplets and contact. In addition, Omicron BA.2, which was prevalent in Shanghai during this outbreak, has an excellent transmission capacity of $R_0 = 9.5$ [44], where $R_0$ represents the number of transmissible new infections in 100 per cent of the susceptible population of

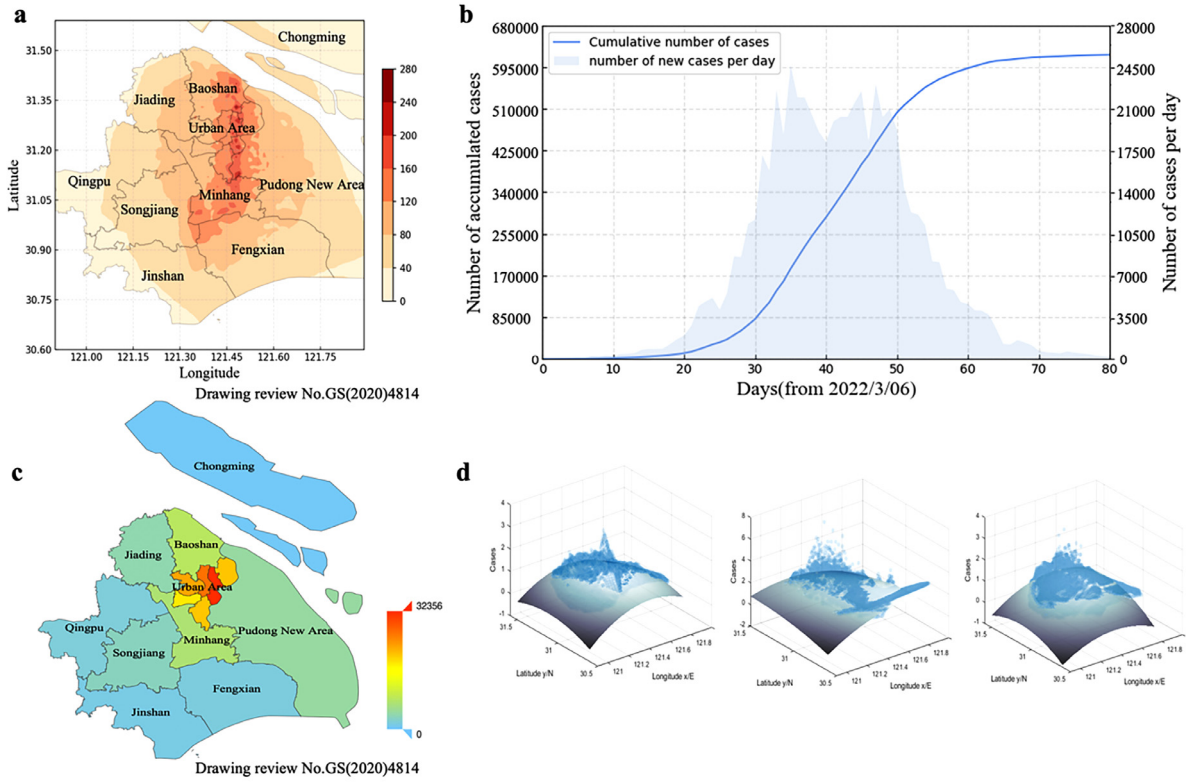


**Fig. 3. Spatial-temporal overview and statistical characteristics of infections in Shanghai**. (a) Spatial distribution of cumulative infections. (b) Time series of cumulative infections. (c) Population density distribution (person/km$^2$). (d) Trend surface characteristics by the epidemic three stages.

an infectious patient without any preventive measures. $R_0 = 9.5$ meaning that one positive infected person can infect roughly ten healthy susceptible persons. The spreading capacity of Omicron BA.2 is 3–5 folds higher than the earlier virus. An autocorrelation analysis of Shanghai cases distribution was conducted using Moran's *I* to find the statistical significance of the positive correlation between cases distribution and location. Moran's *I* shows 0.9567, indicating that there was a strong tendency for the high-density gathering of cases spreading to surroundings.

This instance of the epidemic was divided into three periods. Starting from March 1, Day 10 to Day 29 was the brewing period, Day 30 to Day 50 was the sustained peak period, and Day 51 to Day 80 was the upturn period of the epidemic. The analysis of dataset general trend (Fig. 3d) from the generated 3D maps of the spatial trends of cases in the three periods reveals: the first and third periods have the same elliptical parabolic distribution pattern, with the infected clustered in research region, and the second period has an elliptical cylindrical distribution pattern, where the characteristics of a uniform distribution are more remarkable. This is consistent with the experiential process and spatial and temporal diffusion characteristics of the epidemic occurrence, rapid development, and stabilization processes, respectively.

### 3.3. Shanghai 3.15 epidemic prediction

In this section, we conducted a backward study of Shanghai 3.15 epidemic, to make theoretical predictions of the spatial and temporal epidemic and analyze the influence of relevant meteorological factors on epidemic development and prediction improvement. The prediction of cases distribution at the infection peak is a major concern in all sectors. This Shanghai epidemic peaked on April 13 (i.e., the 43rd day after March 1). In the well-trained model, the numbers of cases from Days 35 to 40 were inputted to make predictions for Days 41 to 45, and the result (Fig. 4a) provides insight into the model potential in predicting the peak, infection-intensive bands. In addition, to provide a baseline for an overall epidemic control over the entire period, the average prediction for each five-day was evaluated (Fig. 4b), which shows that the predicted spatial distribution is roughly similar to the actual, and the seriousness roughly in line with the actual, but the overall prediction is abnormally high. One of the reasons for the deviation between the predicted of ConvLSTM model and the actual is that meteorological factors effect is not considered, so we developed a modified COVID-19 prediction model (Meteor-ConvLSTM) that takes meteorological factors into accounts.

Correlation analysis is a valid method for recognizing the interrelationships between variables with spatial properties and their patterns of association, and such correlation can provide a physical explanation of disease impact mechanisms to some extent. The study analyzed the correlation heatmap between cases and each meteorological factor (Fig. 5a), and selected the three factors (2m_dt, sk_t, v10) with the most significant correlations, which affect disease spread the most, for further correlation analysis. The spatial correlation analysis between cases and meteorological factors facilitates a reasonable physical explanation of meteorological factors' influences on model predictions: skin temperature (Fig. 5b), humidity (Fig. 5c) and cases were highly correlated in areas located in and around Shanghai central district, indicating that humidity conditions and skin temperature conditions in these highly correlated areas are key factors in the distribution of cases as predicted by model. Human coronavirus activeness is susceptible to the humidity and temperature conditions, and suitable ambient temperature and humidity conditions can cause abnormal virus multiplication, providing a favorable climate for the breeding and transmission of infectious disease viruses. Due to the difference in epidemic spread and prevention conditions in each district, Shanghai central district and its neighboring areas with high population densities and high case concentrations are vulnerable to becoming hotbeds for virus multiplication. The suitable temperature and humidity conditions for virus spread and reproduction in urban areas have higher explanatory significance for effective

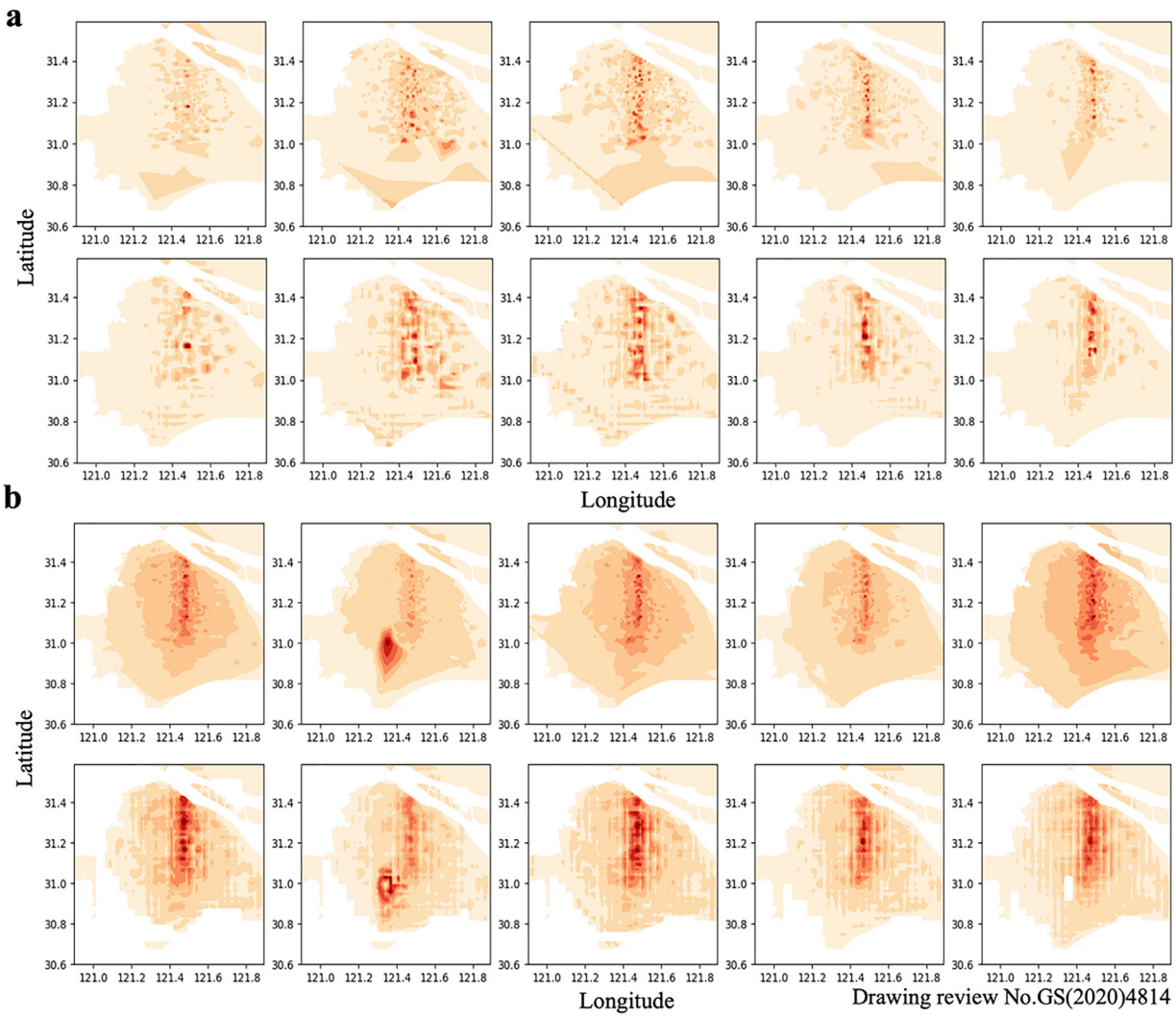


**Fig. 4. Actual cases distribution (pictured above) and the prediction (pictured below) using ConvLSTM**. (a) Number of cases at peak. (b) Average number of cases every 5 days on entire period.

epidemic prediction. The highly correlated areas about 10 m southeast wind field (Fig. 5d) were in the Jiading, Minhang, and Pudong districts, implying that ground-level winds over these areas were a key factor in the successful case prediction.

Meteor-ConvLSTM is used to explore the spatial-temporal case distributions based on both meteorological factors and historical case counts. The predicted case distribution at the peak (Fig. 6a), averaged over the entire study period (Fig. 6b), is shown compared with the actual distribution. The spatial distribution of the Meteor-ConvLSTM predicted results is similar to the actual distribution with a goodness of fit $R^2$ reaching 0.692, which is improved by 0.125 compared to ConvLSTM. The severity of the epidemic is consistent with the actual, which means a inclusion of meteorological factors corrects the overall abnormal high prediction of ConvLSTM.

The Meteor-ConvLSTM, which combines the epidemiological characteristics and the transmission dynamics influenced by meteorological factors, can be used to obtain more accurate and higher resolution prediction results than the original ConvLSTM. The following analysis of scatter density plots and evaluation indicators was used to objectively and quantitatively assess the improvement of Meteor-ConvLSTM compared to the original model. From the scatter density analysis of the two models (Fig. 7), the scatter points are both distributed in a tilted elliptical, and Meteor-ConvLSTM (Fig. 7b) has a narrower ellipse and denser scatter distribution compared to ConvLSTM (Fig. 7a). The error indicators of Meteor-ConvLSTM (MAE = 0.409, MSE = 0.351, RMSE = 0.592) are lower than ConvLSTM (MAE = 0.491, MSE = 0.492, RMSE = 0.702), and the $R^2$ has a significant improvement (ConvLSTM: $R^2$ = 0.567; Meteor-ConvLSTM: $R^2$ = 0.692). In summary, as mentioned earlier, this study optimized the model fit based on additional high-precision meteorological factors, and the Meteor-ConvLSTM has high prediction accuracy compared to the original ConvLSTM.

### 3.4. Discussion

Shanghai has a high international flight interchange frequency, with nearly 40% of international flights in China landing in Shanghai since 2020 [45]. Under the severe global epidemic environment in the first half of 2022, it suffers from the severe impact of the global epidemic and has the highest risk of infection import compared to other cities in mainland China. Therefore, the study of Shanghai as a typical case is of practical significance. After the release of the city-wide lockdown policy, the spread of the epidemic gradually slowed down, and the implementation of several intervention policies kept the spread rate below the epidemic threshold, allowing the Shanghai 3.15 epidemic to be finally arrested. The empirical evidence that the city-wide containment and control implemented in Shanghai successfully prevented further epidemic spread [46], which suggests that reasonable prevention measures for major public health events with high population concentrations and

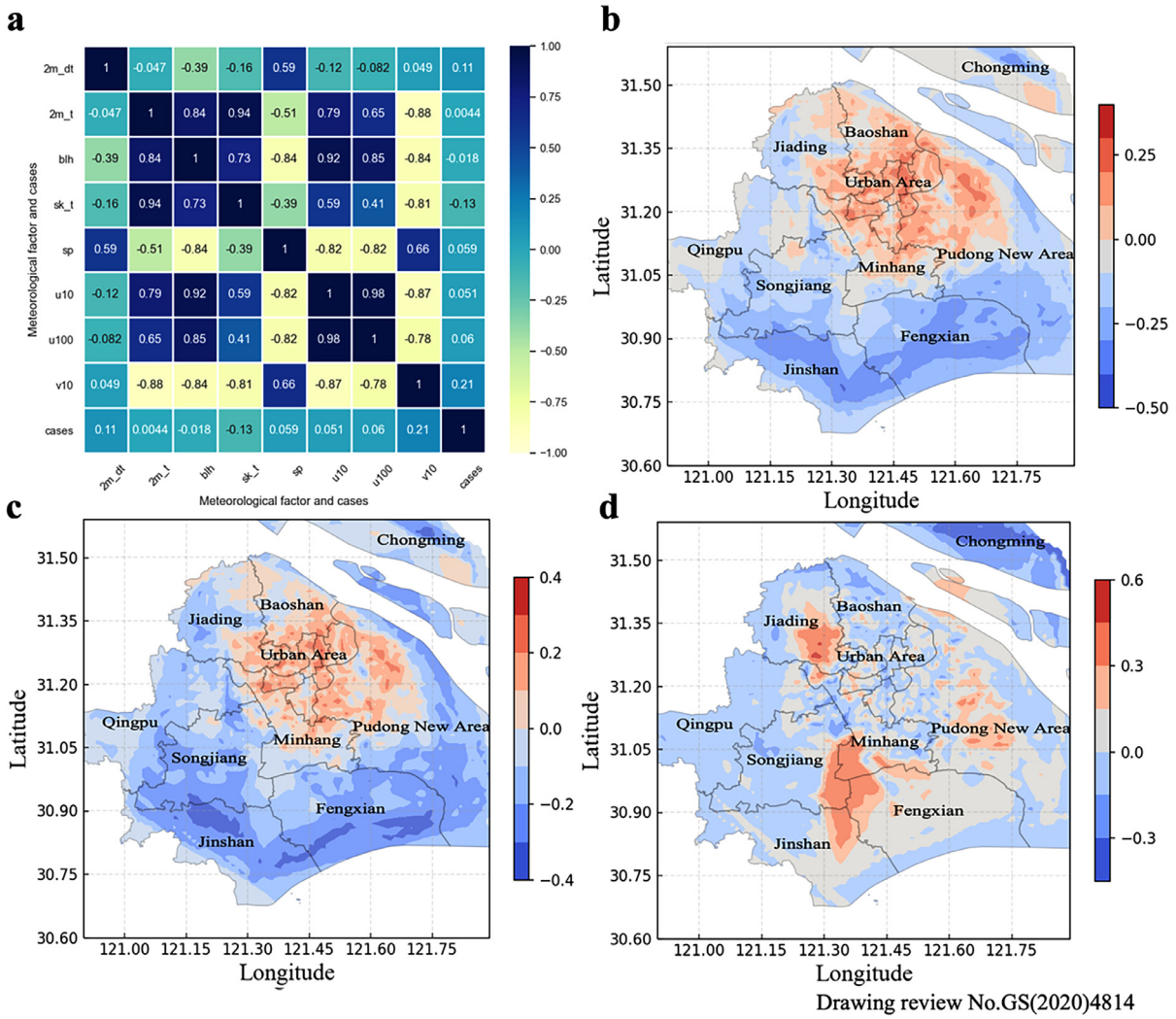


**Fig. 5. Correlation analysis between meteorological factors and cases**. (a) Heatmap of correlation. (b) Distribution of $R$ between skin temperature and the number of cases. (c) Distribution of $R$ between 2 m dew point temperature and the number of cases. (d) Distribution of $R$ between 10 m south-east wind and the number of cases.

highly transmissible diseases are effective. However, quantitative assessment of the socio-economic impact of prevention measures, the burden of disease, and a comprehensive measure for balancing population immunity are necessary, and will serve as future research directions.

The general overview of the previous studies' prediction effects of each statistical model and deep learning model is shown in Table 2. Compared with using traditional models providing one-dimensional time-series forecasts, the prediction error is significantly reduced using the ConvLSTM model. With a temporal resolution at the daily scale and a spatial grid of $0.01^\circ \times 0.01^\circ$ for the city scale, a high accuracy spatial-temporal prediction about new cases was provided. Because of the higher resolution refinement than previous studies, and the interruption of the transmission chain due to the control policy of epidemic prevention in Shanghai, the $R^2$ of the ConvLSTM model failed to exceed that of each statistical, deep learning model based on one-dimensional time-series prediction. Therefore, it is of great practical importance for the enhancement of model refinement prediction effects at high spatial and temporal resolutions. However, the Meteor-ConvLSTM model that considers the influence of meteorological factors can significantly improve the simulation accuracy, even though it exists in the premise of refined COVID-19 spatial-temporal grid prediction.

Other models exhibit a higher $R^2$ at the national level, indicating their suitability for large-scale data variations. However, these models only consider time series information, simplifying training and prediction problems. Therefore, an overall gain is observed in $R^2$, while the high $R^2$ resulting from the simplified time prediction does not necessarily demonstrate high performance in spatiotemporal prediction. Additionally, existing models show a relatively large RMSE, suggesting poor performance in predicting results.

Compared to existing research models, our model has a lower $R^2$ and lower RMSE. However, facing a more complex prediction problem, the lower $R^2$ of ConvLSTM suggests that the model can not perfectly explain the variability of the target variable. Yet, considering its responsibility to provide estimates at both the temporal and spatial levels, the lower $R^2$ performance is acceptable. The lower RMSE indicates accurate predictions on an average level. Still, it can not rule out the possibility that the low RMSE is due to the study's limitation to a small area (Shanghai). Therefore, a lateral comparison with existing research is conducted: within the premise of a small study area, the RMSE and MAE of LSTM case predictions in various states in Malaysia are higher than those of our research model [33], indicating that within a small study area, the selected ConvLSTM model performs better, making predictions closer to the actual observed values.

In conclusion, considering model performance and the necessity of spatial prediction, the choice of the ConvLSTM model is advantageous.

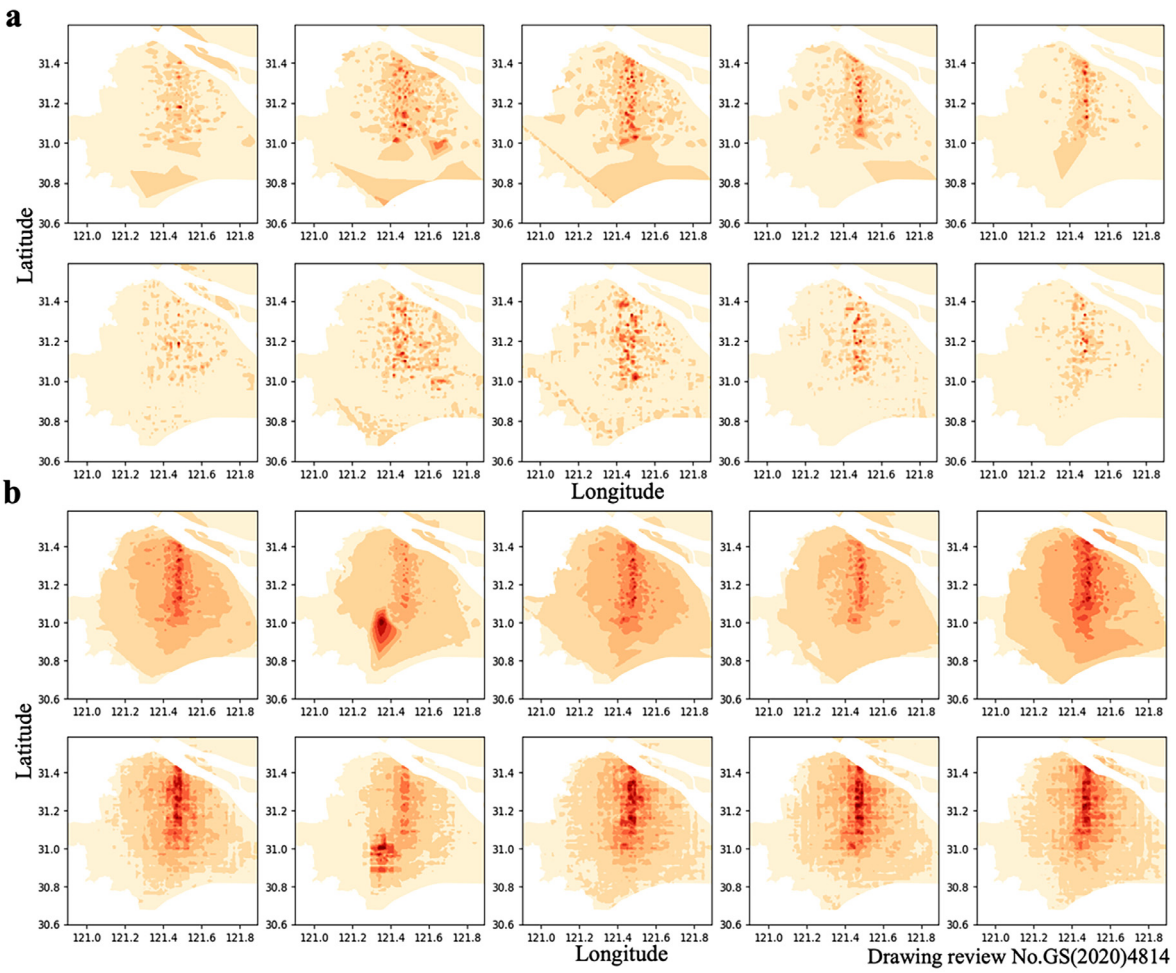


**Fig. 6. Actual cases distribution (pictured above) and prediction (pictured below) using Meteor-ConvLSTM.** (a) The number of cases at peak. (b) Average number of cases every 5 days on entire period.

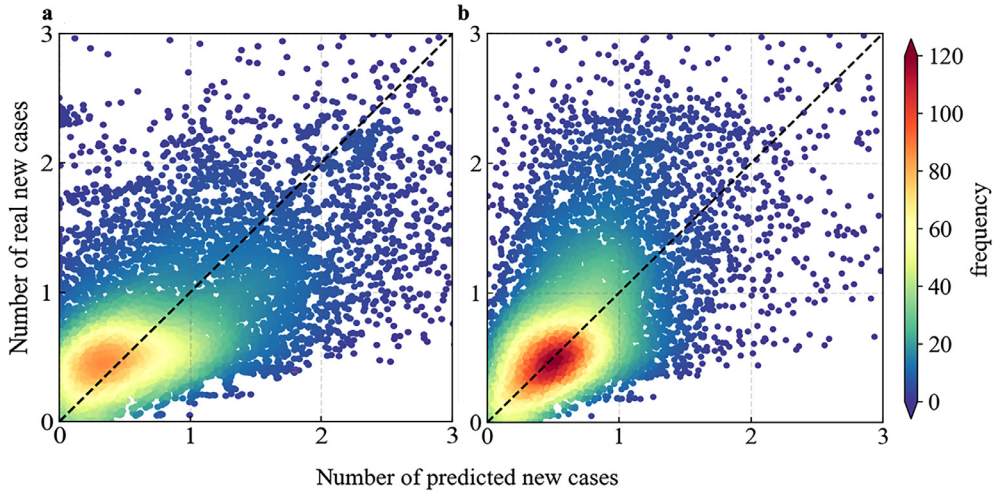


**Fig. 7. Scatter plot of the correlation between the actual and the predicted.** (a) Using ConvLSTM. (b) Using Meteor-ConvLSTM.

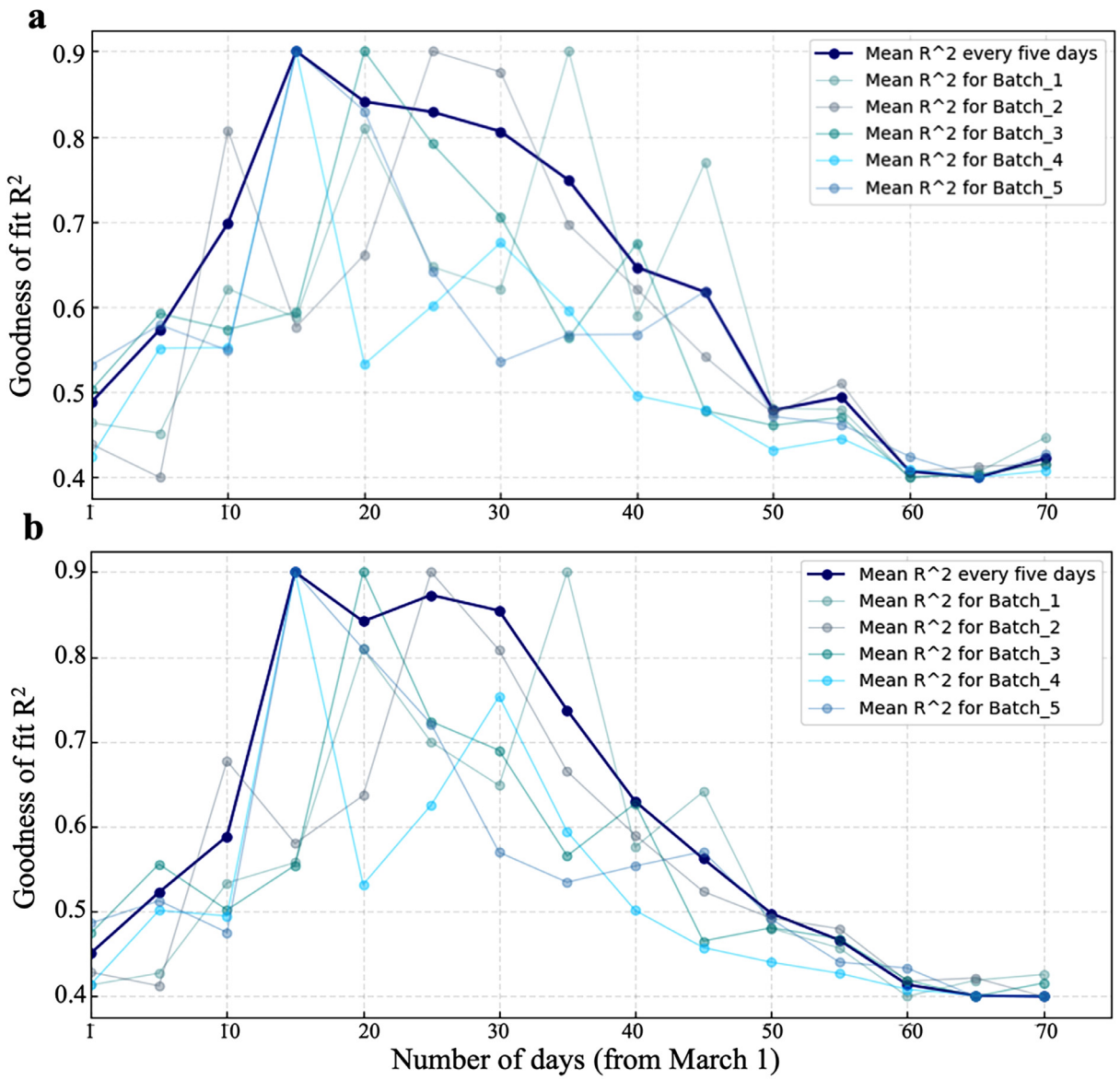


**Fig. 8. Model performance shown by time series $R^2$.** (a) ConvLSTM $R^2$. (b) Meteor-ConvLSTM $R^2$.

**Table 2**
**Overview of prediction effect of existing research.**

| Model | $R^2$ | RMSE | MAE | Region | Period | Research object | Resolution |
|---|---|---|---|---|---|---|---|
| TXGB [19] | / | 338.32 | 69.60 | USA | 2020.10–2021.1 | Cases | Weekly |
| STXGB-FB [19] | / | 308.49 | 64.33 | USA | 2020.10–2021.1 | Cases | Weekly |
| STXGB-SG [19] | / | 318.00 | 64.86 | USA | 2020.10–2021.1 | Cases | Weekly |
| ANN [20] | 0.98 | 17,554.00 | 12,229.00 | Global | 2020.1–2020.11 | Cases | Daily |
| ANN [20] | 0.86 | 631.80 | 463.70 | Global | 2020.1–2020.11 | Deaths | Daily |
| LR [21] | 0.95 | 916.64 | 723.11 | Global | 2020.1–2020.3 | Deaths | Daily |
| LASSO [21] | 0.81 | 1801.12 | 1430.29 | Global | 2020.1–2020.3 | Deaths | Daily |
| ES [21] | 0.97 | 813.77 | 406.08 | Global | 2020.1–2020.3 | Deaths | Daily |
| LR [21] | 0.83 | 383,910.51 | 30,279.55 | Global | 2020.1–2020.3 | Cases | Daily |
| LASSO [21] | 0.98 | 15,322.11 | 60,177.90 | Global | 2020.1–2020.3 | Cases | Daily |
| ES [21] | 0.98 | 16,828.58 | 8867.43 | Global | 2020.1–2020.3 | Cases | Daily |
| RF [25] | 0.84 | 135.57 | 34.59 | China | 2020.1–2020.3 | Cases | Daily |
| GBDT [25] | 0.82 | 178.10 | 41.37 | China | 2020.1–2020.3 | Cases | Daily |
| GWR [25] | 0.81 | 194.01 | 52.98 | China | 2020.1–2020.3 | Cases | Daily |
| ARIMA [14] | / | 1654.66 | 984.17 | Italy | 2020.2–2020.4 | Cases | Daily |
| ARIMA [14] | / | 2031.12 | 1147.89 | Spain | 2020.2–2020.4 | Cases | Daily |
| ARIMA [14] | / | 971.93 | 635.87 | France | 2020.2–2020.4 | Cases | Daily |
| ConvLSTM | 0.57 | 0.70 | 0.49 | Shanghai | 2022.3–2022.5 | Cases | Daily |
| Meteor-ConvLSTM | 0.69 | 0.59 | 0.41 | Shanghai | 2022.3–2022.5 | Cases | Daily |

Since 2023, the new wave of COVID-19 in China has exhibited a low-level, fluctuating pattern, with the overall situation of the pandemic remaining stable. It is now imperative to maintain the overall stability of the nationwide epidemic situation during clustered holidays, while implementing various infectious disease prevention and control measures. The focus is on preventing and controlling the potential occurrence of clustered epidemics such as COVID-19, dengue fever [47,48], and monkeypox [43] in individual regions during holidays. In this environment of domestic epidemic prevention and control policy transition, the in-depth research and judgment for disease transmission ability and refined prediction of new cases are still meaningful. The assurance and judgment of the development tendency reduce the uncertainty of people's judgment and estimation of the future situations, and it can reduce the negative impact of secondary transmission of the epidemic on regular production and livelihood order. It is indicative in the future when facing the challenges of similar major public health events, and how to achieve more efficient prevention with lower management resources and less socio-economic costs will be a key concern from now on.

The $R^2$ temporal variation by ConvLSTM and Meteor-ConvLSTM was used to understand the status of the influence of meteorological factors under different periods of epidemic development (Fig. 8). This research aspect enables a direct investigation of the influence of meteorological contributions on the spatial and temporal spread of the Omicron BA.2 variant. $R^2$ is at a high level in the peak stage of epidemic spread, which fully proves the reliability of the ConvLSTM in predicting the high infectious period; There are different effects at different stages of the epidemic development, meteorological factors have a positive effect on the model prediction in the interval starting on March 26 and ending on April 16. Meteorological conditions had a positive interpretation of virus spread during the peak of the epidemic development. It is shown that the advantage of the meteorological contribution is that the inclusion of meteorological factors leads to higher accuracy of model prediction when the infectious disease infection density is higher.

In the context of research on epidemic-related topics, historical case data serves as crucial research material. The daily epidemic reports released by health commissions at various levels are considered the most authoritative data source. Many studies related to the epidemic rely on these reports as fundamental supporting materials [31,32,42]. Therefore, this study is based solely on health commission data for predictive experiments, discussing the predictive effects of spatiotemporal models using the available reported data.

In addition to the meteorological factors used in this study, various elements such as population migration, contact patterns, and virus variations also play crucial roles in the dynamic process of the COVID-19 pandemic. To specifically explore the individual impact of meteorological factors on the epidemic, research collected the Migration Scale Index (MSI) for 30 cities, controlling for population migration [30]. They employed a nonlinear regression analysis, controlling for a 3-day moving average of MSI, to independently assess the influence of meteorological factors on the number of COVID-19 cases. Population migration and contact patterns directly affect the transmission pathways and speed of the virus. Population migration leads to population movement, exacerbating the spread of the epidemic, and areas with high population flow, such as transportation hubs, city centers, and markets, may be more prone to virus transmission. The aerosol transmission of COVID-19 is of significant concern, emphasizing the need for early monitoring in densely populated areas to prevent outbreaks [31]. Different social and contact networks impact the virus's transmission pathways within communities, thus influencing the spread rate of cluster outbreaks. A crucial study, based on statistical methods and an improved epidemiological model, developed the Global Pandemic Early Warning System (GPEP-2), introducing parameterized schemes for natural and anthropogenic factors, successfully predicting the global and regional spread of COVID-19 [32]. Additionally, policies and intervention measures also influence the virus transmission patterns. Research employed a dose-response relationship to assess the relationship between infection risk and exposure dose [33]. In future studies on cluster outbreaks, our team will comprehensively consider influencing factors to conduct a comprehensive predictive analysis of epidemic development.

## 4. Conclusion

This study utilized aggregated data of new cases collected from the 3.15 epidemic in Shanghai, China, along with the high-resolution ERA5 daily reanalysis of meteorological data. We innovatively proposed a ConvLSTM model that combines CNN image processing capabilities with a traditional LSTM time-series predictor to extract spatial-temporal features of case information by dividing the study area into a $0.01^\circ \times 0.01^\circ$ grid structure. In general, we constructed the following models respectively: a ConvLSTM based on autoregression of historical case numbers, and a Meteor-ConvLSTM that takes meteorological factors into account to modify predictions. Firstly, we described the high aggregation characteristics of the current epidemic in an overview, and specifically characterized the spatial patterns of different periods of the epidemic to provide workable spatial analysis for follow-up predictions. As changes in the signals of meteorological factors (temperature, humidity, wind, air pressure, diffusion conditions, etc.) are a part of climate and environmental changes, which to a large extent affect the rate of virus transmission and lethality, therefore, Meteor-ConvLSTM is based on ConvLSTM, using meteorological factors to simulate the driving mode of case growth and spread, and establishes an integrated model using an ANN is established to learn the characteristics of meteorological factors on disease spread, and collaboratively puts multiple meteorological factors into the model for training in order to make accurate predictions of the spatial-temporal cases distribution.

Evaluating the spatial-temporal prediction effects of the ConvLSTM and Meteor-ConvLSTM, the results showed the following:

- Case clustering hotspots correspond to areas of high population density, with a strong tendency for high-density infected concentrations to spread to the surrounding areas.
- Skin temperature, 2 m dew point temperature and 10 m south-east wind field were the three factors that most affected epidemic transmission. The high correlation areas of skin temperature, humidity with cases were located in and around the Shanghai central district, and the high correlation areas of the 10 m wind field were located in the Jiading, Minhang, and Pudong New Districts. The meteorological conditions in the highly correlated areas had high explanatory significance for effective epidemic prediction.
- ConvLSTM can provide additional spatial prediction compared to traditional dynamic and deep learning models with one-dimensional time-series prediction. The modified prediction model can effectively depict the process of meteorological factors affecting infectious disease transmission. Moreover, Meteor-ConvLSTM, which considers meteorological factors, has higher prediction accuracy than the original ConvLSTM.

## Declaration of competing interest

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

## Acknowledgments

The work was jointly supported by the Self-supporting Program of Guangzhou Laboratory (SRPG22-007), the Collaborative Research Project of the National Natural Science Foundation of China (L2224041), and the Chinese Academy of Sciences (XK2022DXC005): Frontier of Interdisciplinary Research on Monitoring and Prediction of Pathogenic Microorganisms in the Atmosphere, and the Fundamental Research Funds for the Central Universities (lzujbky-2023-ey10). This study was completed with the updated and timely, high-geographic information resolution daily positive patient summary tables enhanced

by the Shanghai Municipal Health and Wellness Commission (https://wsjkw.sh.gov.cn); and daily high-resolution reanalysis information from the 2022 ERA5 (https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview), we would like to express our sincere gratitude to the above.

## References

[1] W.J. Guan, Z.Y. Ni, Y. Hu, et al., Clinical characteristics of coronavirus disease 2019 in China, N. Engl. J. Med. 382 (2020) 1708–1720.
[2] W.J. Guan, Z.Y. Ni, Y. Hu, et al., Clinical characteristics of coronavirus disease 2019 in China, J. Integrat. Med. 80 (6) (2020) 656–665.
[3] A. Subramanian, K. Nirantharakumar, S. Hughes, et al., Symptoms and risk factors for long COVID in non-hospitalized adults, Nat. Med. 28 (2022) 1706–1714.
[4] D. Guan, D. Wang, S. Hallegatte, et al., Global supply-chain effects of COVID-19 control measures, Nat. Hum. Behav. 4 (2020) 577–587.
[5] D. Wang, K. Hubacek, X. Liang, et al., Reply to: Observed impacts of the COVID-19 pandemic on global trade, Nat. Hum. Behav. (5) (2021) 308–309.
[6] A. Josephson, T. Kilic, J.D. Michler, Socioeconomic impacts of COVID-19 in low-income countries, Nat. Hum. Behav. (5) (2021) 557–565.
[7] D. Laborde, A. Herforth, D. Headey, et al., COVID-19 pandemic leads to greater depth of unaffordability of healthy and nutrient-adequate diets in low- and middle-income countries, Nat. Food 2 (2021) 473–475.
[8] H. Tian, Y. Liu, Y. Li, et al., An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China, Science 368 (2020) 638–642.
[9] A. Zeroual, F. Harrou, A. Dairi, et al., Deep learning methods for forecasting COVID-19 time-series data: A comparative study, Chaos, Solitons Fractals 140 (2020) 110121.
[10] S.M. Kissler, C. Tedijanto, E. Goldstein, et al., Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period, Science 368 (2020) 860–868.
[11] M.H.D.M. Ribeiro, R.G. da Silva, V.C. Mariani, et al., Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil, Chaos, Solitons Fractals 135 (2020) 109853.
[12] H. Jo, J. Kim, T.-C. Huang, et al., condLSTM-Q: A novel deep learning model for predicting Covid-19 mortality in fine geographical scale 10 (2022) 125.
[13] A. Mollalo, R. Kiara, V BehzadTI, Artificial neural network modeling of novel coronavirus incidence rates across the continental United States, Int. J. Environ. Res. Public Health 17 (2020) 4204.
[14] Z. Ceylan, Estimation of COVID-19 prevalence in Italy, Spain, and France, Sci. Total Environ. 729 (2020) 138817.
[15] F. Shahid, A. Zameer, M. Muneeb, Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM, Chaos, Solitons Fractals 140 (2020) 110212.
[16] J. Huang, X. Liu, L. Zhang, et al., The oscillation-outbreaks characteristic of the COVID-19 pandemic, Natl. Sci. Rev. 8 (2021) nwab100.
[17] C. Liu, J. Huang, S. Chen, et al., The impact of crowd gatherings on the spread of COVID-19, Environ. Res. 213 (2022) 113604.
[18] K.C. Santosh, COVID-19 prediction models and unexploited data, J. Med. Syst. 44 (2020) 170.
[19] B. Vahedi, M. Karimzadeh, H. Zoraghein, Spatiotemporal prediction of COVID-19 cases using inter- and intra-county proxies of human interactions, Nat. Commun. 12 (2021) 6440.
[20] Q. Guo, Z. He, Prediction of the confirmed cases and deaths of global COVID-19 using artificial intelligence, Environ. Sci. Pollut. Res. 28 (2021) 11672–11682.
[21] F. Rustam, A.A. Reshi, A. Mehmood, et al., COVID-19 future forecasting using supervised machine learning models, IEEE Access. 8 (2020) 101489–101499.
[22] R.K. Singh, M. Rani, A.S. Bhagavathula, et al., Prediction of the COVID-19 pandemic for the top 15 affected countries: Advanced autoregressive integrated moving average (ARIMA) model, JMIR Public Health Surveill. 6 (2020) e19115.
[23] K.E. ArunKumar, D.V. Kalaga, C. Mohan Sai Kumar, et al., Comparative analysis of Gated Recurrent Units (GRU), long Short-Term memory (LSTM) cells, autoregressive Integrated moving average (ARIMA), seasonal autoregressive Integrated moving average (SARIMA) for forecasting COVID-19 trends, Alexandria Eng. J. 61 (2022) 7585–7603.
[24] Y. Han, J. Huang, R. Li, et al., Impact analysis of environmental and social factors on early-stage COVID-19 transmission in China by machine learning, Environ. Res. 208 (2022) 112761.
[25] Q. Shao, Y. Xu, H. Wu, Spatial prediction of COVID-19 in China based on machine learning algorithms and geographically weighted regression, Comput. Math. Methods Med. 2021 (2021) 7196492.
[26] C. Mora, T. McKenzie, I.M. Gaw, et al., Over half of known human pathogenic diseases can be aggravated by climate change, Nat. Clim. Chang. 12 (2022) 869–875.
[27] M. Kang, J. Wei, J. Yuan, et al., Probable evidence of fecal aerosol transmission of SARS-CoV-2 in a high-rise building, Ann. Intern. Med. 173 (2020) 974–980.
[28] M.Z.R. Sabuj, M.A. Rashid, T.R. Dargaville, et al., Stability of inhaled ciprofloxacin-loaded poly(2-ethyl-2-oxazoline) nanoparticle dry powder inhaler formulation in high stressed conditions, Pharmaceuticals 15 (2022) 1223.
[29] Y. Xia, Airborne transmission of pathogen-laden expiratory droplets in open outdoor space, Sci. Total Environ. 773 (2021) 145537 v.–2021 v.773.
[30] J. Liu, J. Zhou, J. Yao, et al., Impact of meteorological factors on the COVID-19 transmission: A multi-city study in China, Sci. Total Environ. 726 (2020) 138513.
[31] J. Huang, D. Wang, Y. Zhu, et al., An overview for monitoring and prediction of pathogenic microorganisms in the atmosphere, Fund. Res. (2023). https://doi.org/10.1016/j.fmre.2023.05.022.
[32] J. Huang, L. Zhang, B. Chen, et al., Development of the second version of the global prediction system for epidemiological pandemic, Fund. Res. (2023). https://doi.org/10.1016/j.fmre.2023.02.030.
[33] N.W. Shen, A. Abu Bakar, H. Mohamad, Univariate and Multivariate Long Short Term Memory (LSTM) model to predict Covid-19 cases in Malaysia using integrated meteorological data, Malaysian J. Fundam. Appl. Sci. 19 (4) (2023) 653–667.
[34] C. Sun, L. Chao, H. Li, et al., Modeling and preliminary analysis of the impact of meteorological conditions on the COVID-19 epidemic, Int. J. Environ. Res. Public Health 19 (2022) 6125.
[35] G. Mühlbach, M. Gasca, Multivariate polynomial interpolation under projectivities part I: Lagrange and newton interpolation formulas, Numer. Algorithms. 1 (1991) 375–399.
[36] C. Accadia, S. Mariani, M. Casaioli, et al., Sensitivity of precipitation forecast skill scores to bilinear interpolation and a simple nearest-neighbor average method on high-resolution verification grids, Weather Forecast. 18 (2003) 918–932.
[37] C.-C. Wang, C.-H. Hsu, C.-C. Lee, et al., A ROM-less DDFS based on a parabolic polynomial interpolation method with an offset, J. Signal. Process. Syst. 64 (2011) 351–359.
[38] R. Van Beeumen, W. Michiels, K Meerbergen, Linearization of Lagrange and Hermite interpolating matrix polynomials, IMA J. Numer. Anal. 35 (2014) 909–930.
[39] X. Shi, Z. Chen, H. Wang, et al., Convolutional LSTM network: A machine learning approach for precipitation nowcasting, MIT Press (2015). https://doi.org/10.1007/978-3-319-21233-3_6.
[40] Z.H. Hu, J.B. Zhou, K.J. Huang, et al., A data-driven approach for traffic crash prediction: A case study in Ningbo, China, Int. J. Intell. Transport. Syst. Res. 20 (2022) 508–518.
[41] S.T. Mubarrat, S. Chowdhury, Convolutional LSTM: A deep learning approach to predict shoulder joint reaction forces, Comput. Methods Biomech. Biomed. Engin. 26 (2023) 65–77.
[42] J. Huang, Y. Zhao, W. Yan, et al., Multi-source dynamic ensemble prediction of infectious disease and application in COVID-19 case, J. Thorac. Dis. 15 (2023) 4040–4052.
[43] L. Zhang, J. Huang, W. Yan, et al., Global prediction for mpox epidemic, Environ. Res. 243 (2024) 117748.
[44] X. Zhang, W. Zhang, S. Chen, Shanghai's life-saving efforts against the current omicron wave of the COVID-19 pandemic, Lancet 399 (2022) 2011–2012.
[45] Z. Chen, X. Deng, L. Fang, et al., Epidemiological characteristics and transmission dynamics of the outbreak caused by the SARS-CoV-2 Omicron variant in Shanghai, China: A descriptive study, medRxiv (2022). https://doi.org/10.1016/j.lanwpc.2022.100592.
[46] X. Chen, X. Yan, K. Sun, et al., Estimation of disease burden and clinical severity of COVID-19 caused by Omicron BA.2 in Shanghai, February-June 2022, medRxiv (2022). https://doi.org/10.1080/22221751.2022.2128435.
[47] T.L. Thein, Y.S. Leo, V.J. Lee, et al., Validation of probability equation and decision tree in predicting subsequent dengue hemorrhagic fever in adult dengue inpatients in Singapore, Am. J. Trop. Med. Hyg. 85 (2011) 942–945.
[48] V.J. Lee, D.C. Lye, Y. Sun, et al., Decision tree algorithm in deciding hospitalization for adult patients with dengue haemorrhagic fever in Singapore, Trop. Med. Int. Health 14 (2009) 1154–1159.

**Bin Chen** (BRID: 08567.00.50608) is a professor and doctoral supervisor in College of Atmospheric Sciences Lanzhou University. He has long been dedicating to the study of satellite remote sensing of atmospheric aerosols, atmospheric pollutants, and their environmental and climatic effects; application of machine learning and deep learning in atmospheric science; epidemic macro data and machine learning prediction and impact studies.