RESEARCH

Study on the impact of meteorological factors on influenza in different periods and prediction based on artificial intelligence RF-Bi-LSTM algorithm: to compare the COVID-19 period with the non-COVID-19 period

Hansong Zhu[1,8*†], Si Chen[2,9,10†], Weixia Qin[3†], Joldosh Aynur[4†], Yuyan Chen[5†], Xiaoying Wang[4†], Kaizhi Chen[6], Zhonghang Xie[1,8], Lingfang Li[1], Yu Liu[7*], Guangmin Chen[1,8*], Jianming Ou[1,8*] and Kuicheng Zheng[1,8*]

Abstract

Objective At different times, public health faces various challenges and the degree of intervention measures varies. The research on the impact and prediction of meteorology factors on influenza is increasing gradually, however, there is currently no evidence on whether its research results are affected by different periods. This study aims to provide limited evidence to reveal this issue.

Methods Daily data on influencing factors and influenza in Xiamen were divided into three parts: overall period (phase AB), non-COVID-19 epidemic period (phase A), and COVID-19 epidemic period (phase B). The association between influencing factors and influenza was analysed using generalized additive models (GAMs). The excess risk (ER) was used to represent the percentage change in influenza as the interquartile interval (IQR) of meteorology factors increases. The 7-day average daily influenza cases were predicted using the combination of bi-directional long

†Hansong Zhu, Si Chen, Weixia Qin, Joldosh Aynur, Yuyan Chen, and Xiaoying Wang contributed equally to this work.

*Correspondence:
Hansong Zhu
hszhu33@126.com
Yu Liu
ly8923028@126.com
Guangmin Chen
231147698@qq.com
Jianming Ou
ojmfj@vip.sina.com
Kuicheng Zheng
kingdadi9909@126.com
Full list of author information is available at the end of the article

Background

Influenza is a major global public health problem. In early 2019, a publication from the Global Burden of Disease Study (GBD) estimated a range of 99,000 to 200,000 annual deaths from lower respiratory tract infections directly attributable to influenza [1–3]. Although the public health and social measures (PHSM) taken to curb the spread of coronavirus disease 2019 (COVID-19) after its outbreak led to a sharp decline in the number of global influenza cases from 2020 to 2021, the level of influenza activity has rebounded significantly since then, and the number of reported influenza cases in southern

short memory (Bi-LSTM) and random forest (RF) through multi-step rolling input of the daily multifactor values of the previous 7-day.

Results In periods A and AB, air temperature below 22 ℃ was a risk factor for influenza. However, in phase B, temperature showed a U-shaped effect on it. Relative humidity had a more significant cumulative effect on influenza in phase AB than in phase A (peak: accumulate 14d, AB: ER = 281.54, 95% CI = 245.47 ~ 321.37; A: ER = 120.48, 95% CI = 100.37 ~ 142.60). Compared to other age groups, children aged 4–12 were more affected by pressure, precipitation, sunshine, and day light, while those aged ≥ 13 were more affected by the accumulation of humidity over multiple days. The accuracy of predicting influenza was highest in phase A and lowest in phase B.

Conclusions The varying degrees of intervention measures adopted during different phases led to significant differences in the impact of meteorology factors on influenza and in the influenza prediction. In association studies of respiratory infectious diseases, especially influenza, and environmental factors, it is advisable to exclude periods with more external interventions to reduce interference with environmental factors and influenza related research, or to refine the model to accommodate the alterations brought about by intervention measures. In addition, the RF-Bi-LSTM model has good predictive performance for influenza.

Keywords COVID-19, Influenza, Meteorological, Bi-LSTM, Random forest (RF)

provinces of China increased abnormally in the summer of 2022, and the intensity of influenza activity in China in spring and winter of 2023 was also higher than that in the natural epidemic years before the COVID-19 pandemic. [4–8].

Meteorology is recognized as an important influencing factor of influenza. Earlier studies have confirmed that meteorology factors are associated with influenza activity [3, 9–12]. However, different research results had both consistency and conflict. Park et al. 's study showed that low temperature would increase the prevalence of influenza [13], while Wang et al. reported that high temperature would also increase the risk of influenza [14]. Studies have shown that influenza A (H3N2) transmissibility was observed to be positively correlated with absolute humidity when absolute humidity was greater than 19g/m³ [15], but Zhu et al. 's study showed that both high humidity and low humidity would increase the risk of influenza [3]. Gomez Barroso et al. reported a positive correlation between precipitation and influenza [10], however, Soebiyanto et al. reported that the association between influenza and precipitation was location dependent [16]. The differences in research reports may be due to climate heterogeneity [17] or the use of different data types (e.g., daily data, weekly data) and analysis model schemes [17]. However, there is currently no evidence to suggest whether research on the impact of influencing factors on influenza and prediction is influenced by different periods.

The 2023 Report of the *Lancet* Countdown on Health and Climate Change points out that the correlation between climate action and health still needs to be improved, and suggests the establishment of a health oriented meteorological risk early warning system [18]. Thus, the in-depth study of the relationship between meteorology and infectious diseases is becoming increasingly important and urgent.

The powerful and effective PHSM had led to a sharp decline in the global influenza cases during the COVID-19 epidemic, leading to an influenza epidemic that did not conform to the natural law. Although there may be some extreme weather during this period, it often only presents a phased pattern [19, 20]. Hence, the analysis of the mode of action of meteorology on influenza during this period may have been distorted, however, there is currently no evidence of the degree of impact. Therefore, this study aims to provide limited evidence, using excess risk (ER) and predictive evaluation indicators to reveal the degree of this impact based on the same city and method, in order to select the research object period more appropriately. This is also an important issue that scholars will face in collecting data when studying the relationship between influencing factors and respiratory infectious diseases and prediction in the future.

Influenza is not only affected by meteorology, but also related to other potential environmental and demographic factors. Hoogeveen et al.'s study showed that there was a highly significant inverse correlation between pollen and flu-like incidence [21]. Day light is understood to regulate melatonin levels, and subsequently circadian immunity [22, 23]. The influence transmission rates of respiratory viruses might through the frequency and type of social contacts (e.g. holidays, school periods, international traveling, etc) [22, 24]. Annual influenza vaccination is an effective way to prevent influenza and can reduce the risk of influenza [25]. Therefore, we have analyzed the influencing factors including seasonal pollen allergens, day light, population mobility, and vaccination in this study, in order to have a more complete correlation analysis and more accurate prediction.

Random forest is a widely used method for data prediction and classification calculation, which can effectively predict the onset of influenza. LSTM is an artificial intelligence deep learning algorithm that is suitable for time series data analysis [3]. To overcome the limitations of LSTM cell which is able to work on previous content but cannot use the future one, Schuster and Paliwal proposed bidirectional recurrent neural networks [26], which is better than LSTM in predicting COVID-19 [27]. This study aims to integrate the advantages of RF pre-processing and Bi-LSTM for accurate prediction to construct an influenza prediction model, namely RF-Bi-LSTM, which has not been reported yet.

Xiamen, located in southeast China, has a land area of only 1,698.78 square kilometres [28] (Fig. 1). Therefore, monitoring by meteorological stations reflect the real situation of the city more accurately and comprehensively. Xiamen is an island with a high urbanization rate (90.10%) [28], confounding factors are reduced and the study of the impact of meteorology factors on influenza is facilitated.

Materials and methods
Data sources
The influenza data and resident demographic data of Xiamen from January 1, 2010, to March 31, 2022, were from the China Disease Prevention and Control Information System. The population of influenza was stratified by sex (male and female) and age (0 ~ 3 years, 4 ~ 12 years, and ≥ 13 years old), of which the age-stratified was divided according to the epidemiological characteristics of influenza in Xiamen. Due to the high urbanization rate (90.10%) of Xiamen, the population of influenza had not been stratified into urban and rural by areas in this study. The annual influenza vaccination data was downloaded from the Fujian Provincial Immunization Planning Information Management System.

The data of meteorology factor, seasonal allergens pollen (abbreviated as allergen), and day light (h) were provided by the Fujian Climate Center. The meteorology factors in this study included 8 indicators, of which air pressure (hPa), relative humidity (%), air temperature (℃), and wind speed (m/s) were the average values of 24 h a day, abbreviated as pressure, humidity, temperature, and wind respectively. Precipitation (mm) and sunshine duration (abbreviated as sunshine, h) were the daily cumulative value, air pressure difference (abbreviated as pres-difference, hPa) and daily temperature difference (abbreviated as temp-difference, ℃) were the difference between the maximum and minimum values of daily. The allergenic period was calculated based on the annual flower season from March to May in Xiamen. Day light was calculated based on the duration between sunrise and sunset on each day in Xiamen.

Due to the fact that the influenza cases among children and adolescents aged 0–12 in Xiamen accounted for about 85% of the total population, and other family members often travel and gather based on their children's holidays, we used student holidays as a population mobility indicator in this study. The holiday time obtained from Xiamen Education Bureau (https://edu.xm.gov.cn/jyfw/
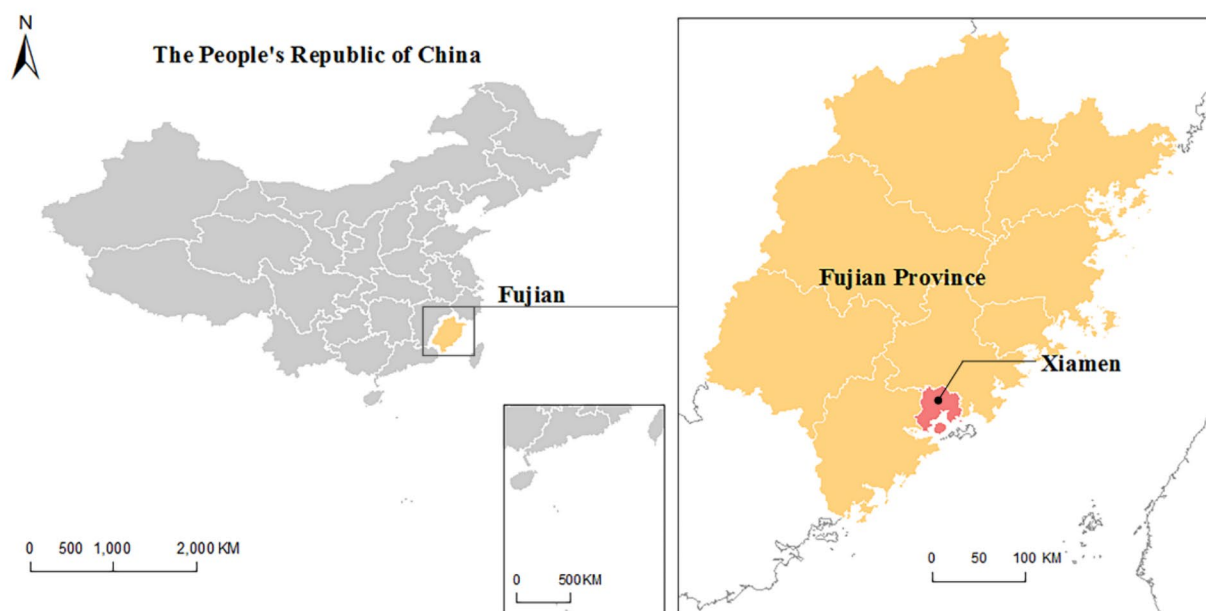

Fig. 1 The geographical location of Xiamen city

ndxl/). Allergen and holiday were dummy variables with values of 0 or 1 in this study.

Because the first case of COVID-19 in Fujian Province was reported on January 22, 2020, we divided the data on influenza and meteorology factors into three parts in this study: the overall year (January 1, 2010, to December 31, 2021, hereinafter referred to as phase AB), the non-COVID-19 epidemic period (January 1, 2010, to January 21, 2020, hereinafter referred to as phase A) and the COVID-19 epidemic period (January 22, 2020, to December 31, 2021, hereinafter referred to as phase B).

Statistical data analysis

The map in Fig. 1 was drawn using ArcGIS software (version 10.3, ESRI, Redlands, CA, USA).

R software (version 4.2.2, R Foundation for Statistical Computing, Vienna, Austria) was used to statistically analyse and visualize data on influenza and influencing factors. The effects of influencing factors on influenza were analysed using Spearman and generalized additive models (GAMs). The differences between subgroups of the population, the differences in lagged cumulative effects between groups in three phases, and the differences in evaluation indicators for predictive effects between groups in three phases were anslysed. When the econometric data exhibited a normal distribution and homogeneity of variance, the independent sample t-test and one-way analysis of variance were employed (the statistical measures for measuring the differences between samples are t and F values, respectively). In instances where the data displayed a skewed distribution and failed to demonstrate homogeneity of variance, nonparametric rank sum tests, such as the Kruskal-Wallis test and Mann-Whitney test, were utilized for statistical analysis (the statistical measures for measuring the differences between samples are H and Z values, respectively). Differences were considered statistically significant at $P < 0.05$. The GAM is an extension of the generalized linear model, a free and flexible statistical model, which can be used to detect the impact of nonlinear regression [29]. The GAM can fit influencing factors and unknown confounding factors with parametric and nonparametric methods, respectively, control for holidays and other confounding factors through a smoothing function, and estimate the risk degree on the premise of removing confounding factors [30]. The GAM formula is as follows [31]:

$$Log\left[E\left(Yi\right)\right] = \beta\rho i + NS\left(t, df\right) + s\left(Zj, df\right) + Dow + \alpha \quad (1)$$

Yi is the actual number of influenza cases on the i-th day. $E(Yi)$ is the expected number of influenza cases on the i-th day. β is the exposure response coefficient, which refers to the increase in influenza cases caused by each

increase of 1 unit in influencing factors. ρi is the meteorological factor on the i-th day. *NS (...)* is a natural spline function (used to control for seasonal and long-term trends, the day of the week effect and other influencing factors). *t* is a date variable. *df* is the degrees of freedom. *s* is a spline function. *Zj* is other influencing factors related to the influencing factors studied (used to control the mixing effect of other influencing factors). *Dow* is the dummy variable for the effect of the day of the week (controlling for the day of the week effect), and *α* is the intercept term.

According to the akaike information criterion (AIC) minimum principle, *df* was determined to be 3.

The impact of influencing factors on influenza were estimated as the relative risk (RR) associated with per interquartile range (IQR) increase of influencing factors values. The RR formula is as follows:

$$RR = e^{\beta \times \mathrm{IQR}} \quad (2)$$

β is the partial effect.

The ER and its 95% confidence interval (95% CI) of influenza associated with per impact of influencing factors values were indicated as the percentage change of influenza and its 95% CI with per IQR increase in influencing factors values. The formula of ER and its 95% CI are as follows:

$$ER = (RR - 1) \times 100\% \quad (3)$$

$$95\%CI = e^{(\beta \pm 1.96 \times SE) \times IQR} - 1 \quad (4)$$

SE is the standard error.

TensorFlow 2.8.0 software (Google Brain Team, Mountain View, CA, USA) and Python 3.8.13 software (Python Software Foundation, Delaware, USA) were used to predict the cases of influenza through RF-Bi-LSTMmodel predicted combined with influencing factors.

RF is an algorithm that integrates multiple trees through the idea of integrated learning. Its basic unit is the decision tree, and its essence is a major branch of machine learning [32]. The operation process includes five steps. The first step is feature splitting, which is used to split data and build supervised learning data. Then, random sampling is performed, and N samples are obtained by randomly sampling N times from the original dataset and placing them back. Third, a decision tree is constructed and trained for each sampled sample dataset. The fourth step is decision-making, in which each tree makes its own decisions based on the data. Finally, there is decision aggregation, with the average value of the tree predicted as the final result [33]. The RF process is shown in Fig. 2.

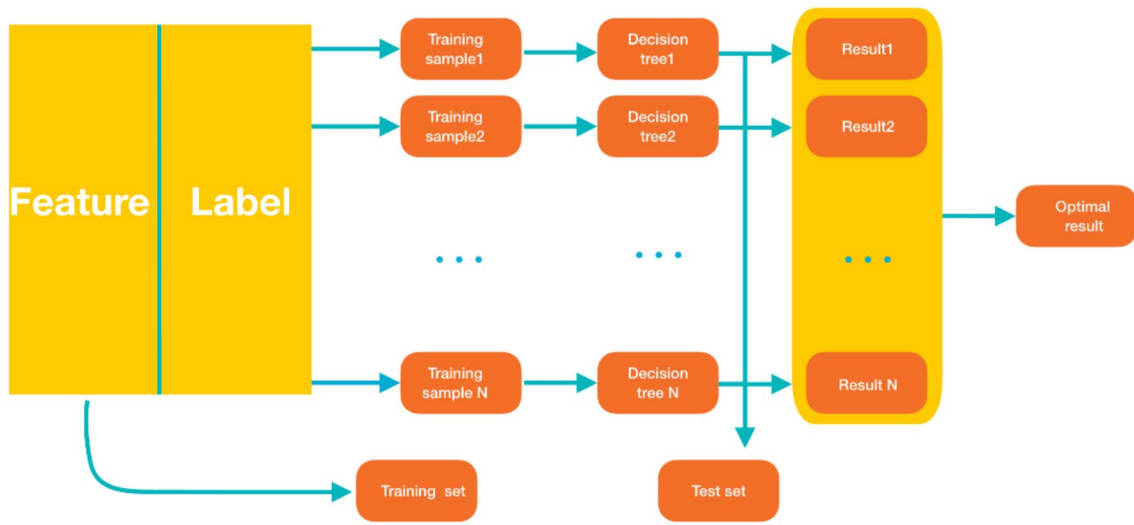The result of the final decision tree are as follows.

Fig. 2 A brief operation process of the RF in this study

$$\mathrm{D} = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\} \quad (5)$$

$$h_i(x) = \quad (\mathrm{D}, D_{bs}) \quad (6)$$

$$\mathrm{H}(x) = \frac{1}{T} \sum_{t=1}^{T} h_i(x) \quad (7)$$

(Note: *D*: all data; *x*: feature; *y*: label; $h_i(x)$: the output of each base model; ζ: base learning algorithm; D_{bs}: Sample set generated by self-service sampling.)

The core of LSTM concepts are the cell state and "gate" structure. The whole process is mainly divided into three parts.

The first step is to determine what information to discard from the neuron state through the sigmoid layer of the "forgetting gate". The sigmoid layer outputs a value (between -1 and 1) to determine how much information each part can pass (0 means completely discarded, 1 means completely reserved).

The second step is to determine what new information is stored in the neuron state. First, the sigmoid layer of the "input gate" determines which value will be updated. Second, the tanh layer creates a new candidate value vector *t* (between -1 and 1), adds it to the state and multiplies the value of the sigmoid function to update the old neuron state. Third, the output determines the information to be output.

Finally, the "output gate" determines the output information. First, which part of the output neuron state passes through the sigmoid layer is determined. Second, the neuron state is processed by tanh (between -1 and 1) and multiplied by the output of the sigmoid gate. Third, the determined part is output.

The LSTM calculation formulas are as follows:

$$f_t = \sigma(W_f [h_{t-1}, x_t] + b_f) \quad (8)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (9)$$

$$\tilde{C}_t = tanh(W_c \cdot [h_{t-1}, x_t] + b_c \quad (10)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (11)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (12)$$

$$h_t = o_t \cdot tanh(C_t) \quad (13)$$

(Note: h_{t-1}: the output of the previous neuron state; X_t: represents the input of the current neuron state; σ: the sigmoid function; C_{t-1} is updated to C_t.)

Bi-LSTM is comprised of two distinct LSTM hidden layers with similar output in opposite directions [27]. With this architecture, previous and future information is exploited in output layer [32]. An input sequence $X = (X_1, X_2, ., X_n)$ in Bi-LSTM is calculated in forward direction as $\vec{h} = (\vec{h}_1, \vec{h}_2, ..., \vec{h}_n)$ and backward directions as $\overleftarrow{h} = (\overleftarrow{h}_1, \overleftarrow{h}_2, ..., \overleftarrow{h}_n)$. The final out of this cell y_t is formed by both $\vec{h}$ and $\overleftarrow{h}$, the final sequence of out looks like $Y = (Y_1, Y_2, . Y_t, Y_n)$ [30]. The single cell of LSTM and Bi-LSTM are displayed in Fig. 3.

(Note: h_{t-1}: the output of the previous neuron state; X_t: represents the input of the current neuron state; σ: the sigmoid function; C_{t-1} is updated to C_t).

RF-Bi-LSTM was used for influenza prediction in this study, which can fully utilize the advantages of RF and Bi-LSTM models to improve the accuracy and stability of time series prediction. The four operational steps were as follows:

Data preparation: Time series data, which included time, climate data and influenza incidence data, were
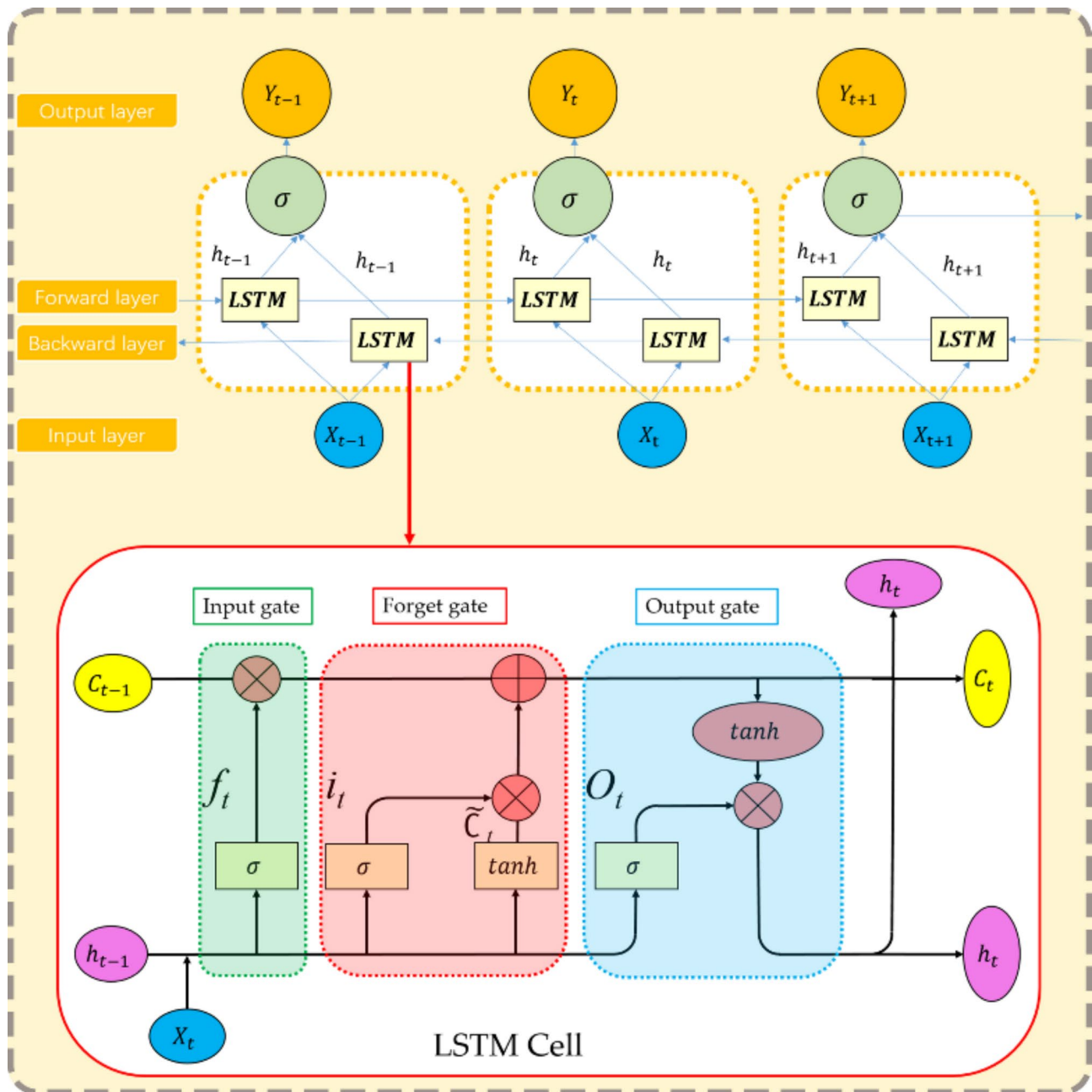
Fig. 3 The single cell of LSTM and Bi-LSTM

prepared into a format suitable for model training; that is, the data were divided into three input sequences (AB phase: January 1, 2010, to December 31, 2021; A phase: January 1, 2010, to January 21, 2020; B phase: January 22, 2020, to December 31, 2021) and three sequential target sequences (the three months following the input sequence, AB: January 1 to March 31, 2022; A: January 22 to April 21, 2021; B: January 1 to March 31, 2022). The input sequence was the historical time step data used for prediction, while the target sequence was the future time step data after the input sequence.

RF feature extraction: The RF model was used to extract features from the input sequence. Through the RF training process, the importance of influencing factors that affected the number of influenza outbreaks was ranked, redundant information was eliminated, and the resulting model was subsequently used as a new feature combination input model.

Bi-LSTM model training: The Bi-LSTM model was trained using input sequences.

Prediction: The influenza cases in the three target sequences were predicted using the trained Bi-LSTM

model. Based on our previous experience, predicting the daily average number of cases for the next 7 days by inputting multi factor values for 7 days was the best method of prediction [34]. Thus, the method involved inputting the multifactor value of 7 days to predict the daily average cases of influenza in the next 7 days, and the prediction was realized through multistep rolling in this study.

Figure 4 shows the RF-Bi-LSTM process.

Ten indicators reflecting different characteristics were used to comprehensively evaluate the prediction performance. The root mean square error (RMSE) is widely used to measure the deviation between predicted values and actual values, while the mean absolute error (MAE) is the average of absolute errors, which can better reflect the actual situation of prediction errors. The mean absolute percentage error (MAPE) has a very intuitive explanation for relative error, but it is not suitable for prediction models with large expected errors, while the symmetric mean absolute percentage error (SMAPE) can correct this drawback of MAPE [35]. The RMSE-observations standard deviation ratio (RSR) reflects the root mean square error to observation's standard deviation ratio [36]. The correlation coefficient (CC) is used to evaluate the strength and direction of the linear relationship between the predicted results of the model and the actual observed values, thereby helping to assess the predictive performance of the model. The Nash-Sutcliffe efficiency (NSE) normalizes the relative magnitude of the residual variance between the predicted values and actual values, indicating how well a plot of the two data values fits along the 1:1 line [37], mainly used to evaluate the overall fit of the model. The Kling-Gupta efficiency (KGE) takes into account multiple aspects of performance, including the mean, standard deviation, and correlation of research data, and provides a good reflection of the model's performance under different data conditions. The Willmott's index of agreement (IA) is a descriptive measure, and it is both a relative and bounded measure which can be widely applied in order to make cross-comparisons between models [38]. A lower bound of zero for MAE, RMSE, and MAPE means a perfect fit, but for models with poor performance, the values gradually increase infinitely, as these values largely depend on the range of the descriptive variables, making them incomparable to each other within the same metric [35]. RSR also changes from the optimal value of zero to positive infinity [37]. However, the Legate and McCabe's Index (LMI) can robustly address the predictive limitations and it ranges between 0 and 1, where 1 is an ideal value [39]. Same as it, a model with CC=1,
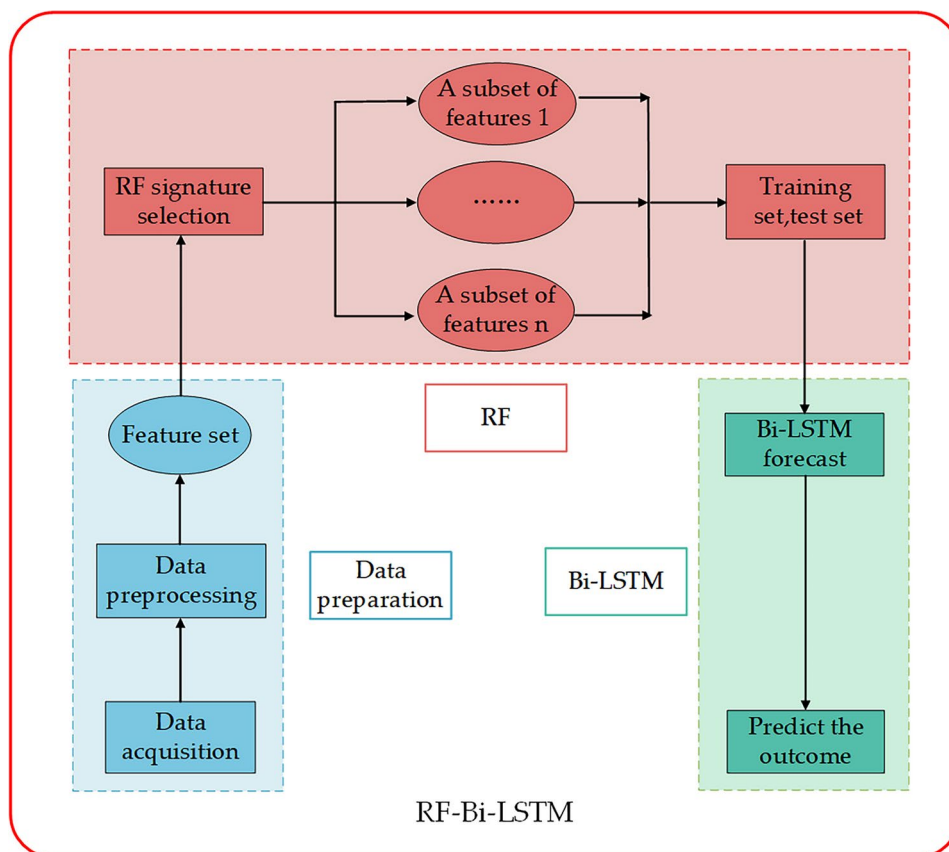

Fig. 4 The RF-Bi-LSTM process

Zhu *et al. BMC Infectious Diseases*   (2024) 24:878

Page 8 of 22

NSE=1, KGE=1, and IA=1 is a great model [36, 39, 40]. In addition, SMAPE values are bounded, with the lower bound 0% implying a perfect fit, and the upper bound 200% reached when all the predictions and the actual target values are of opposite sign [35]. The calculation formulas of the ten evaluation indicators are as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum \mathop{}_{i=1}^{n} (P_i - X_i)^2} \quad (14)$$

$$MAE = \frac{\sum_{i=1}^{n} \left| X_i - \hat{X}_i \right|}{n} \quad (15)$$

$$MAPE = \frac{100\%}{n} \sum \mathop{}_{i=1}^{n} \left| \frac{P_i - X_i}{P_i} \right| \quad (16)$$

$$SMAPE = \frac{100\%}{n} \sum \mathop{}_{i=1}^{n} \left| \frac{|P_i - X_i|}{\frac{|P_i| + |X_i|}{2}} \right| \quad (17)$$

$$RSR = \frac{RMSE}{P} \quad (18)$$

$$CC = \frac{\sum_{i=1}^{n} \left(P_i - \bar{P} \right) \left(X_i - \bar{X} \right)}{\sqrt{\sum_{i=1}^{n} \left(P_i - \bar{P} \right)^2} \sqrt{\sum_{i=1}^{n} \left(X_i - \bar{X} \right)^2}} \quad (19)$$

$$NSE = 1 - \frac{\sum_{i=1}^{n} (P_i - X_i)^2}{\sum_{i=1}^{n} \left(P_i - \bar{P} \right)^2} \quad (20)$$

$$KGE = 1 - \sqrt{(r-1)^2 + (\quad -1)^2 + (\quad -1)^2} \quad (21)$$

$$IA = 1 - \frac{\sum_{i=1}^{n} (P_i - X_i)^2}{\sum_{i=1}^{n} \left(\left| X_i - \bar{P} \right| + \left| P_i - \bar{P} \right| \right)^2} \quad (22)$$

$$LMI = 1 - \frac{\sum_{i=1}^{n} |P_i - X_i|}{\sum_{i=1}^{n} \left| P_i - \bar{P} \right|} \quad (23)$$

(Note: P_i: the observed daily incidence of influenza cases on day i; X_i: the predicted daily incidence of influenza cases on day i, where $i=1\ldots, n$; σ_P: the standard deviation of the observed values; r: the correlation coefficient; α: the ratio of the standard deviations; β: the ratio of the means).

Results

Descriptive statistics

The numbers of influenza cases in phases AB, A, and B were 21,324, 19,431, and 1893, with an average of 4.87, 5.45, and 2.67 daily cases and a maximum of 227, 227, and 133 cases, respectively. The minimum values of humidity (35.00%) and temperature (6.60 ℃) in phase B were significantly higher than those in phases AB and A (23.00% and 3.90 ℃), while the maximum values of pres-difference (10.50 hPa) and precipitation (123.80 mm) were significantly lower than those in phases AB and A (39.7 hPa and 172.70 mm). Figure 5 shows more descriptive statistics on influenza and meteorology factors.

From 2010 to 2021, Xiamen reported 9 influenza cluster incidents (a total of 721 cases), including 1 (56 cases), 2 (139 cases), 4 (358 cases), and 2 (168 cases) in 2013, 2017, 2018, and 2019, respectively. One incident occurred in junior high schools, while 8 incidents occurred in primary schools.

There were significant differences between the sex and age groups in the influenza-affected population (*P*<0.001). The descriptive statistics for the daily-based cases of influenza are presented in Table 1.

Influenza fluctuations in phase A had a specific seasonal periodicity. The number of influenza cases reported in the first ten days of phase B was high, and the seasonal periodicity in other periods was not significant. The fluctuations in influencing factors in both periods A and B were cyclical. The detailed trends are presented in Fig. 6.

Correlation analysis

The connecting line on the right side of Fig. 7 shows that the holiday and allergen in AB, A, and B phases were not significant correlated with influenza (-0.05<*r*<0.05 and *P*<0.05), while pressure, humidity, temperature, and day light were significantly correlated with influenza.

The heatmap on the lower left side of Fig. 7 shows a strong negative correlation between pressure and temperature (*r*=-0.853, *P*<0.01), as well as between precipitation and sunshine (*r*=-0.564, *P*<0.01). There was a strong positive correlation between temp-difference and sunshine (*r*=0.651, *P*<0.01), as well as between humidity and precipitation (*r*=0.592, *P*<0.01). Day light was positively correlated with temperature (*r*=0.72, *P*<0.01) and negatively correlated with pressure (*r*=-0.78, *P*<0.01).

The detailed correlations between influencing factors and influenza, as well as between influencing factors, are presented in Fig. 7.

Vaccine coverage rates in 2020 (2.39%), 2021 (1.51%) and 2019 (0.88%) ranked top three, while influenza incidence rates were one year earlier than those, respectively, and in 2019 (2.05‰), 2020 (1.10‰) and 2018 (1.00‰) ranked top three. There was no significant correlation (*r*=0.41, *P*=0.18) between the vaccination rate and the
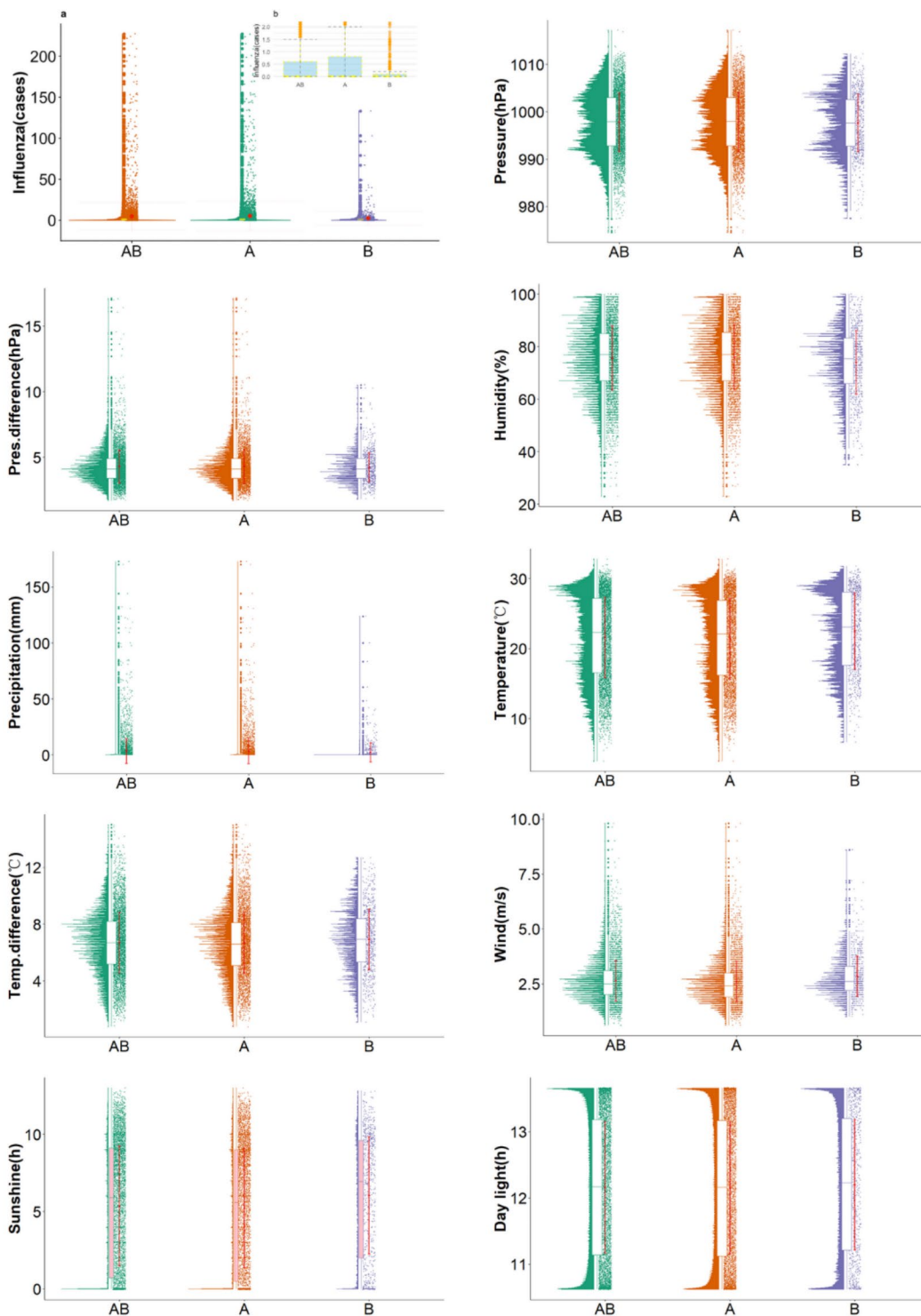
Fig. 5 Violin cloud and rain box chart of influenza and meteorology factors (Note: The influenza in this figure include two images, a and b, where a is the main image and b is an enlargement of the box diagram. Since the holiday and allergen in this study are dummy variables, this figure is not shown)

Table 1 Descriptive statistics for the daily-based cases of influenza

Variables	Cases	Constituent ratio (%)	Min	P25	M50	P75	Max	Mean ± SD	t/F	p
Sex										
AB									24.92	0.00
Male	11,848	55.56	0.00	0.00	0.00	1.00	134.00	2.70 ± 9.86		
Female	9476	44.44	0.00	0.00	0.00	1.00	110.00	2.16 ± 8.35		
A									23.25	0.00
Male	10,766	55.41	0.00	0.00	0.00	1.00	134.00	2.93 ± 10.51		
Female	8665	44.59	0.00	0.00	0.00	1.00	110.00	2.36 ± 8.92		
B									10.62	0.00
Male	1082	57.16	0.00	0.00	0.00	1.00	71.00	1.52 ± 5.22		
Female	811	42.84	0.00	0.00	0.00	1.00	62.00	1.14 ± 4.18		
Ages(years)										
AB									25.96	0.00
0~3	6284	29.47	0.00	0.00	0.00	1.00	83.00	1.43 ± 10.51		
4~12	11,477	53.82	0.00	0.00	0.00	1.00	139.00	2.62 ± 8.92		
BBBB			0.000							
≥13	3563	16.71	0.00	0.00	0.00	1.00	48.00	0.81 ± 8.92		
A									24.14	0.00
0~3	5735	29.51	0.00	0.00	0.00	1.00	83.00	1.56 ± 5.40		
4~12	10,615	54.63	0.00	0.00	0.00	1.00	139.00	2.89 ± 11.64		
BBBB										
≥13	3081	15.86	0.00	0.00	0.00	1.00	44.00	0.84 ± 3.12		
B									11.93	0.00
0~3	549	29.00	0.00	0.00	0.00	1.00	43.00	0.77 ± 3.18		
4~12	862	45.54	0.00	0.00	0.00	1.00	51.00	1.21 ± 4.12		
BBBB										
≥13	482	25.46	0.00	0.00	0.00	1.00	48.00	0.684 ± 2.87		

Note: Min stands for minimum value, P25 stands for 25th percentile, P50 stands for 50th percentile and P75 stands for 75th percentile, Max stands for maximum value, SD stands for standard deviation
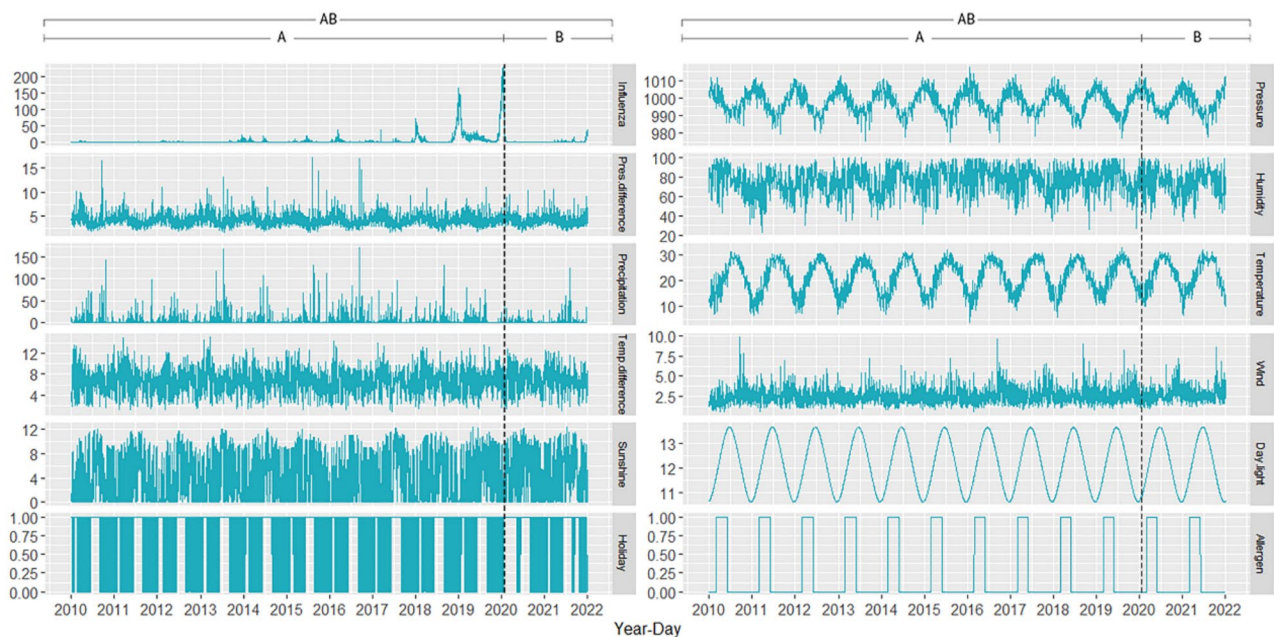

Fig. 6 Time series of influenza and influencing factors (Note: The units of influenza, pressure, pres-difference, humidity, precipitation, temperature, temp-difference, wind, sunshine and day light are case, hPa, hPa, %, mm, ℃, ℃, m/s, h and h, respectively. Holiday and allergen are dummy variables in this study.)
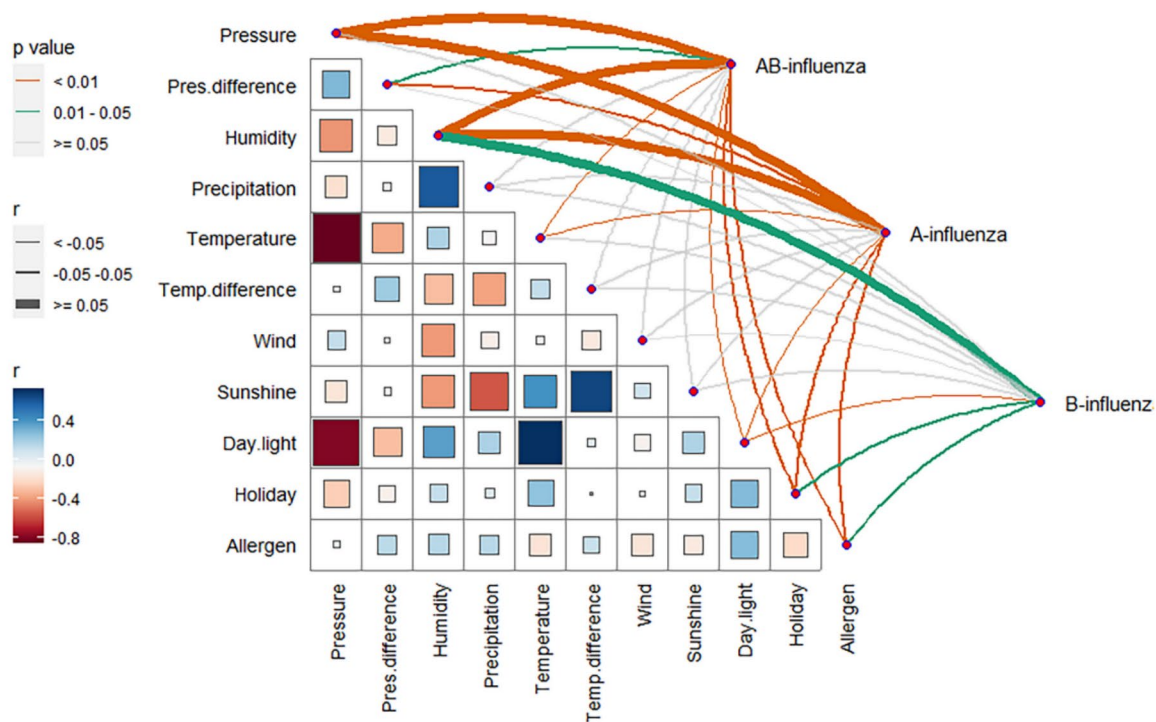
Fig. 7 Heatmap of Spearman correlation analysis of influenza and influencing factors

incidence rate of annual influenza from 2010 to 2021. More information of influenza incidence rate and vaccination rate are presented in Fig. 8.

The association between meteorology factors and influenza

In phase AB and phase A, pressure had a U-shaped impact on influenza, however, the value ranges were not entirely consistent (AB: < 991 hPa, > 1004 hPa; A: < 993 hPa, > 1006 hPa). Moreover, in phase B, the association between pressure and influenza presented a linear trend, which was positive when pressure was above 998 hPa.

In phase AB and phase A, the pres-difference was a risk factor for influenza when the pres-difference was below 4 hPa and gradually shifted from a risk effect to a protective effect as the pres-difference increased. Conversely, in phase B, the pres-difference shifted from a protective factor to a risk factor as it increased.

The relationship between humidity and influenza in the three phases increased monotonically, and all presented risk factors at high humidity levels (>72%). However, unlike the curves in the AB and A periods, phase B presented a linear pattern.

The relationship between precipitation and influenza in phase AB and phase A was somewhat similar, that is, it decreased with increasing precipitation, but in phase A, the precipitation was above 100 mm and

presented a gentle upward trend. However, phase B differed greatly from phase AB and phase A, and the relationship between precipitation and influenza increased monotonically.

The relationship between temperature and influenza in both phase AB and phase A was arcuate, and when temperature < 22 °C, it was a risk factor for influenza. However, the relationship between temperature and influenza in phase B was U-shaped.

The relationships between influenza and the temp-difference, wind, and sunshine during the three phases were similar. However, the relationship between the temp-difference and influenza in phases AB and A presented an arcuate shape, while the relationship in phase B decreased monotonically and linearly. The relationship curve between wind and influenza in phase A was not as obvious as the U-type curve in phases AB and B.

In phase AB and phase A, day light had a U-shaped impact on influenza, however, the value ranges were not entirely consistent (AB: < 11.2 h, > 13.3 h; A: < 11.1 h, > 13.2 h). Moreover, in phase B, the association between day light and influenza presented an almost linear trend, which was negative when day light was above 11.6 h.

Additional characteristics of the impact of meteorology factors on influenza in the three phases are presented in Fig. 9.
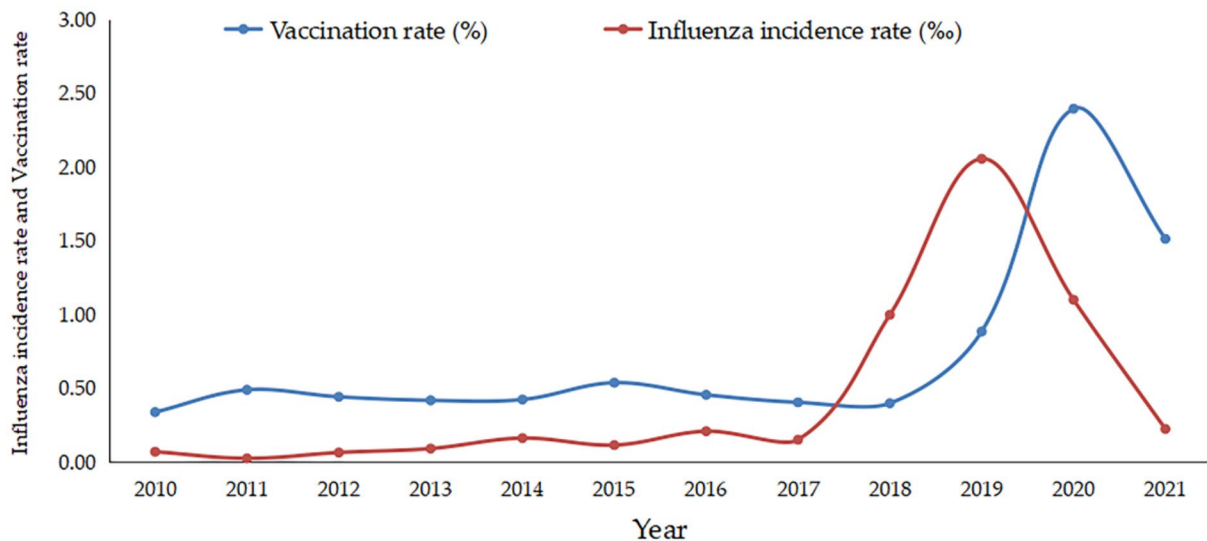
Fig. 8 Spearman correlation analysis of annual influenza incidence rate and vaccination rate

The lag and cumulative effects of the association between meteorology factors and influenza

In all three periods, pressure had a multi-day cumulative effect on influenza, and the more accumulated the number of days, the greater the excess risk (peak: accumulate 14d, AB: ER=228.48, 95% CI=184.82~278.84; A: ER=219.56, 95% CI=173.37~273.56; B: ER=162.14, 95% CI=84.17~273.12). For phase AB and phase A, there was a correlation between pressure and influenza in each lag period ($P<0.05$), while for phase B, there was no correlation between them in some lag days ($P>0.05$).

In phase AB and phase A, the relationship between pres-difference and influenza showed a W-shaped pattern with increasing lag time (low peak: lag 4d, AB: ER = -12.56, 95% CI = -16.34~-8.60; A: ER = -6.75, 95% CI = -11.29~-1.98), while in phase B, it showed the opposite characteristic (M-shaped) and a significant cumulative effect (peak: lag 14d, ER=8.96, 95% CI = -5.73~25.95; accumulate14d, ER=19.80, 95% CI = -18.17~75.40). There was a statistically significant difference in the cumulative lag effect among the three phases ($P<0.05$). In phase AB, the more cumulative the number of days, the lower the excess risk of pres-difference to influenza, while in phase A, the opposite was true.

In all three phases, the more lag days there were, the lower the excess risk of humidity to influenza (peak: lag 0d, AB: ER=141.08, 95% CI=123.53~160.02; A: ER=87.10, 95% CI=72.87~102.50; B: ER=101.61, 95% CI=39.93~190.48). Humidity had a more significant cumulative effect on influenza in phase AB than in phase A (peak: accumulate 14d, AB: ER=281.54, 95% CI=245.47~321.37; A: ER=120.48, 95% CI=100.37~142.60). There was a significant difference in

the impact of humidity on influenza between phase AB and phase A ($P=0.005$).

In phase B, there was no correlation between precipitation and influenza at each lag time ($P<0.05$), however, in phase AB and phase A, as the lag time increased, the correlation between precipitation and influenza showed an inverted U-shaped pattern (peak: lag 11d, AB: ER=0.28, 95% CI=0.07~0.48; A: ER=0.55, 95% CI=0.25~0.84). In phase A, precipitation had a significant cumulative effect on influenza (peak: accumulate 14d, ER=0.86, 95% CI = -0.05~1.80), however, in phase AB, its cumulative effect did not increase the risk of influenza.

In phase AB and phase A, for every one IQR unit increase in temperature and temp-difference, influenza cases showed a continuous downward trend with the increase of cumulative days (temperature low peak: accumulate 14d, AB: ER = -73.96, 95% CI = -77.15~-70.32; A: ER = -84.29, 95% CI = -86.35~-81.91; temp-difference low peak: accumulate 14d, AB: ER = -44.16, 95% CI= -49.16~-38.68, A: ER = -49.60, 95% CI = -54.66~-43.97). The trend in phase B was opposite to that in phase AB and phase A. In addition, the cumulative lag effect of temperature on influenza showed significant differences among the three phases ($P<0.05$). The impact of temperature in phase A on influenza was significantly higher than that in phase AB.

Compared with phase AB and phase A, wind had a more significant impact on influenza in phase B. The impact of wind on influenza in all three phases had a significant multi day cumulative effect. For every increase of one IQR unit in wind during each cumulative time, the influenza cases would decrease (low peak: AB: accumulate 7d, ER = -42.47, 95% CI = -47.10~-37.43; A:
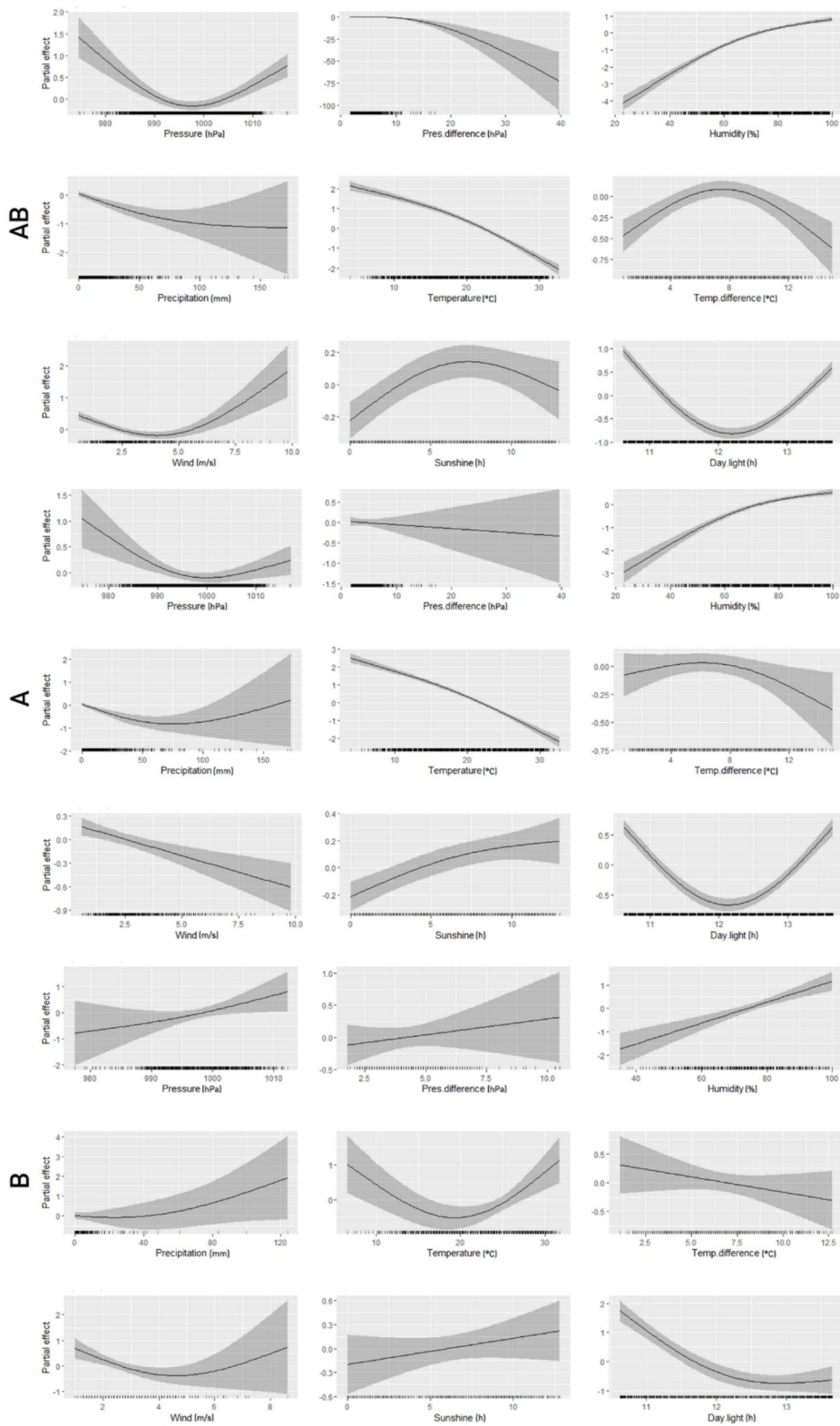
Fig. 9 Correlation between influencing factors and influenza based on GAM analysis

Table 2 IQR of various meteorology factors during different research phases

Phase	Pressure (hPa)	Pres-difference (hPa)	Humidity (%)	Precipitation (mm)	Temperature (℃)	Temp-difference (℃)	Wind (m/s)	Sunshine (h)	Day light (h)
AB	10.20	1.50	18.00	0.60	10.70	3.00	1.10	8.40	2.04
A	10.20	1.50	18.50	0.80	10.70	3.00	1.10	8.50	2.05
B	9.88	1.50	17.30	0.10	10.48	3.08	1.10	7.60	1.99

Table 3 Test for differences in ER values across three phases with different cumulative lag times

Variables	A-AB-B		A-AB		A-B		AB-B	
	H	P	Z	P	Z	P	Z	P
Pressure	11.84	**0.00**	-1.33	0.19	-2.34	**0.02**	-3.23	**0.00**
Pres-difference	24.72	**0.00**	-3.20	**0.00**	-2.56	**0.01**	-4.52	**0.00**
humidity	6.54	**0.04**	-2.78	**0.01**	-0.32	0.77	-1.61	0.11
Precipitation	2.91	0.23	-1.58	0.12	-1.01	0.32	-0.92	0.37
temperature	37.95	**0.00**	-3.30	**0.00**	-5.13	**0.00**	-4.71	**0.00**
Temp-difference	1.27	0.53	-1.04	0.31	-0.76	0.46	-0.76	0.46
Wind	5.54	0.06	-1.30	0.20	-2.15	**0.03**	-1.42	0.16
Sunshine	5.04	0.08	-0.48	0.65	-1.93	0.06	-1.90	0.06
Day light	26.91	**0.00**	-4.97	**0.00**	-2.28	**0.02**	-3.77	**0.00**

(*Note*: Non parametric Kruskal-Wallis test is used for differences between the three phases, and Mann-Whitney test is used for comparison between the two phases)

accumulate 7d, ER = -34.51, 95% CI = -39.68~-28.89; B: accumulate 14d, ER = -33.76, 95% CI = -48.58~-14.67).

In phase AB and phase A, the correlation between sunshine and influenza showed a W-shaped pattern with increasing lag time (low peak: lag 11d, AB: ER = -24.15, 95% CI = -29.88~-19.36; A: ER = -25.14, 95% CI = -30.55~-19.32), while in phase B, it showed a U-shaped pattern (low peak: lag 5d, B: ER= -28.89, 95% CI = -43.06~-11.19). The cumulative effect trend of phase B was opposite to that of phase AB and phase A.

In phase AB and phase A, for every one IQR unit increase in day light, influenza cases showed a continuous downward trend with the increase of cumulative days (low peak: accumulate 14d, AB: ER = -75.60, 95% CI = -78.51~-72.29; A: ER = -58.51, 95% CI = -64.25~-51.58). The trend in phase B was opposite to that in phase AB and phase A. In addition, the cumulative lag effect of day light on influenza showed significant differences among the three phases (*P*<0.05). The impact of day light in phase A on influenza was significantly higher than that in phase AB.

Overall, there were differences in the analysis results among these three phases.

Compared to other age groups, children aged 4–12 were more affected by pressure, precipitation, sunshine, and day light, while those aged ≥ 13 were more affected by the accumulation of humidity over multiple days.

In addition, three phases within the same age group also had different effects. In phase B, for every one IQR unit increase in pres-difference, the influenza cases in the children aged ≥ 4 increased, while in phase AB and phase A, the opposite is true. In all age groups, the impact of

temperature on influenza in phase B was smaller than that in phases AB and phase A.

There was no significant difference in the impact of meteorology factors on influenza between genders. However, in some lag times for females, the influenza cases of phase B increased for every increase of one IQR unit in temperature, while the opposite was true for males. And for every increase of one IQR unit in precipitation, as the cumulative number of days increased, influenza cases in phases AB and phase A also increased, while in phase B, the opposite trend was observed. The remaining detailed information is shown in the Tables 2 and 3; Fig. 10.

RF-Bi-LSTM forecasts

Figure 11 shows that the evaluation index values of using RF-Bi-LSTM algorithms to predict the AB, A, and B phases of influenza through meteorology factors were close to the lowest value 0 or to the highest value 1, indicating high prediction accuracy. The optimal values of the ten evaluation indicators RMSE, MAE, MAPE, SMAPE, RSR, CC, NSE, KGE, IA, and LMI were 1.05, 0.59, 0.08, 0.12, 0.12, 0.99, 0.98, 0.99, 0.88, and 0.95, respectively.

The predictive evaluation indicators RMSE, MAE, MAPE, SMAPE, and RSR values of phase A were lower than those of phase AB and phase B, while CC, NSE, KGE, IA, and LMI were closer to 1 (Excluding the influenza prediction KGE for male in phase A). The values of multiple predictive evaluation indicators in phase B were higher than those in phase AB and phase A or further away from 1, especially MAPE and SMAPE. Table 4 shows that compared with phase B, all ten evaluation indicators in phase A showed significant differences (*P*<0.05), while compared with phase AB, six indicators
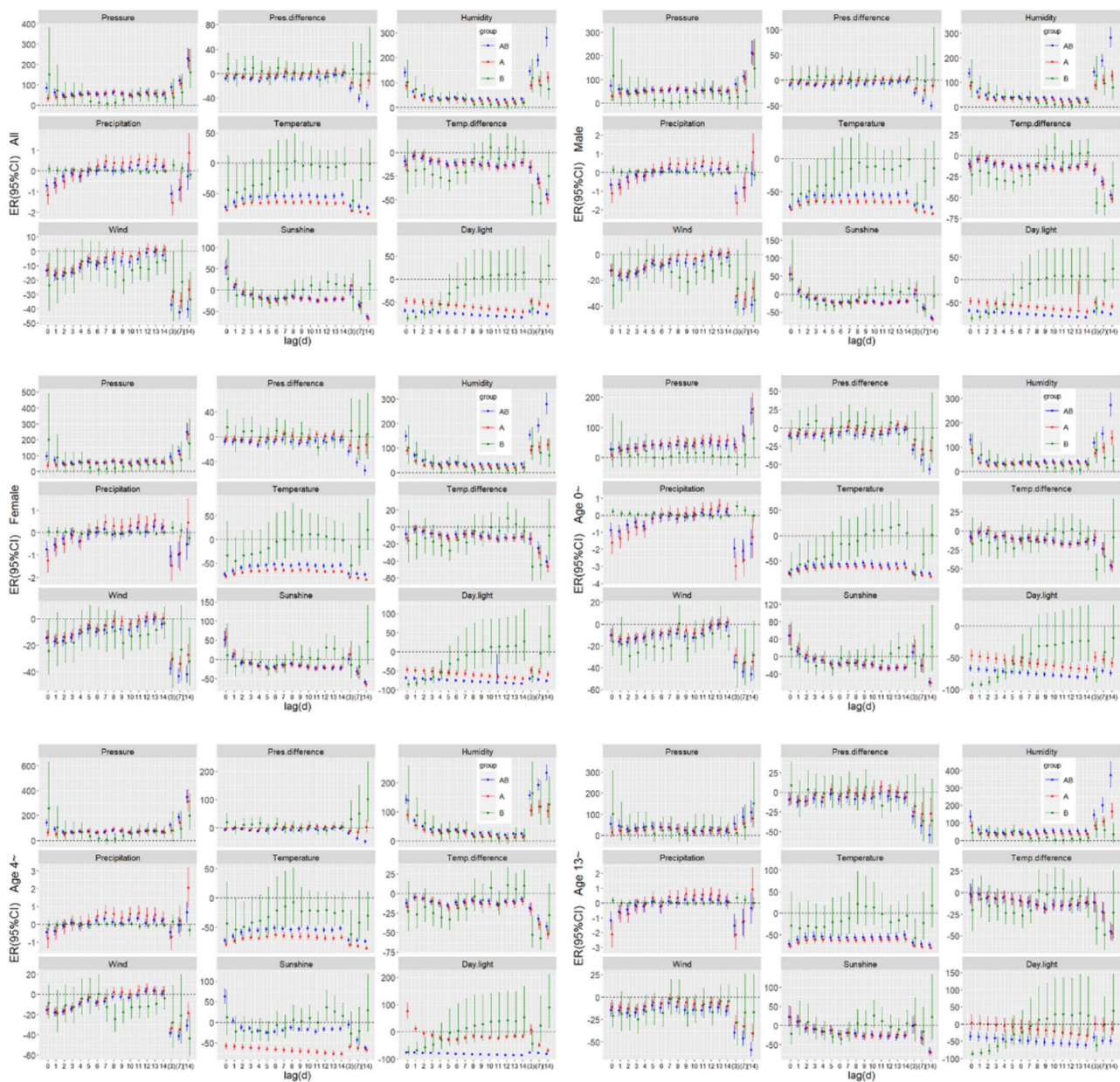
Fig. 10 The lag and cumulative effects of the association between meteorology factors and influenza based on GAM analysis (Note: lag0-14 represents a single day lag of 0-14d, (3) represents a cumulative 3-day lag over multiple days, (7) represents a cumulative 7-day lag over multiple days, and (14) represents a cumulative 14-day lag over multiple days)

showed significant differences. And it was also confirmed by the predicted and actual values in Fig. 12. Which indicates that phase A had the best prediction effect, and phase B had the worst one.

The prediction effect for the ≥13 age group in phase A was the best, with the lowest RMSE, MAE, MAPE, SMAPE, and RSR values of 1.05, 0.59, 0.08, 0.12, and 0.12, respectively, and the highest IA and LMI values of 0.99 and 0.88. The NSE (0.98) and KGE (0.95) values for females in phase A were the highest.

More detailed evaluation indicator values are shown in Fig. 11. More detailed actual and predicted influenza values are shown in Fig. 12.

Discussion

Although Fig. 6 shows that the fluctuations in meteorology factors, day light, holiday and allergen in both phases A and B were cyclical, and might affect seasonal influenza, Fig. 7 shows that not all of these factors were significantly correlated with influenza. The correlation between allergens and holidays with influenza is very
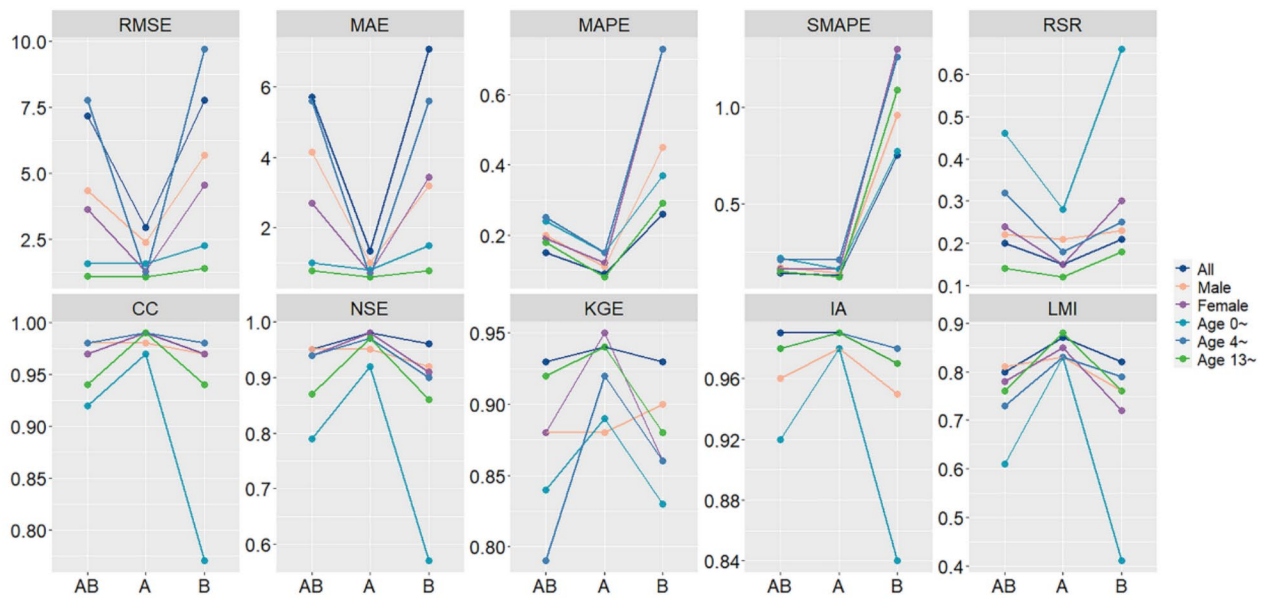
Fig. 11 Evaluation indicators based on RF-Bi-LSTM predictions

Table 4 Test for differences in evaluation indicator values across three phases with different groups of people

Evaluation indicator	A-AB-B		AB-A		AB-B		A-B	
	H	P	Z	P	Z	P	Z	P
RMSE	6.75	**0.03**	31.00	**0.04**	13.00	0.48	4.00	**0.03**
MAE	6.94	**0.03**	31.00	**0.04**	17.00	0.94	3.00	**0.02**
MAPE	14.06	**0.00**	33.00	**0.02**	0.00	**0.00**	0.00	**0.00**
SMAPE	12.19	**0.00**	26.00	0.22	0.00	**0.00**	0.00	**0.00**
RSR	5.19	0.07	28.00	0.13	14.50	0.63	4.50	**0.04**
CC	7.63	**0.02**	5.00	**0.04**	22.50	0.51	33.00	**0.02**
NSE	7.65	**0.02**	5.00	**0.04**	23.00	0.47	33.50	**0.02**
KGE	5.76	0.06	5.50	0.05	18.50	1.00	31.00	**0.04**
IA	7.86	**0.02**	7.00	0.07	24.00	0.36	34.00	**0.01**
LMI	11.51	**0.00**	0.00	**0.00**	20.00	0.81	36.00	**0.00**

(*Note*: Non parametric Kruskal-Wallis test is used for differences between the three phases, and Mann-Whitney test is used for comparison between the two phases)

weak (-0.05<*r*<0.05, *P*<0.05), indicating that they may not be important influencing factors for influenza in Xiamen. Therefore, we did not include them in subsequent GAM analysis and predictive studies.

Table 1 shows that the 0–12 age group accounts for nearly 85% of all influenza cases in Phase A. Out of the 9 influenza cluster incidents reported from 2010 to 2021 in Xiamen, 8 incidents occurred in primary schools and 1 incident occurred in junior high schools. Therefore, suggesting the mobility of children is crucial for studying the impact of human behavior on influenza. During school, students are in relatively closed classrooms and susceptible groups gather, making it easy for influenza to spread quickly. However, due to limited participation in extracurricular group activities, there are fewer opportunities for infection and introduction of influenza virus. During holidays, children usually go to public places to

participate in activities, play, shopping, etc., and are often accompanied by other family members, which increases the chance of regional influenza transmission. Although there is an increase in student mobility during holidays, the density and frequency of gatherings are often low, and they are mainly concentrated outdoors, such as tourist attractions [41, 42]. During the COVID-19 epidemic, students in Xiamen were repeatedly required to study at home for a long time due to the prevention and control measures. Table 1 shows that the proportion of influenza in the 0–12 age group has decreased from 84.14% in phase A to 74.54% in phase B, but it is uncertain whether studying from home has reduced transmission in school, reduced transmission in public places, or both. Therefore, these situations may help to understand that the correlation between holidays and influenza is not significant.
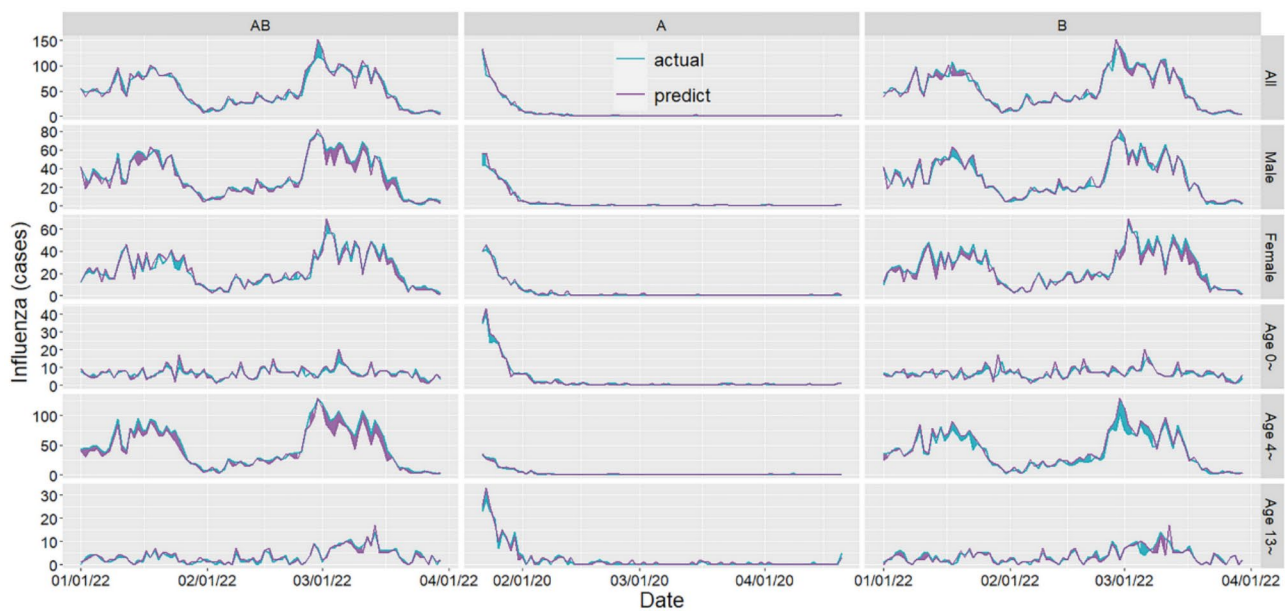
Fig. 12 Predicted true influenza values over three months based on RF-Bi-LSTM

The high incidence period of seasonal influenza in Xiamen often occurs in late winter and early spring from December to January of the following year, and in summer from May to July, while seasonal pollen is distributed from March to May each year. During the pollen period, there was a sustained moderate precipitation and high humidity, as confirmed by Fig. 6. Precipitation and relative humidity are negatively correlated with pollen concentration [43]. Rain makes pollen less airborne and very high humidity levels are even detrimental to pollen [21]. Pollen bio-aerosol and UV light exposure lead to immuno-activation, and sometimes allergic symptoms, which seem to protect against flu-like viruses, or at least severe outcomes from them [21]. Therefore, pollen may inhibit the spread of influenza, but the meteorological conditions have suppressed the concentration and pathogenicity of pollen in Xiamen. There was no significant correlation between allergens and influenza in this study, which may be due to the important role played by meteorological conditions during the pollen period.

Vaccination is currently the most effective response to influenza [44]. However, the average vaccination rate of influenza vaccine in Chinese Mainland from 2014 to 2020 was 2.43%, while the adult vaccination rate of the United States in the influenza season in 2020 was 48.4%, suggesting that the vaccination rate of influenza vaccine in Chinese Mainland is extremely low [44, 45]. The highest influenza vaccination rate in Xiamen from 2010 to 2021 was only 2.39% (2020), indicating that vaccination cannot form a sufficient immune barrier in the population. Moreover, the existing influenza vaccination protection rate is only 40–60%, and influenza viruses are prone to mutation [44]. Therefore, the current vaccination has a very limited effect on suppressing the influenza epidemic in Xiamen. The correlation analysis also showed that there was no significant correlation between the vaccination rate and the incidence rate of influenza ($r=0.41$, $P=0.18$) in this study. Therefore, the vaccination rate was not used for subsequent GAM analysis and prediction.

Figures 5 and 7 show that the descriptive statistics and correlation analysis results of influenza and meteorology factors in phase AB and phase A were generally consistent. However, the influenza and precipitation values in phase B significantly decreased, the minimum values of humidity and temperature increased, and the correlation between precipitation and the pres-difference and precipitation and temperature decreased. The abnormalities of influenza, pres-difference, precipitation, humidity, and temperature in phase B might have an impact on the correlation analysis between meteorology factors and influenza.

GAM analysis showed that although the relationship trend and risk value of meteorology factors with influenza in phase AB and phase A were somewhat similar, the risk effect of precipitation > 100 mm on influenza in phase A presented a mild upward trend, which was different from the continuous downward trend in phase AB, and the risk effect curve of wind on influenza in phase A was not as obvious as the U-shaped curve in phase AB, while the difference between phase B and phase AB and phase A was more significant.

GAM analysis showed that both low and high pressure in phase AB and phase A were risk factors for influenza, and they were generally consistent with the high

incidence of influenza (Fig. 6), mainly due to the significant correlation between high pressure and the influenza epidemic season, as well as between low pressure and the subepidemic season. However, in phase B, pressure is positively correlated with the onset of influenza. Therefore, due to the impact of phase B, high pressure in phase AB has a higher risk of influenza than in phase (A) Since the minimum, M_{50}, maximum, and $\bar{X}$ values of pressure in these three phases were very close, and no extreme pressure events were detected from 2020 to 2021 [19, 20], this was mainly due to the abnormal number of influenza cases in phase (B) In addition, the impact of pressure on influenza shows a significant cumulative effect. Wang et al. believe that pressure is closely related to temperature and humidity, thereby increasing the risk of influenza [9].

A high pres-difference was mainly distributed from January to February, and the initial stage of phase B had a high number of influenza cases, which led to a high pres-difference (8-10.5 hPa) being a risk factor for influenza. This shows an opposite trend to the phase A and phase AB. Affected by Phase B, high pres-difference in phase AB has a greater impact on influenza than in phase A.

The incidence of influenza in phase B, especially the non-epidemic influenza season in 2021 (late summer and early autumn), the number of influenza cases was relatively high in that year, when the humidity in Xiamen was also higher. Therefore, high humidity in phase B poses a more significant risk of influenza. In 2021, Xiamen experienced extreme weather with less cold and more heat and severe drought in winter and spring [19], resulting in a low humidity in the winter and spring influenza epidemic season. Although the overall fluctuation in phase B influenza was not significant, Fig. 6 shows that there were more influenza cases in the winter of 2021, and compared to Phase A, the humidity in the winter of phase B was lower, leading to this extreme weather being an important reason for the increased risk effect of low humidity on influenza in phase B. The impact of these phenomena in phase B on phase AB was not significant. The range of humidity levels that had a risk effect on influenza in phase A and phase AB was generally consistent with the humidity value characteristics reported by Ng et al. [46]. and Wang et al. [9].

Many extreme weather events occurred in Fujian Province in phase B, including low precipitation and weak rainstorm intensity in each quarter of 2020 and low precipitation in the winter and spring of 2021 [19, 20], which was consistent with the situation in which the P75, maximum and average values of precipitation in phase B were lower than those in phase A in this study. In 2021, the rainy season (from April to July) lasted a long time, and the average precipitation in Fujian Province was 12% higher than that in the same period of the previous year [19]. Although the overall number of phase B influenza

cases decreased substantially, the number of influenza cases in the summer and autumn of 2021 was relatively high, and the risk effect of high precipitation on influenza was enhanced in phase B. This is opposite to the trend of phase A and phase AB. However, compared with phase B and phase AB, phase A has a more significant cumulative and lag effect. Although the trends, value ranges, peaks, and lag times of risk effect of the four meteorology factors, including temperature, temp-difference, wind, and sunshine, on influenza in phase AB and phase A were basically consistent, there were significant differences in the risk characteristics of these four meteorology factors on influenza in phase B compared to phase AB and phase A (Figs. 9 and 10).

In phase AB and phase A, for every one IQR unit increase in day light, influenza cases showed a continuous downward trend with the increase of cumulative days in this study (low peak: accumulate 14d, AB: ER = -75.60; A: ER = -58.51), indicating that increasing daylight hours would be beneficial for suppressing influenza infection. Day light can affect immune function by regulating people's circadian rhythm [22, 23, 47]. The circadian rhythm is also important for coordinating complex biological processes such as immunity, which is most evident in the respiratory system and can be understood from a molecular perspective [23]. This study showed that the trend in phase B is opposite to that in phases AB and A, suggesting that this may have distorted the true relationship between sunlight and influenza. Therefore, it led to the impact of day light in phase A on influenza being significantly higher than that in phase AB.

The impact of meteorological factors on influenza is not significantly different between males and females in this study. However, compared to other age groups, children aged 4–12 were more affected by pressure, precipitation, sunshine, and day light, which may be related to the immune and behavioral characteristics of different age groups. During the early childhood period from 0 to 3 years old, the fetus has a stronger immune system due to obtaining more immunoglobulins in the mother's body. Children aged 4–12 have weaker immune function due to being in the immune system building phase, and frequent mobility and aggregation behaviors.

In summary, the sharp decline in influenza activity affected the analysis of the association between meteorology factors and influenza in phase AB, and this effect was more significant when only phase B was analyzed, although extreme weather conditions such as humidity and precipitation were also influencing factors.

The significant decline in the number of influenza cases in phase B might have been influenced by various factors, but the PHSM during the COVID-19 pandemic might be the most important reason. Many previous research reports have confirmed that PHSM during COVID-19

significantly limit influenza transmission [48–55]. PHSM includes wearing masks, maintaining interpersonal distance, restricting travel and gatherings, suspending classes, hand hygiene, etc., aiming to reduce the risk and scale of infectious disease transmission, as well as alleviate the burden on the health system, economy, and society [6, 50, 56, 57]. The effectiveness depends on intensity, the timing of implementation and/or de-implementation, socio-cultural aspects (such as societal compliance and trust in authority), etc. [6, 56]. For example, when the implementation time of PHSM coincides with the peak activity of the influenza season, the impact is significant [6]. Overall, COVID-19 PHSMs had reduced influenza transmissibility by a maximum of 17.3–40.6% and attack rate by 5.1–24.8% in the 2019–20 influenza season: for 11 different locations and countries [6].

Children were the most important group affected by influenza in Xiamen, with nearly 85% of children aged ≤ 12 years suffering from influenza [58]. Clearly, measures to prevent and control the COVID-19 epidemic, such as wearing masks, school closures, restricting gatherings, hand hygiene, etc. reduced the risk of transmission among the most vulnerable groups. Table 1 shows that the influenza proportion of children aged ≤ 12 years dropped from 84.14% before the COVID-19 period to 74.54% during the COVID-19 period, indicating that PHSM had a more significant impact on children in Xiamen. These interventions may alter the natural pattern of influenza transmission, thereby affecting the association features between meteorology factors and influenza incidence.

The method we used in this study was to input multifactor values for 7 days to predict the daily average influenza cases for the next 7 days and to achieve predictions through multistep rolling. This not only met the demand for daily predicted values in daily work but also avoided the situation of large relative errors between daily predicted values and actual values through the 7-day daily average method.

This study showed that RF-Bi-LSTM had low evaluation indicators, which means it had high prediction accuracy for influenza through meteorology factors. And regardless of whether before or after stratification, the prediction effect of period A is the best.

Although the time data and influenza data used to predict influenza reflect the characteristics of influenza epidemic caused by comprehensive factors such as vaccination, demographic characteristics, prevention and control policies, and influenza activity, the sharp decline of influenza in the context of PHSM during the COVID-19 epidemic significantly affected the association between meteorology factors and influenza, so meteorology factors data that cannot explain the truth of the correlation was captured when predicting influenza. This might

result in the prediction performance of phase B not being as good as phase AB or phase A. The prediction effect of the AB period was not as good as that of the A period, indicating that the AB period was also affected by the B period. The difference test of the predicted evaluation indicators in Table 4 also confirms this result.

The use of RF-Bi-LSTM for influenza prediction in Phase A achieved the best results, with MAPE values (0.08–0.14) significantly lower than similar studies that previously used LSTM for influenza prediction (0.47–0.88) [3], and also lower than the prediction effect of deep learning hybrid model for influenza (0.13–0.22), consistent with the autoregressive moving average-generalized autoregressive conditional heteroscedasticity (ARMA—GARCH) model prediction effect (0.08–0.14) [59]. The CC value (0.97–0.99) is significantly higher than the evaluation value (0.72–0.96) of predicting influenza using machine learning models such as RF, support vector regression (SVR), and extreme gradient boosting (XGBoost) [60]. Therefore, RF- Bi-LSTM can be used as a good model combination for predicting influenza.

Due to environmental factors such as meteorology and pollen allergens, the impact on respiratory infectious diseases such as influenza is continuously receiving high attention, and high-intensity government regulation will seriously affect the natural spread of respiratory infectious diseases. For example, PHSM on a global scale has severely suppressed the spread of influenza, thereby seriously affecting the analysis and prediction of the association between environmental factors and respiratory infectious diseases such as influenza. This limitation will persist for a long time, but such special periods can be excluded in the study. Ali et al.'s study showed that they can simulat influenza activity by constructing the standard susceptible–exposed–infected–recovered transmission model under the counterfactual scenario of implementation timing of COVID-19 PHSMs, and simulat to predict the incidence under no effect of PHSMs [6, 61].

This study also has some limitations. Firstly, solar radiation including ultraviolet (UV) light is another important indicator affecting infectious diseases [22], however, due to the lack of complete data, this study was not included in the analysis. Secondly, due to the inability to collect complete data on seasonal allergen pollen, this study replaced it with dummy variables based on pollen period. Thirdly, there is a lack of data on indoor and outdoor population mobility in public places, and online platforms such as Google and Baidu cannot well reflect the mobility data of children aged 0–12 who are most susceptible to influenza. Therefore, this study only analyzed the indicators that best reflect the degree of child mobility, holidays.

Conclusion

The varying degrees of intervention measures adopted during different phases can lead to significant differences in the impact of meteorology factors on influenza and in the influenza prediction. The sharp decline in influenza activity in the context of PHSM during the COVID-19 pandemic had significantly affected the long-term multi-year analysis of the association between meteorology factors and influenza and of the prediction of influenza.

This hints that when studying the correlation and prediction between meteorology factors and respiratory infectious diseases, it is important to select the data year span cautiously. In association studies of respiratory infectious diseases, especially influenza, and environmental factors, it is advisable to exclude periods with more external interventions to reduce interference with environmental factors and influenza related research, or to refine the model to accommodate the alterations brought about by intervention measures. In addition, the RF-Bi-LSTM model has good predictive performance for influenza.

Abbreviations

AIC	Akaike information criterion
ARMA-GARCH	Autoregressive moving average-generalized autoregressive conditional heteroscedasticity
Bi-LSTM	Bi-directional long short memory
CC	Correlation coefficient
CI	Confidence interval
COVID-19	Corona virus disease 2019
DLNM	Distribution lag nonlinear models
GAM	Generalized additive models
IA	Willmott's index of agreement
KGE	Kling-Gupta efficiency
LMI	Legate and McCabe's Index
MAE	Mean absolute error
MAPE	Mean absolute percentage error
NSE	Nash-Sutcliffe efficiency
PHSM	Public health and social measures
RF	Random forest
RMSE	Root mean squared error
RR	Risk ratio
RSR	RMSE-observations standard deviation ratio
SMAPE	Symmetric mean absolute percentage error
SVR	Support vector regression
XGBoost	Extreme gradient boosting

Author contributions
HZ, KZ, JO, GC and YL conceived and designed the study. HZ, SC, WQ, JA, YC, and XW carried out data analysis and draw pictures through software, and drafted the manuscript. HZ, SC, WQ, JA, YC, XW, KC, ZX, and LL revised the article. All authors read and approved the final manuscript.

Data availability
The datasets that support the findings of this study are available from Fujian Provincial Centre of Disease Control and Prevention, Fujian Climate Center and meteorological data network of the China Meteorological Administration, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of these three institutions (E-mail: hszhu33@126.com).

Declarations

Ethics approval and consent to participate
The Ethics Committee of Fujian Provincial Center for Disease Control and Prevention approved all the study procedures (2020-029), and waived the necessity for informed consent of participants according to the Infectious Diseases Law of the People's Republic of China. All methods were carried out in accordance with the relevant guidelines and regulations (Declaration of Helsinki). The data used in this study was always anonymized (including before this study).

Consent for publication
Not Applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Fujian Provincial Center for Disease Control and Prevention, Fuzhou, Fujian 350012, China
[2]Fujian Institute of Meteorological Sciences, Fuzhou, Fujian 350007, China
[3]The First Affiliated Hospital of Xiamen University, Xiamen, Fujian 361003, China
[4]School of Public Health, Xiamen University, Xiamen, Fujian 361100, China
[5]Fujian Provincial Judicial Drug Rehabilitation Hospital, Fuzhou, Fujian 350007, China
[6]Fuzhou University, Fuzhou, Fujian 350108, China
[7]Xiangnan University, Chenzhou, Hunan 423001, China
[8]School of Public Health, Fujian Medical University, Fuzhou, Fujian 350011, China
[9]Fujian Key Laboratory of Severe Weather, Fuzhou, Fujian 350007, China
[10]Key Laboratory of Straits Severe Weather, China Meteorological Administration, Fuzhou, Fujian 350007, China

References
1. Paget J, Spreeuwenberg P, Charu V, Taylor RJ, Iuliano AD, Bresee J, Simonsen L, Viboud C. Global mortality associated with seasonal influenza epidemics: New burden estimates and predictors from the GLaMOR Project. J Glob Health. 2019;9(2):020421.
2. Mortality morbidity. Hospitalisations due to influenza lower respiratory tract infections, 2017: an analysis for the global burden of Disease Study 2017. Lancet Respir Med. 2019;7(1):69–89.
3. Zhu H, Chen S, Lu W, Chen K, Feng Y, Xie Z, Zhang Z, Li L, Ou J, Chen G. Study on the influence of meteorological factors on influenza in different regions and predictions based on an LSTM algorithm. BMC Public Health. 2022;22(1):2335.
4. Song S, Li Q, Shen L, Sun M, Yang Z, Wang N, Liu J, Liu K, Shao Z. From outbreak to Near Disappearance: how did non-pharmaceutical interventions against COVID-19 affect the transmission of Influenza Virus? Front Public Health. 2022;10:863522.
5. Zipfel CM, Colizza V, Bansal S. The missing season: the impacts of the COVID-19 pandemic on influenza. Vaccine. 2021;39(28):3645–8.
6. Ali ST, Lau YC, Shan S, Ryu S, Du Z, Wang L, Xu XK, Chen D, Xiong J, Tae J, et al. Prediction of upcoming global infection burden of influenza seasons after relaxation of public health and social measures during the COVID-19 pandemic: a modelling study. Lancet Glob Health. 2022;10(11):e1612–22.

7. Influenza monitoring weekly report [. https://ivdc.chinacdc.cn/cnic/zyzx/lgzb/202204/P020220506749036060201.pdf]
8. Influenza. monitoring weekly report [https://ivdc.chinacdc.cn/cnic/zyzx/lgzb/202306/P020230615281509963224.pdf]
9. Wang J, Zhang L, Lei R, Li P, Li S. Effects and Interaction of Meteorological Parameters on Influenza Incidence during 2010–2019 in Lanzhou, China. Front Public Health. 2022;10:833710.
10. Gomez-Barroso D, León-Gómez I, Delgado-Sanz C, Larrauri A. Climatic factors and Influenza Transmission, Spain, 2010–2015. Int J Environ Res Public Health 2017, 14(12).
11. Zheng Y, Wang K, Zhang L, Wang L. Study on the relationship between the incidence of influenza and climate indicators and the prediction of influenza incidence. Environ Sci Pollut Res Int. 2021;28(1):473–81.
12. Zhang L, Ma C, Duan W, Yuan J, Wu S, Sun Y, Zhang J, Liu J, Wang Q, Liu M. The role of absolute humidity in influenza transmission in Beijing, China: risk assessment and attributable fraction identification. Int J Environ Health Res. 2024;34(2):767–78.
13. Park JE, Son WS, Ryu Y, Choi SB, Kwon O, Ahn I. Effects of temperature, humidity, and diurnal temperature range on influenza incidence in a temperate region. Influenza Other Respir Viruses. 2020;14(1):11–8.
14. Wang D, Lei H, Wang D, Shu Y, Xiao S. Association between temperature and influenza activity across different regions of China during 2010–2017. Viruses 2023, 15(3).
15. Yan ZL, Liu WH, Long YX, Ming BW, Yang Z, Qin PZ, Ou CQ, Li L. Effects of meteorological factors on influenza transmissibility by virus type/subtype. BMC Public Health. 2024;24(1):494.
16. Soebiyanto RP, Gross D, Jorgensen P, Buda S, Bromberg M, Kaufman Z, Prosenc K, Socan M, Vega Alonso T, Widdowson MA, et al. Associations between Meteorological Parameters and Influenza Activity in Berlin (Germany), Ljubljana (Slovenia), Castile and León (Spain) and Israeli districts. PLoS ONE. 2015;10(8):e0134701.
17. Si X, Wang L, Mengersen K, Hu W. Epidemiological features of seasonal influenza transmission among 11 climate zones in Chinese mainland. Infect Dis Poverty. 2024;13(1):4.
18. The Lancet Countdown Report on Population Health and Climate Change. 2023 - Health risks from climate change continue to rise [https://www.cma.gov.cn/2011xwzx/2011xqxkj/2011xkjdt/202311/t20231122_5900714.html]
19. Climate Bulletin of Fujian Province in. 2021 [http://fj.weather.com.cn/zxfw/qhgb/03/3527254_2.shtml]
20. Climate Bulletin of Fujian Province in. 2020 [http://www.weather.com.cn/fujian/zxfw/qhgb/03/3445765_2.shtml]
21. Martijn J, Hoogeveen, Eric CM, van Gorp EK. Hoogeveen: can pollen explain the seasonality of flu-like illnesses in the Netherlands? Sci Total Environ. 2021;755(2):143182.
22. Martijn J, Hoogeveen, Aloys CM, Kroes EK. Hoogeveen: environmental factors and mobility predict COVID-19 seasonality in the Netherlands. Environ Res. 2022;211:113030.
23. Nosal C, Ehlers A, Haspel JA. Why lungs keep time: circadian rhythms and lung immunity. Annu Rev Physiol. 2020;82:391–412.
24. Varela-Lasheras I, Perfeito L, Mesquita S, Gonçalves-Sá J. The effects of weather and mobility on respiratory viruses dynamics before and during the COVID-19 pandemic in the USA and Canada. PLOS Digit Health. 2023;2(12):e0000405.
25. Zou H, Huang Y, Chen T, Zhang L. Influenza vaccine hesitancy and influencing factors among university students in China: a multicenter cross-sectional survey. Ann Med. 2023;55(1):2195206.
26. Schuster M, Paliwal KK. Bidirectional recurrent neural networks. IEEE Trans Signal Process. 1997;45(11):2673–81.
27. Shahid F, Zameer A, Muneeb M. Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. Chaos Solitons Fractals. 2020;140:110212.
28. Guo H, Chen B. Yearbook of Xiamen Special Economic Zone; 2022.
29. Koyak RA. Generalized additive models. J Am Stat Assoc. 1991;86(416):1140–1.
30. Fan L, Gu Q, Zeng Q. Progress in the application of generalized additive model in epidemiologic studies on air pollution. J Environ Occup Med. 2019;36(7):676–81.
31. Wu C, Yan Y, Chen X, Gong J, Guo Y, Zhao Y, Yang N, Dai J, Zhang F, Xiang H. Short-term exposure to ambient air pollution and type 2 diabetes mortality: a population-based time series study. Environ Pollut. 2021;289:117886.
32. Behr A, Schiwy C, Weinblat J. Investment, default propensity score and cash flow sensitivity in six EU member states: evidence based on firm-level panel data. Appl Econ. 2019;51(49):5345–68.
33. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32.
34. Zhu H, Chen S, Liang R, Feng Y, Joldosh A, Xie Z, Chen G, Li L, Chen K, Fang Y, Ou J. Study of the influence of meteorological factors on HFMD and prediction based on the LSTM algorithm in Fuzhou, China. BMC Infect Dis. 2023;23(1):299.
35. Chicco D, Warrens MJ, Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput Sci. 2021;7:e623.
36. Hosseini S, Khatti J, Taiwo BO, Fissha Y, Grover KS, Ikeda H, Pushkarna M, Berhanu M, Ali M. Assessment of the ground vibration during blasting in mining projects using different computational approaches. Sci Rep. 2023;13(1):18582.
37. Jeon S, Kim S, Lee M, An H, Jung K, Um M-J, An K, Park D. Insights into the Pollutant removal performance of stormwater green infrastructures: a Case Study of Detention basins and Retention ponds. Int J Environ Res Public Health. 2021;18:10104.
38. Willmott CJ. Some comments on the evaluation of Model Performance. Bull Am Meteorol Soc. 1982;63:1039–313.
39. Prasad SS, Deo RC, Salcedo-Sanz S, Downs NJ, Casillas-Pérez D, Parisi AV. Enhanced joint hybrid deep neural network explainable artificial intelligence model for 1-hr ahead solar ultraviolet index prediction. Comput Methods Programs Biomed. 2023;241:107737.
40. Ghimire S, Yaseen ZM, Farooque AA, Deo RC, Zhang J, Tao X. Streamflow prediction using an integrated methodology based on convolutional neural network and long short-term memory networks. Sci Rep. 2021;11:17497.
41. Zhenjun Luo. Reviewer comments: a study on the Safety issues of students during holidays. J Xiangtan Normal Univ (Social Sci Edition). 2009;31(3):69–71.
42. Xueyi, Wang. How to manage students during summer and winter vacations and holidays. Res Educational Theory. 2018;798(24):166–7.
43. Rahman A, Khan MHR, Luo C, Yang Z, Ke J, Jiang W. Variations in airborne pollen and spores in urban Guangzhou and their relationships with meteorological variables. Heliyon. 2021;7(11):e08379.
44. Li J, Zhang Y, Zhang X, Liu L. Influenza and Universal Vaccine Research in China. Viruses, 2022, 15(1).
45. Fan J, Cong S, Wang N, Bao H, Wang B, Feng Y, Lv X, Zhang Y, Zha Z, Yu L, Yang T, Wang L, Fang L. Influenza vaccination rate and its association with chronic diseases in China: results of a national cross-sectional study. Vaccine. 2020;38(11):2503–11.
46. Ng H, Li Y, Zhang T, Lu Y, Wong C, Ni J, Zhao Q. Association between multiple meteorological variables and seasonal influenza A and B virus transmission in Macau. Heliyon. 2022;8(11):e11820.
47. Kim S, Casement MD. Promoting adolescent sleep and circadian function: a narrative review on the importance of daylight access in schools. Chronobiol Int. 2024;41(5):725–37.
48. Arellanos-Soto D, Padilla-Rivas G, Ramos-Jimenez J, Galan-Huerta K, Lozano-Sepulveda S, Martinez-Acuña N, Treviño-Garza C. Montes-De-ocaluna R, de-la OCM, Rivas-Estilla AM: decline in influenza cases in Mexico after the implementation of public health measures for COVID-19. Sci Rep. 2021;11(1):10730.
49. Song S, Wang P, Li J, Nie X, Liu L, Liu S, Yin X, Lin A. The indirect impact of control measures in COVID-19 pandemic on the incidence of other infectious diseases in China. Public Health Pract (Oxf). 2022;4:100278.
50. Kuitunen I. Influenza season 2020–2021 did not begin in Finland despite the looser social restrictions during the second wave of COVID-19: a nationwide register study. J Med Virol. 2021;93(9):5626–9.
51. Liu Y, Yue S, Hu X, Zhu J, Wu Z, Wang J, Wu Y. Associations between feelings/behaviors during COVID-19 pandemic lockdown and depression/anxiety after lockdown in a sample of Chinese children and adolescents. J Affect Disord. 2021;284:98–103.
52. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, Ho RC. Immediate psychological responses and Associated Factors during the initial stage of the 2019 Coronavirus Disease (COVID-19) epidemic among the General Population in China. Int J Environ Res Public Health 2020, 17(5).
53. Horton R. Offline: 2019-nCoV-A desperate plea. Lancet. 2020;395(10222):400.
54. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. Lancet. 2020;395(10223):470–3.
55. Xie X, Xue Q, Zhou Y, Zhu K, Liu Q, Zhang J, Song R. Mental Health Status among children in Home Confinement during the Coronavirus Disease 2019 Outbreak in Hubei Province, China. JAMA Pediatr. 2020;174(9):898–900.
56. Rehfuess EA, Movsisyan A, Pfadenhauer LM, Burns J, Ludolph R, Michie S, Strahwald B. Public health and social measures during health emergencies such as the COVID-19 pandemic: an initial framework to conceptualize and classify measures. Influenza Other Respir Viruses. 2023;17(3):e13110.

57. Emborg HD, Carnahan A, Bragstad K, Trebbien R, Brytting M, Hungnes O, Byström E, Vestergaard LS. Abrupt termination of the 2019/20 influenza season following preventive measures against COVID-19 in Denmark, Norway and Sweden. Euro Surveill. 2021;26(22):2001160.

58. Zhu H, Wang M, Xie Z, Huang W, Lin J, Ye W, Chen S. Effect of meteorological factors on influenza incidence in Xiamen city. Chin J Public Health. 2019;35(10):1404.

59. Zhang X, Yang L, Chen T, Wang Q, Yang J, Zhang T, Yang J, Zhao H, Lai S, Feng L, Yang W. Predicting influenza-like illness trends based on sentinel surveillance data in China from 2011 to 2019: a modelling and comparative study. Infect Dis Model. 2024;9(3):816–27.

60. Cheng HY, Wu YC, Lin MH, Liu YL, Tsai YY, Wu JH, Pan KH, Ke CJ, Chen CM, Liu DP, et al. Applying machine learning models with an Ensemble Approach for Accurate Real-Time Influenza Forecasting in Taiwan: Development and Validation Study. J Med Internet Res. 2020;22(8):e15394.

61. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. Am J Epidemiol. 2013;178:1505–12.

Publisher's Note