The Journal of Macau University of Science and Technology VOL.17 NO.2
June 2023, pp.1-30
DOI: 10.58664/mustjournal.2023.06.001

Review of Influenza Prediction*

Wei Qu[136#], Yuanfang Lai[3#], Zhijie Lin[2], Ying Zhang[2], Jingyi Liang[2], Ruihan Chen[2], Ruilan Deng[3], Zhiqi Zeng[13*], Zifeng Yang[134*], Chitin Hon[245*], Nanshan Zhong[1]

1 State Key Laboratory of Respiratory Disease, National Clinical Research Center for Respiratory Disease, Guangzhou Institute of Respiratory Health, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, Guangdong 510120, People's Republic of China.

2 Department of Engineering Science, Faculty of Innovation Engineering, Macau University of Science and Technology, Taipa, Macau, China.

3 Guangzhou Key Laboratory for Clinical Rapid Diagnosis and Early Warning of Infectious Diseases, King Med School of Laboratory Medicine, Guangzhou Medical University, Guangzhou, People's Republic of China.

4 Guangzhou Laboratory, Guangzhou, Guangdong 510120, People's Republic of China.

5 Respiratory Disease AI Laboratory on Epidemic Intelligence and Medical Big Data Instrument Applications, Macau, China.

6 College of Sciences, China Jiliang University, China

\# **These authors contributed equally to this work.**

Abstract: Influenza is a common respiratory disease caused by the influenza virus, and each outbreak can result in a large number of infections and deaths worldwide. Predicting the dynamics of influenza outbreaks could be useful for decision-making regarding the allocation of public health resources. Due to the rapid advancement of science and technology in recent years, significant improvements have been made in influenza early warning and prediction, and the original systems have been improved and diversified. In this review, we gather several common influenza prediction systems and early warning methods, as well as some strategies to minimize prediction errors. In addition, we discuss the various data sources and collect the prediction situation on local, regional, national, or global level. Last but not least, we summarized the accuracy and advantages and disadvantages of various prediction methods.

Keywords: Influenza prediction; Surveillance systems; Compartmental models; Time series analysis; Machine learning methods

---

* Received and Accepted: 20 May 2023.
This paper is Special Article for *Journal of Macau University of Science and Technology*.

1. Introduction

Influenza (flu), is an acute respiratory infection caused by the influenza virus that is highly contagious and spreads quickly. Historically, every clearly documented influenza pandemic had resulted in a large scale of infections and deaths. Today, influenza still causes about 3 to 5 million severe cases and about 250,000 to 500,000 fatalities every year.[1] Influenza was and continues to be a serious public health issue confronting human society. If influenza outbreaks could be predicted in advance, health authorities would gain more valuable time to prepare invention measures against the epidemic for the sake of reducing influenza hazards and mortality.

There are two main types of viruses that cause influenza, namely, influenza A virus and influenza B. Influenza early warning forecasting involves converting surveillance data from clinical laboratories for confirmed influenza and outpatient visits for suspected influenza disease into weekly incidence forecasts, while the forecast data may include crowdsourced data from social media, epidemiology, or evolution, through which models are built to track when, where, and at what ages influenza occurs and form real-time forecasts, from which modelers predict the influenza seasonal trend and peak intensity.[2]

Timely monitoring and analysis of influenza flu characteristics and virus activity patterns, to effectively determine the development trend of the epidemic and formulate scientific control strategies, are currently important tools for preventing influenza outbreaks and pandemics. Additionally, the theoretical study of infectious disease transmission processes, prevention and control impacts, and prediction and early warning models all benefit from the use of mathematical tools in conjunction with statistical methods. Influenza early warning prediction reflects the occurrence, development and transmission characteristics of influenza disease through mathematical models, shows its transmission mechanism and risk, analyzes the social or environmental and other factors

---

1 Dugas, A. F.; Jalalpour, M.; Gel, Y.; Levin, S.; Torcaso, F.; Igusa, T.; Rothman, R. E., "Influenza forecasting with Google Flu Trends," *PLoS One* 8.2 (2013): e56176.

2 Osthus, D.; Moran, K. R., "Multiscale Influenza Forecasting," *Nat Commun* 12.1 (2021): 2991.

associated with it, predicts the trend, realizes early warning of influenza, and evaluates the effectiveness of control measures. Meanwhile, influenza early warning prediction applies complex statistical models to the actual control of influenza, which has become a key direction in the development of public health statistics and theoretical epidemiology.[3]At present, the models applied to the study of influenza transmission characteristics such as prevention and control effects mainly include time series models,[4] SIR dynamics models,[5] neural network models,[6] Bayesian-Markov chain-Monte Carlo models, regression models, etc. With the advancement of computer technology in recent years, individual-based simulation models, metacellular automata, multi-intelligent body systems, wavelet neural networks and other technologies are also gradually emerging.

The main aims of the present study are to (i) summarize existing approaches to

---

3 Reich, N. G.; Brooks, L. C.; Fox, S. J.; Kandula, S.; McGowan, C. J.; Moore, E.; Osthus, D.; Ray, E. L.; Tushar, A.; Yamana, T. K.; Biggerstaff, M.; Johansson, M. A.; Rosenfeld, R.; Shaman, J.,"A Collaborative Multiyear, Multimodel Assessment of Seasonal Influenza Forecasting in the United States," *Proc Natl Acad Sci USA* 116.8 (2019): 3146-3154.

4 Reich, N. G.; Brooks, L. C.; Fox, S. J.; Kandula, S.; McGowan, C. J.; Moore, E.; Osthus, D.; Ray, E. L.; Tushar, A.; Yamana, T. K.; Biggerstaff, M.; Johansson, M. A.; Rosenfeld, R.; Shaman, J., "A Collaborative Multiyear, Multimodel Assessment of Seasonal Influenza Forecasting in the United States," *Proc Natl Acad Sci USA* 116,8 (2019): 3146-3154; Samaras, L.; Garcia-Barriocanal, E.; Sicilia, M. A., "Syndromic Surveillance Models Using Web Data: The Case of Influenza in Greece and Italy Using Google Trends," *JMIR Public Health Surveill* 3.4 (2017): e90; Osthus, D.; Gattiker, J.; Priedhorsky, R.; Del Valle, S. Y., "Dynamic Bayesian Influenza Forecasting in the United States with Hierarchical Discrepancy (with Discussion)." *Bayesian Analysis* 14.1 (2019); Li, R.; Bai, Y.; Heaney, A.; Kandula, S.; Cai, J.; Zhao, X.; Xu, B.; Shaman, J., "Inference and forecast of H7N9 influenza in China, 2013 to 2015," *Euro Surveill* 22.7 (2017).

5 Pei, S.; Shaman, J., "Counteracting Structural Errors in Ensemble Forecast ofInfluenza Outbreaks." *Nat Commun* 8.1 (2017), 925; Yang, W.; Cowling, B. J.; Lau, E. H.; Shaman, J., "Forecasting Influenza Epidemics in Hong Kong." *PLoS Comput Biol* 11,7 (2015): e1004383; Pei, S.; Kandula, S.; Yang, W.; Shaman, J., "Forecasting the spatial transmission of influenza in the United States." *Proc Natl Acad Sci U S A* 115.11 (2018): 2752-2757.

6 Jiang-Ning, L.; Xian-Liang, S.; An-Qiang, H.; Ze-Fang, H.; Yu-Xuan, K.; Dong, L., "Forecasting Emergency Medicine Reserve Demand with a Novel Decomposition-ensemble Methodology," *Complex Intell Systems* 2021, 1-11; Paul, S.; Mgbere, O.; Arafat, R.; Yang, B.; Santos, E., "Modeling and Forecasting Influenza-like Illness (ILI) in Houston, Texas Using Three Surveillance Data Capture Mechanisms." *Online J Public Health Inform* 9.2 (2017): e187; Kandula, S.; Shaman, J., "Near-term Forecasts of Influenza-like Illness: An Evaluation of Autoregressive Time Series Approaches." *Epidemics* 2019, 27, 41-51; Lee, K.; Ray, J.; Safta, C., "The Predictive Skill of Convolutional Neural Networks Models for Disease Forecasting," *PLoS One* 16,7 (2021): e0254319.

influenza prediction, (ii) present differences in measures of accuracy and evaluate the degree to which various performance measures are met, (iii) discuss limitations in the data sources and different methods. The motivation of this review is to inform further research on influenza prediction and provide researchers and public health practitioners with a summary of the accomplishments and limitations in influenza prediction.

2. Article selection and evaluation

The scope of this review includes studies on forecasting methods, data sources for model predictions, and the prediction of influenza dynamics at the local, regional, national, or global level. Firstly, we searched articles on influenza prediction on the Web of Science using the keywords “flu Prediction OR flu forecasting OR influenza Prediction OR influenza forecasting”, resulting in a total of 2522 literatures. After obtaining a general understanding of the content related to influenza prediction, we conducted a second search to identify different models used in influenza prediction. We narrowed down the search range with specific keywords, including “influenza forecast” and “SIR model”, “influenza forecast” and “SEIR model”, “influenza forecast” and “SEIRS model”, “influenza forecast” and “LSTM”, “influenza forecast”and “ARIMA”, “influenza forecast” and “GARMA”, and “influenza forecast” and “CNN”. The results showed that there were 33 literatures for the SIR model, 17 literatures for the SEIR model, 3 literatures for the SEIRS model, 21 literatures for LSTM, 47 literatures for ARIMA, 1 literature for GARMA, and 3 literatures for CNN. To identify relevant data sources for influenza prediction, we further adjusted the keywords to “influenza forecast” and “data sources”, “influenza forecast” and “ILInet”, “influenza forecast” and “Google trends”, and “influenza forecast” and “Internet”. After screening, we selected a total of 79 articles for analysis and grouped and presented the studies based on the prediction indicators.

3. Data sources

From the 79 sort-listed papers, we summarize various data sources for influenza prediction researches from 2006 to 2023, which can be mainly divided into 7 categories: Public health data, Hospital visit data, Virology surveillance network, Self-reported data from the general public, Social media data, Human behavior patterns, and Weather data **(Table 1)**. These data sources demonstrate the importance of different sources in monitoring and analyzing influenza activity. CDC's ILI data and laboratory-confirmed influenza cases are the most used sources in public health data, with a total of 64 research papers. In addition, social media and internet search data, such as Google Flu Trend, Twitter, Baidu Index, etc., account for a total of 43 research papers. Weather data, hospital visit data and human behavior patterns are also widely used, appearing 13, 13 and 5 times, respectively. Some data sources were only used once, such as Google search query data, WordPress flu-related blogs, OTC surveillance, School-based ILI surveillance, Yahoo, US DOD AFHSB'ILI data, Anonymized mobility map (AMM), Sina Weibo, National Health and Family Planning Commission (NHFPC) **(Figure 1, supplement table 1)**. These data sources can provide seasonal and regional patterns of influenza transmission and disease trends, thereby helping to formulate relevant prevention and control strategies.

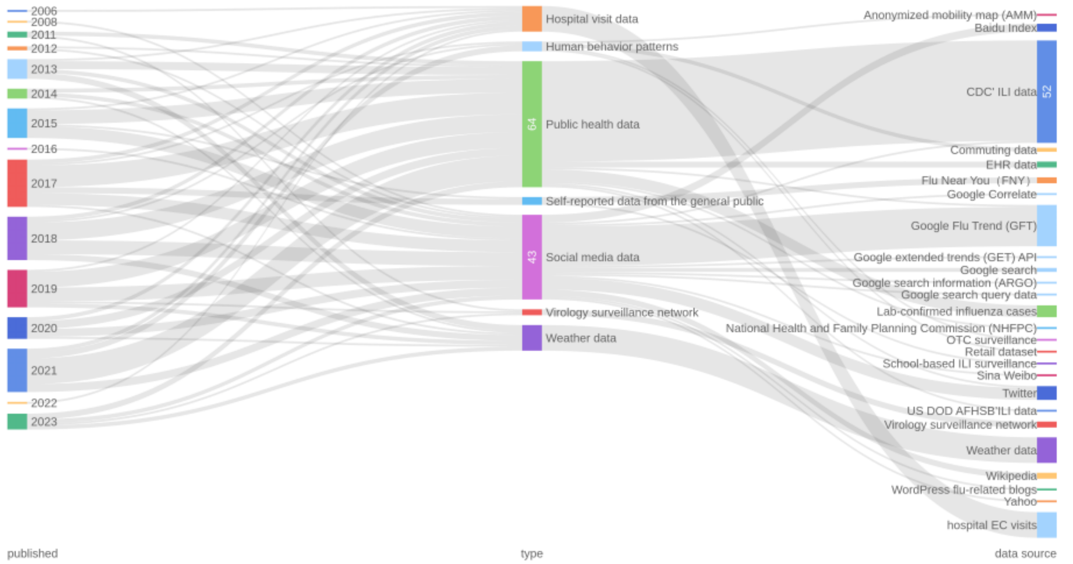
Table 1. Summary of data source characteristics

Data source	Characteristics	Reference
Public health data	China-CDC or CDC national sentinel surveillance system（ILINet）weekly influenza surveillance reports, [2,4,6,14] Japanese historical ILI data (NIID data): [76] ILI datasets are available, as well as patient epidemiology of influenza outbreaks information, influenza activity levels, virus types, etc. **Electronic health record data（EHR data）:** [49] Includes structured (lab test results, prescriptions, ICD-10 diagnoses) and unstructured (discharge letters, pathology reports, surgery reports) patient data。 **National Home Doctor Service (NHDS)**: [54]Estimating influenza activity levels based on influenza-like	2,4,6,11,14,31,4954,76

	symptoms found during home visits. **Lab-confirmed influenza cases:** [31] Percentage of laboratory analyses that include positive tests for all respiratory specimens reported from multiple clinical laboratory facilities. **US DOD AFHSB**: [11] Includes health data on military personnel and their families serving within the US Department of Defense.	
Hospital visit data	**Hospital emergency center (EC) visits**: [3,12] The information on age, gender, severity of illness, medication, etc. of influenza patients.	3,12
Virology surveillance network	**National Respiratory and Enteric Virus Surveillance System (NREVSS)**: [69]Virological surveillance reports the total number of respiratory specimens tested weekly and the number of influenza A and B viruses detected.	69
Self-reported data from the general public	**Flu Near You（FNY）:** [36] A community-engaged influenza surveillance system that collects information on flu-like illness symptoms reported by the public through an online questionnaire and uses these data to monitor influenza activity levels nationwide. **School-based ILI surveillance**: [13] School-based influenza-like illness data includes ILI cases and leave of absence associated with ILI.	3,36
Social media data	Google extended trends (GET) AP, [I6,5] Google Flu Trend (GFT), [14,12] Twitter, [15,38] Google search information (ARGO), [75] Google search query data, [52]Baidu Index, [53]Wikipedia, [60]Google Correlate, [70] WordPress flu-related blogs, [31] Yahoo: [37] Extracting timeline data on the frequency of queries for search terms or terms over a specific time period, the web tool allows users to access search trend data for various time and geographical areas to generate real-time estimates (nowcasts) of flu-like illnesses that can be used for	5-6,12,14-15,31,37-38,52-53,60,70,75

	analysis and research.	
Human behavior patterns	Shopping behavior: **Retail dataset:** [46] As a proxy for predicting seasonal influenza, changes in customer behaviour are reflected in the items purchased in the shopping basket, thus providing a valuable proxy for the spread of seasonal influenza. **OTC surveillance**[36]**:** Sales data from designated over-the-counter medicines for the treatment of colds and acute respiratory infections, including tablet, powder, granule and syrup forms. Commuting behavior: **Commuting data:** [50] Considers the spread of visitors between any two regions to predict the spread of intracity epidemics using the volume of traffic in the city. **Anonymize mobility map (AMM):** [24]The dataset consists of an aggregation from users who have turned on the location history setting, which aggregates the traffic of people from one area to another.	24,36,46,50
Weather data	**Weather data provided by the weather forecasting agencies**: [14,12] including temperature, local humidity, precipitation, sunlight, wind speed and other meteorological parameters that can help researchers analyse the seasonal and geographical patterns of influenza virus transmission.	12,14

Figure1. Sankey diagram for classification of data sources



4. Methods

Influenza prediction methods can be divided into the following categories **(also can be seen in table 2)**.

Time-series analysis-based methods can identify trends, seasonality, and periodicity in the data, and then apply these trends and periodicity to future predictions. The main advantage of this method is simplicity, but because it is based on historical data, the influence of other important factors, such as climate change and population migration.

SEIR models and their various can consider many factors, such as population migration, climate change, influenza virus variation, etc., thus with high accuracy and prediction accuracy. However, the method requires substantial data and computational resources, along with having expertise in infectious disease dynamics.

Models of infectious disease dynamics can be combined with social media-based methods to consider the characteristics of data and models of infectious disease dynamics on social media.

4.1 Methods based on the time-series analysis

Time series analysis is one of the earliest methods adopted in influenza prediction, mainly including autoregressive, multiple regression, etc. Samaras, L., et al. used multiple regressions of the terms submitted in the Google search engine related to influenza for the period from 2011 to 2012 in Greece and Italy (sample data for 104 weeks for each country).[7] They then used the autoregressive integrated moving average statistical model to determine the correlation between the Google search data and the real influenza cases confirmed by the aforementioned authorities. Dugas, A. F., et al. developed a practical influenza forecast model based on real-time, [8]geographically focused, and easy to access data, designed to provide individual medical centers with advanced warning of the expected number of influenza cases, thus allowing for sufficient time to implement interventions. Zhang, Y., et al. collected influenza notifications, temperature and Google Trends (GT) data between January 1st, 2011 and December 31st, 2016.[9] They performed time-series cross correlation analysis and temporal risk analysis to discover the characteristics of influenza epidemics in the period. The seasonal autoregressive integrated moving average (SARIMA) model and regression tree model were developed to track influenza epidemics using GT and climate data. Guolo, A. and C. Varin propose a practical approach to analyze bounded time series, through a beta regression model.[10] The method allows the direct interpretation of the regression parameters on the original response scale, while properly accounting for the heteroskedasticity typical of bounded variables. The methodology is motivated by an application to the influenza-like-illness incidence estimated by the Google Flu Trends project.

---

7 Samaras, L.; Garcia-Barriocanal, E.; Sicilia, M. A., "Syndromic Surveillance Models Using Web Data: The Case of Influenza in Greece and Italy Using Google Trends." *JMIR Public Health Surveill* 3.4 (2017): e90.

8 Dugas, A. F.; Jalalpour, M.; Gel, Y.; Levin, S.; Torcaso, F.; Igusa, T.; Rothman, R. E., "Influenza Forecasting with Google Flu Trends." *PLoS One* 8.2 (2013): e56176.

9 Zhang, Y.; Bambrick, H.; Mengersen, K.; Tong, S.; Hu, W., "Using Google Trends and Ambient Temperature to Predict Seasonal Influenza Outbreaks." *Environ Int* 117 (2018): 284-291.

10 Guolo, A.; Varin, C., "Beta regression for time series analysis of bounded data, with application to Canada Google® Flu Trends," *The Annals of Applied Statistics* 8.1 (2014).

4.2 Machine learning-based approach

With the development of computer technology, machine learning-based methods are getting more and more attention. This approach utilizes machine learning algorithms to identify trends in influenza outbreaks. Su, K., et al. collect Multi-source electronic data, including historical percentage of influenza-like illness (ILI%), weather data, Baidu search index and Sina Weibo data of Chongqing, China, were collected and integrated into an innovative Self-adaptive AI Model (SAAIM), which was constructed by integrating Seasonal Autoregressive Integrated Moving Average model and XGBoost model using a self-adaptive weight adjustment mechanism.[11] Venkatramanan, S., et al. focus on a machine-learned anonymized mobility map (hereon referred to as AMM) aggregated over hundreds of millions of smartphones and evaluate its utility in forecasting epidemics.[12] They factor AMM into a metapopulation model to retrospectively forecast influenza in the USA and Australia and show that the AMM model performs on-par with those based on commuter surveys, which are sparsely available and expensive. They also compare it with gravity and radiation-based models of mobility and find that the radiation model's performance is quite similar to AMM and commuter flows. Additionally, they demonstrate the model's ability to predict disease spread even across state boundaries. Yang, L., et al. established a new multiattention-long short-term memory (LSTM) deep-learning model (MAL model), which was used to predict the percentage of ILI (ILI%) cases and the product of ILI% and the influenza-positive rate (ILI%xpositive%), respectively.[13] They

---

11 Su, K.; Xu, L.; Li, G.; Ruan, X.; Li, X.; Deng, P.; Li, X.; Li, Q.; Chen, X.; Xiong, Y.; Lu, S.; Qi, L.; Shen, C.; Tang, W.; Rong, R.; Hong, B.; Ning, Y.; Long, D.; Xu, J.; Shi, X.; Yang, Z.; Zhang, Q.; Zhuang, Z.; Zhang, L.; Xiao, J.; Li, Y., “Forecasting Influenza Activity Using Self-adaptive AI Model and Multi-source Data in Chongqing, China.” *EBioMedicine* 47 (2019): 284-292.

12 Venkatramanan, S.; Sadilek, A.; Fadikar, A.; Barrett, C. L.; Biggerstaff, M.; Chen, J.; Dotiwalla, X.; Eastham, P.; Gipson, B.; Higdon, D.; Kucuktunc, O.; Lieber, A.; Lewis, B. L.; Reynolds, Z.; Vullikanti, A. K.; Wang, L.; Marathe, M., “Forecasting Influenza Activity Using Machine-learned Mobility Map.” *Nat Commun* 12.1 (2021), 726.

13 Yang, L.; Li, G.; Yang, J.; Zhang, T.; Du, J.; Liu, T.; Zhang, X.; Han, X.; Li, W.; Ma, L.; Feng, L.; Yang, W., “Deep-Learning Model for Influenza Prediction From Multisource Heterogeneous Data in a Megacity: Model Development and Evaluation.” *J Med Internet Res* 25 (2023): e44238.

also combined the data in different forms and added several machine-learning and deep-learning models commonly used in the past to predict influenza trends for comparison. Lu, F. S., et al. introduce a methodological framework which dynamically combines two distinct influenza tracking techniques, using an ensemble machine learning approach, to achieve improved state-level influenza activity estimates in the United States.[14] Clemente, L., et al. use machine learning-based methodology that uses flu-related Internet search activity and historical information to monitor flu activity, named ARGO (AutoRegression with Google search), was extended to generate flu predictions for 8 Latin American countries (Argentina, Bolivia, Brazil, Chile, Mexico, Paraguay, Peru, and Uruguay) for the time period: January 2012 to December 2016.[15]

4.3 Methods based on a kinetic model of infectious diseases

Pei, S. and J. Shaman conducted an investigation on the error growth of a compartmental influenza model, and discovered that a robust error structure emerges naturally from the nonlinear dynamics of the model.[16] They developed a novel forecasting method by addressing these structural errors identified through error breeding. This approach combines dynamical error correction with statistical filtering techniques.

Yang, W., et al. conducted a retrospective study in which they applied susceptible-infected-recovered (SIR) model-filter systems to predict influenza epidemics in Hong Kong from January 1998 to December 2013, including the 2009 pandemic.[17] The model predicted the timing and magnitude of the peak for 44 epidemics caused by various influenza strains, such as seasonal influenza A(H1N1), pandemic A(H1N1), A(H3N2), and

---

[14] Lu, F. S.; Hattab, M. W.; Clemente, C. L.; Biggerstaff, M.; Santillana, M., "Improved State-level Influenza Nowcasting in the United States Leveraging Internet-based Data and Network Approaches," *Nat Commun*, 10.1 (2019): 147.

[15] Clemente, L.; Lu, F.; Santillana, M., "Improved Real-Time Influenza Surveillance: Using Internet Search Data in Eight Latin American Countries." *JMIR Public Health Surveill* 5.2 (2019): e12214.

[16] Pei, S.; Shaman, J., Counteracting structural errors in ensemble forecast of influenza outbreaks. Nat Commun 2017, 8 (1), 925.

[17] Yang, W.; Cowling, B. J.; Lau, E. H.; Shaman, J., Forecasting Influenza Epidemics in Hong Kong. PLoS Comput Biol 2015, 11 (7), e1004383.

B, as well as 19 epidemics caused by a combination of these strains.

Axelsen, J. B., et al. utilized a basic epidemiological model to demonstrate the multiyear predictability of influenza outbreaks using high-quality surveillance data for Israel.[18] The study confirmed the model's accuracy through metapopulation comparisons within Israel. They found that successful forecasting relied on factors such as temperature, humidity, antigenic drift, and immunity loss.

Dukic, V., et al. incorporated a classical mathematical epidemiology model, specifically a susceptible-exposed-infected-recovered (SEIR) model, into the state-space framework, which allowed for changes in SEIR dynamics over time.[19] The authors used a particle filtering algorithm to implement this model, which learned about the epidemic process sequentially over time and provided updated odds of a pandemic with each new surveillance data point.

Trawicki, M. suggested a novel SEIRS model that generalizes various classical deterministic epidemic models, including SIR, SIS, SEIR, and SEIRS.[20] The SEIRS model incorporated vital dynamics with different birth and death rates, vaccinations for both newborns and non-newborns, and temporary immunity.

Hill, E. M., et al. combined multiple data sources to calibrate a transmission model for seasonal influenza that includes susceptible, latent, infected, and recovered compartments.[21] The model incorporated the four main influenza strains and mechanisms that linked prior season epidemiological outcomes to immunity at the beginning of the following season.

Postnikov, E. B demonstrated that the SIRS (Susceptible-Infected-Recovered-

[18] Axelsen, J. B.; Yaari, R.; Grenfell, B. T.; Stone, L., "Multiannual Forecasting of Seasonal Influenza Dynamics Reveals Climatic and Evolutionary Drivers." *Proc Natl Acad Sci U S A* 111.26 (2014): 9538-42.

[19] Dukic, V.; Lopes, H. F.; Polson, N. G., "Tracking Epidemics With Google Flu Trends Data and a State-Space SEIR Model." *Journal of the American Statistical Association* 107.500 (2012): 1410-1426.

[20] Trawicki, M., "Deterministic Seirs Epidemic Model for Modeling Vital Dynamics, Vaccinations, and Temporary Immunity," *Mathematics* 5.1 (2017).

[21] Hill, E. M.; Petrou, S.; de Lusignan, S.; Yonova, I.; Keeling, M. J., "Seasonal influenza: Modelling Approaches to Capture Immunity Propagation." *PLoS Comput Biol* 15.10 (2019): e1007096.

Susceptible) model accurately reproduced actual flu activity curves.[22] This model contained a variable reaction rate, which was dependent on the mean daily temperature. By representing the SIRS equations as a second-order ODE with an outer excitation, the authors explained the origin of the model's predictive efficiency and analytically justify the 1:1 dynamical resonance, which was a crucial property of epidemic behavior.

4.4 Hybrid method-based approach

In addition to the above methods, several researchers have explored ways to mix the different methods. Scarpino, S. V., et al. utilized a versatile statistical framework to evaluate the efficacy of traditional (ILINet) and advanced (BioSense 2.0 and Google Flu Trends) surveillance systems for assessing the incidence of influenza across different poverty levels.[23] The study used multiple data sources and statistical methods to identify the strengths and weaknesses of each system for situational awareness of influenza.

Baltrusaitis, K., et al. conducted an analysis of the Flu Near You (FNY) participant demographics during the 2014-2015 flu season and compared them to the general US population in terms of sex, age, and Human Development Index (HDI) scores.[24] They also studied the relationship between participant follow-up and demographic and behavioral factors. Additionally, they calculated descriptive statistics of responses from FNY's 2015 and 2016 end-of-season user surveys.

Santillana, M., et al. developed a machine learning-based methodology that utilizes data from various sources such as Google searches, Twitter microblogs, hospital visit records, and participatory surveillance system to provide real-time and forecast estimates of influenza activity in the US.[25]

---

22 Postnikov, E. B.; Tatarenkov, D. V., "Prediction of Flu Epidemic Activity with Dynamical Model Based on Weather Forecast." *Ecol. Complex*. 15 (2013): 109-113.

23 Scarpino, S. V.; Scott, J. G.; Eggo, R. M.; Clements, B.; Dimitrov, N. B.; Meyers, L. A., "Socioeconomic Bias in Influenza Surveillance," *PLoS Comput Biol* 16.7 (2020): e1007941.

24 Baltrusaitis, K.; Santillana, M.; Crawley, A. W.; Chunara, R.; Smolinski, M.; Brownstein, J. S., "Determinants of Participants' Follow-Up and Characterization of Representativeness in Flu Near You, A Participatory Disease Surveillance System." *JMIR Public Health Surveill* 3.2 (2017): e18.

25 Santillana, M.; Nguyen, A. T.; Dredze, M.; Paul, M. J.; Nsoesie, E. O.; Brownstein, J. S., "Combining

Pawelek, K. A., et al. have created a mathematical model to study the impact of tweets on the transmission of influenza among a population.[26] The model considers the dynamics of short Twitter messages and how they can increase disease awareness, change behavior, and decrease transmission. The researchers derive the basic reproductive number of the model and prove the stability of the steady states. They find that a threshold curve is crossed, causing a Hopf bifurcation and the possibility of multiple outbreaks of influenza.

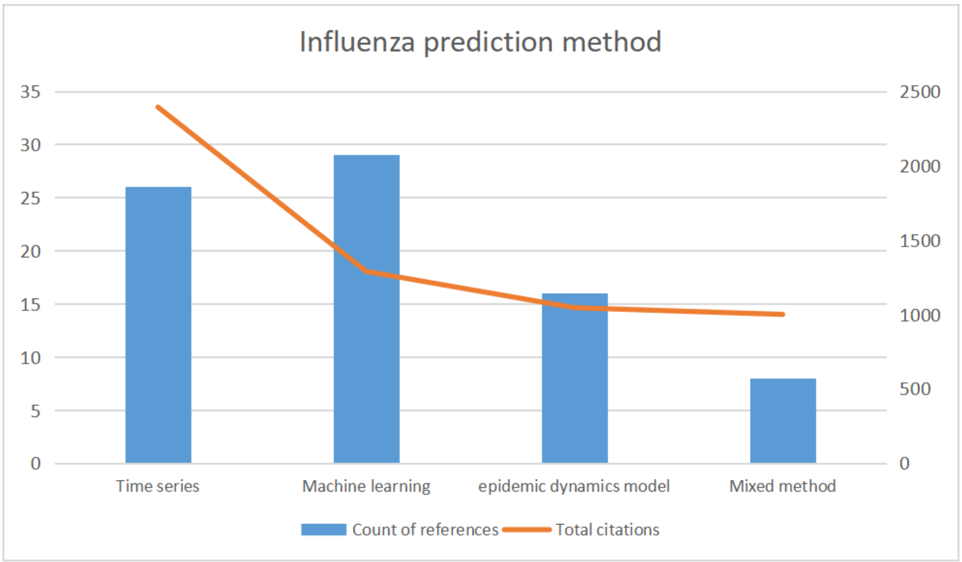
Table2. Comparison of Influenza prediction methods

Influenza prediction method	Count of references	Total citations	References
Time series	26	2393	1，5，8，13，18，28，33，37，39，42，44，46，48，52，53，54，58，61，64，66，68，69，73，76，82，84
Machine learning	29	1289	2，3，4，6，9，12，14，15，16，23，24，25，26，27，29，31，35，38，40，41，45，49，50，59，62，70，74，75，83
epidemic dynamics model	16	1045	7，10，11，17，19，20，21，22，51，56，57，60，63，65，71，81
Mixed method	8	1000	30，32，34，36，47，55，67，72

Search, Social Media, and Traditional Data Sources to Improve Influenza Surveillance." *PLoS Comput Biol* 11.10 (2015): e1004513.

26 Pawelek, K. A.; Oeldorf-Hirsch, A.; Rong, L., "Modeling the Impact of Twitter on Influenza Epidemics." *Math Biosci Eng* 11.6 (2014): 1337-56.

Figure2. Comparison of Influenza prediction methods



5. Accuracy

The accuracy of prediction methods can be assessed using statistical measures such as root mean square error (RMSE), mean absolute error (MAE), mean squared error (MSE), mean absolute percentage Error (MAPE), and Pearson correlation coefficient (PCC), depending on the type of prediction, whether classification or regression.

Dukic et al. reported that Dante outperformed the Dynamic Bayesian Model (DBM) in terms of accuracy, as measured by mean squared error (MSE) of point predictions (posterior means), at all geographic scales.[27] In Catalonia, the accuracy of the method was measured by calculating the mean squared error (MSE) of the predicted rates at one and two weeks ahead for the five different models (ARIMA, LM, GLS, FLM, and FGLS) and for the seven regions in local region.[28]

Su et al. achieved a Mean Absolute Percentage Error (MAPE) of 11.9% between 2014 and 2016, and the study used a specialized bootstrap strategy for time series to obtain 95% prediction intervals, which enclosed 96.2% of the true ILI% data points from 2017 to

---

27 Osthus, D.; Moran, K. R., "Multiscale Influenza Forecasting," *Nat Commun* 12.1 (2021): 2991.

28 Basile, L.; Oviedo de la Fuente, M.; Torner, N.; Martinez, A.; Jane, M., "Real-time Predictive Seasonal Influenza Model in Catalonia, Spain." *PLoS One* 13.2 (2018): e0193651.

2018.[29]

The linear regression model achieved the highest accuracy results using the measure of Pearson correlation for the module of weekly flu rate estimation, demonstrating a strong correlation of 96.29% with available data from Centers for Disease Control (CDC) as the ground truth.[30] Pearson's correlation coefficient (PCC) was 0.931. The other tested queries were based on structured data.[31]

Some articles compare measurement accuracy with many metrics. Yang et al. used the R-value, explained variance scores, MAE, and MSE to evaluate the quality of the models, with the highest correlation coefficients found for the Baidu search data for ILI% and for air quality for ILI% × positive%.[32] Other studies used machine learning algorithms and qualitative measures to improve short-term predictions of flu activity, with MAE values ranging from 0.016 to 0.034 and RMSE values ranging from 0.021 to 0.047, indicating that their models outperformed traditional linear regression models in terms of prediction accuracy.[33] The TL-based model leveraging the combination of ILI surveillance data, weather data, and Twitter data achieved the best performance, with an RMSE of 0.128 and a PCC of 0.822.[34]

---

29 Su, K.; Xu, L.; Li, G.; Ruan, X.; Li, X.; Deng, P.; Li, X.; Li, Q.; Chen, X.; Xiong, Y.; Lu, S.; Qi, L.; Shen, C.; Tang, W.; Rong, R.; Hong, B.; Ning, Y.; Long, D.; Xu, J.; Shi, X.; Yang, Z.; Zhang, Q.; Zhuang, Z.; Zhang, L.; Xiao, J.; Li, Y., "Forecasting Influenza Activity Using Self-adaptive AI Model and Multi-source Data in Chongqing, China." *EBioMedicine* 47 (2019): 284-292.

30 Alessa, A.; Faezipour, M., "Flu Outbreak Prediction Using Twitter Posts Classification and Linear Regression With Historical Centers for Disease Control and Prevention Reports: Prediction Framework Study," *JMIR Public Health Surveill* 5.2 (2019): e12383.

31 Bouzille, G.; Poirier, C.; Campillo-Gimenez, B.; Aubert, M. L.; Chabot, M.; Chazard, E.; Lavenu, A.; Cuggia, M., "Leveraging hospital big data to monitor flu epidemics. Comput," *Meth. Programs Biomed.* 154 (2018): 153-160.

32 Yang, L.; Li, G.; Yang, J.; Zhang, T.; Du, J.; Liu, T.; Zhang, X.; Han, X.; Li, W.; Ma, L.; Feng, L.; Yang, W., "Deep-Learning Model for Influenza Prediction From Multisource Heterogeneous Data in a Megacity: Model Development and Evaluation." *J Med Internet Res* 25 (2023): e44238.

33 Xue, H.; Bai, Y.; Hu, H.; Liang, H., "Regional Level Influenza Study Based on Twitter and Machine Learning Method," *PLoS One* 14.4 (2019): e0215600; Moss, R.; Zarebski, A.; Dawson, P.; Mc, C. J., "Retrospective Forecasting of the 2010-2014 Melbourne Influenza Seasons Using Multiple Surveillance Systems," *Epidemiol Infect* 145.1 (2017): 156-169.

34 Athanasiou, M.; Fragkozidis, G.; Zarkogianni, K.; Nikita, K. S., "Long Short-term Memory-Based Prediction of the Spread of Influenza-Like Illness Leveraging Surveillance, Weather, and Twitter Data:

In summary, evaluating the accuracy of prediction methods requires various statistical measures and methods, depending on the type of prediction. Cross-validation or holdout validation sets are necessary to avoid overfitting and ensure the model generalizes well to unseen data. Many studies have used different metrics and approaches to assess the accuracy of their models, such as correlation analysis, regression analysis, time series modeling, and machine learning algorithms. The use of specialized strategies and multiple measures can provide a more comprehensive evaluation of the model's performance and contribute to the development of more accurate prediction methods.

6. Limitation

At present, the shortcomings of influenza prediction research are mainly divided into three aspects: data, model, and application.

In terms of data, the current prediction research has two problems, such as low data quality and unstable source. The low quality of prediction data is mainly reflected in the lack of time dimension, single data type, and poor representativeness. Gabriel J Milinovich et al, based on seasonal influenza data in the last 5 years, may not monitor the same seasonal data of the same subtype, increasing the difficulty of modeling. Regarding the data source, the lack of laboratory data is an important factor limiting the accuracy of the model.[35] Most of the current studies were the percentage of influenza-like cases (ILI%) monitored by local or national CDC and the main limitation being the inability to distinguish the epidemic intensity of other non-influenza respiratory pathogens, such as respiratory syncytial virus (RSV), with clinical manifestations similar to that of influenza.[36] The lack of real nucleic

Model Development and Validation." *J Med Internet Res* 25 (2023): e42519.

[35] Milinovich, G. J.; Williams, G. M.; Clements, A. C.; Hu, W., "Internet-based Surveillance Systems for Monitoring Emerging Infectious Diseases." *Lancet Infect Dis* 14.2 (2014): 160-8; Polgreen, P. M.; Chen, Y.; Pennock, D. M.; Nelson, F. D., "Using Internet Searches for Influenza Surveillance." *Clin Infect Dis* 47.11 (2008): 1443-8; Li, R.; Bai, Y.; Heaney, A.; Kandula, S.; Cai, J.; Zhao, X.; Xu, B.; Shaman, J., "Inference and forecast of H7N9 influenza in China, 2013 to 2015," *Euro Surveill* 22.7 (2017).

[36] Paul, S.; Mgbere, O.; Arafat, R.; Yang, B.; Santos, E., "Modeling and Forecasting Influenza-like Illness (ILI) in Houston, Texas Using Three Surveillance Data Capture Mechanisms." *Online J Public Health Inform* 9.2 (2017): e187; Yin, R.; Tran, V. H.; Zhou, X.; Zheng, J.; Kwoh, C. K., "Predicting Antigenic

acid detection data in influenza prediction studies can lead to the problem of low lag and low specificity of prediction results. Moreover, these studies lack data considering antigen detection and serological testing, which prevents the prediction results from reflecting the ability of current seasonal vaccination to protect against epidemic strains, thus affecting the accuracy of the prediction model. In addition, Mauricio Santillana et al. have predicted data from specific geographic locations (such as Greece or two US states) to forecast the outbreaks of influenza in additional states and counties.[37] If it is good for national prediction, sampling error will affect the effect of model prediction. To enhance the generalization of prediction models, it is advantageous to make predictions based on publicly available databases. At present, some of the prediction data are from private companies, which are faced with the problem of limited data update at any time or limited access, which poses a great challenge in the process of model promotion.[38]

Variants of H1N1 Influenza Virus Based on Epidemics and Pandemics Using a Stacking Model." *PLoS One* 13.12 (2018): e0207777; Dong, X.; Boulton, M. L.; Carlson, B.; Montgomery, J. P.; Wells, E. V., "Syndromic surveillance for influenza in Tianjin, China: 2013-14." *J Public Health (Oxf)* 39.2 (2017): 274-281; Agor, J. K.; Ozaltin, O. Y., "Models for predicting the evolution of influenza to inform vaccine strain selection." *Hum Vaccin Immunother* 14.3 (2018): 678-683.

[37] Santillana, M.; Nguyen, A. T.; Dredze, M.; Paul, M. J.; Nsoesie, E. O.; Brownstein, J. S., "Combining Search, Social Media, and Traditional Data Sources to Improve Influenza Surveillance." *PLoS Comput Biol* 11.10 (2015): e1004513; Scarpino, S. V.; Scott, J. G.; Eggo, R. M.; Clements, B.; Dimitrov, N. B.; Meyers, L. A., "Socioeconomic Bias in Influenza Surveillance." *PLoS Comput Biol* 16.7 (2020): e1007941; Xu, J.; Wu, Q., "Prediction on Influenza-like Virus Pathogen and its Effects on Prognosis of Patients with Community Acquired Pneumonia under Long and Short Term Memory Neural Network Model." *Results in Physics* 24 (2021); Morita, H.; Kramer, S.; Heaney, A.; Gil, H.; Shaman, J., "Influenza Forecast Optimization when Using Different Surveillance Data Types and Geographic Scale." *Influenza Other Respir Viruses* 12.6 (2018): 755-764; Zhang, Y.; Bambrick, H.; Mengersen, K.; Tong, S.; Hu, W., "Using Google Trends and ambient temperature to predict seasonal influenza outbreaks." *Environ Int* 117 (2018): 284-291; Agor, J. K.; Ozaltin, O. Y., "Models for predicting the evolution of influenza to inform vaccine strain selection." *Hum Vaccin Immunother* 14.3 (2018): 678-683; Venkatramanan, S.; Sadilek, A.; Fadikar, A.; Barrett, C. L.; Biggerstaff, M.; Chen, J.; Dotiwalla, X.; Eastham, P.; Gipson, B.; Higdon, D.; Kucuktunc, O.; Lieber, A.; Lewis, B. L.; Reynolds, Z.; Vullikanti, A. K.; Wang, L.; Marathe, M., "Forecasting Influenza Activity Using Machine-Learned Mobility Map." *Nat Commun* 12.1 (2021): 726.

[38] Zimmer, C.; Leuba, S. I.; Yaesoubi, R.; Cohen, T., "Use of Daily Internet Search Query Data Improves Real-time Projections of Influenza Epidemics." *J R Soc Interface* 15.147 (2018); Yang, L.; Li, G.; Yang, J.; Zhang, T.; Du, J.; Liu, T.; Zhang, X.; Han, X.; Li, W.; Ma, L.; Feng, L.; Yang, W., "Deep-Learning Model for Influenza Prediction from Multisource Heterogeneous Data in a Megacity: Model Development and Evaluation." *J Med Internet Res* 25 (2023): e44238; Schneider, P. P.; Van Gool, C. J.; Spreeuwenberg, P.;

Data processing, modeling and model validation are all an important part of establishing a reliable influenza model. If current influenza prediction studies apply to multiple time series, different data streams may be measured at different time intervals, and they are often standardized for convenience, which may lead to loss of key information and reduce timeliness.[39] In addition, J. Zhang and K. Nawata first predicted multiple data features (such as temperature and humidity weather), and then further constructs influenza ILI% prediction model based on the above prediction data.[40] The mechanism is like MSP, leading to the accumulation of prediction error step by step. Choosing the appropriate prediction model based on the research purpose is the focus of improving the research quality. Multiple influenza prediction studies reflect limitations such as discomfort for emerging infectious diseases, low accuracy, limited observed abundance and availability, and no advantages compared with previous teamwork.[41] Compared with traditional mathematical models, disease transmission dynamics is a certain advantage. This model can quantify the population according to the natural history of infectious disease transmission in infection, severe disease, recovery, and also make subgroup deduction according to the age composition of the population, so as to obtain a more realistic mathematical model of the epidemic.

Hooiveld, M.; Donker, G. A.; Barnett, D. J.; Paget, J., "Using Web Search Queries to Monitor Influenza-like Illness: An Exploratory Retrospective Analysis, Netherlands, 2017/18 Influenza Season." *Euro Surveill* 25.21 (2020); Alessa, A.; Faezipour, M., "A Review of Influenza Detection and Prediction through Social Networking Sites." *Theor Biol Med Model* 15.1 (2018): 2.

39 Dukic, V.; Lopes, H. F.; Polson, N. G., "Tracking Epidemics With Google Flu Trends Data and a State-Space SEIR Model." *Journal of the American Statistical Association* 107.500 (2012): 1410-1426; Sebastiani, P.; Mandl, K. D.; Szolovits, P.; Kohane, I. S.; Ramoni, M. F., "A Bayesian Dynamic Model for Influenza Surveillance." *Stat Med* 25.11 (2006): 1803-16; discussion: 1817-25.

40 Zhang, J.; Nawata, K., "Multi-step prediction for influenza outbreak by an adjusted long short-term memory." *Epidemiol Infect* 146.7 (2018): 809-816.

41 Aiken, E. L.; Nguyen, A. T.; Viboud, C.; Santillana, M., "Toward the Use of Neural Networks for Influenza Prediction at Multiple Spatial Resolutions." *Sci. Adv*. 7.25 (2021): 13; Yamana, T. K.; Kandula, S.; Shaman, J., "Individual Versus Superensemble Forecasts of Seasonal Influenza Outbreaks in the United States." P*LoS Comput Biol* 13.11 (2017): e1005801; Pei, S.; Shaman, J., "Counteracting Structural Errors in Ensemble Forecast of Influenza Outbreaks." *Nat Commun* 8.1 (2017): 925; Cheng, H. Y.; Wu, Y. C.; Lin, M. H.; Liu, Y. L.; Tsai, Y. Y.; Wu, J. H.; Pan, K. H.; Ke, C. J.; Chen, C. M.; Liu, D. P.; Lin, I. F.; Chuang, J. H., "Applying Machine Learning Models with An Ensemble Approach for Accurate Real-Time Influenza Forecasting in Taiwan: Development and Validation Study." *J Med Internet Res* 22.8 (2020): e15394.

Influenza forecasting is ultimately used to support public health decisions. Accurate prediction and prediction-based prevention and control recommendations enable health authorities to make a timely and early response. However, the current research generally reflects the problem of low application, partly based on the relevant studies developed from hospital electronic cases or local health departments, and lack of regional universality verification, making it difficult to achieve flexible retrieval of data from different geographic regions.[42] In addition, the current is mostly for seasonal influenza research, the upcoming outbreak of influenza epidemic research is relatively weak.

7. Conclusion

Reliable predictions of indicators such as trends, peak, and duration during influenza outbreaks can provide a basis for resource allocation in medical institutions. Therefore, government can prepare for a surge in influenza cases by obtaining necessary resources such as vaccines and medical staffs such as nurses and doctors. However, predictions must be interpretable and effective to be useful. Therefore, research must clearly define the temporal and spatial applicability of the predicted events and methods, quantify the likelihood of events in terms of probability statements or relative to other similar events, and emphasize their limitations. In addition, defining a global accuracy measure for evaluating the correctness of various prediction methods will simplify the process of comparing predictions. Thus, challenges still exist in the real-time evaluation and quantification of the performance of these methods.

---

[42] Murayama, T.; Shimizu, N.; Fujita, S.; Wakamiya, S.; Aramaki, E., "Robust Two-stage Influenza Prediction Model Considering Regular and Irregular Trends." *PLoS One* 15.5 (2020): e0233126; Lu, F. S.; Hattab, M. W.; Clemente, C. L.; Biggerstaff, M.; Santillana, M., "Improved State-level Influenza Nowcasting in the United States Leveraging Internet-based Data and Network Approaches." *Nat Commun* 10.1 (2019): 147; Clemente, L.; Lu, F.; Santillana, M., "Improved Real-Time Influenza Surveillance: Using Internet Search Data in Eight Latin American Countries." J*MIR Public Health Surveill* 5.2 (2019): e12214; Bouzille, G.; Poirier, C.; Campillo-Gimenez, B.; Aubert, M. L.; Chabot, M.; Chazard, E.; Lavenu, A.; Cuggia, M., "Leveraging Hospital Big Data to Monitor Flu Epidemics. Comput. Meth." *Programs Biomed*. 154 (2018): 153-160; Osthus, D.; Moran, K. R., "Multiscale Influenza Forecasting." *Nat Commun* 12.1 (2021): 2991.

Bibliography (Reference to table 1-2)

[1]. Reich, N. G.; Brooks, L. C.; Fox, S. J.; Kandula, S.; McGowan, C. J.; Moore, E.; Osthus, D.; Ray, E. L.; Tushar, A.; Yamana, T. K.; Biggerstaff, M.; Johansson, M. A.; Rosenfeld, R.; Shaman, J., "A Collaborative Multiyear, Multimodel Assessment of Seasonal Influenza Forecasting in the United States." *Proc Natl Acad Sci U S A* 116.9 (2019): 3146-3154.

[2]. Jiang-Ning, L.; Xian-Liang, S.; An-Qiang, H.; Ze-Fang, H.; Yu-Xuan, K.; Dong, L., "Forecasting Emergency Medicine Reserve Demand with a Novel Decomposition-Ensemble Methodology." *Complex Intell Systems* (2021):1-11.

[3]. Paul, S.; Mgbere, O.; Arafat, R.; Yang, B.; Santos, E., "Modeling and Forecasting Influenza-like Illness (ILI) in Houston, Texas Using Three Surveillance Data Capture Mechanisms." *Online J Public Health Inform* 9.2 (2017): e187.

[4]. Kandula, S.; Shaman, J., "Near-term Forecasts of Influenza-like Illness: An Evaluation of Autoregressive Time Series Approaches." *Epidemics* 27 (2019): 41-51.

[5]. Samaras, L.; Garcia-Barriocanal, E.; Sicilia, M. A., "Syndromic Surveillance Models Using Web Data: The Case of Influenza in Greece and Italy Using Google Trends." *JMIR Public Health Surveill* 3.4 (2017): e90.

[6]. Lee, K.; Ray, J.; Safta, C., "The Predictive Skill of Convolutional Neural Networks Models for Disease Forecasting." *PLoS One* 16.7 (2021): e0254319.

[7]. Pei, S.; Shaman, J., "Counteracting Structural Errors in Ensemble Forecast of Influenza Outbreaks." *Nat Commun* 8.1 (2017): 925.

[8]. Osthus, D.; Gattiker, J.; Priedhorsky, R.; Del Valle, S. Y., "Dynamic Bayesian Influenza Forecasting in the United States with Hierarchical Discrepancy (with Discussion)." *Bayesian Analysis* 14.1 (2019).

[9]. Su, K.; Xu, L.; Li, G.; Ruan, X.; Li, X.; Deng, P.; Li, X.; Li, Q.; Chen, X.; Xiong, Y.; Lu, S.; Qi, L.; Shen, C.; Tang, W.; Rong, R.; Hong, B.; Ning, Y.; Long, D.; Xu, J.; Shi, X.; Yang, Z.; Zhang, Q.; Zhuang, Z.; Zhang, L.; Xiao, J.; Li, Y., "Forecasting Influenza Activity Using Self-adaptive AI Model and Multi-source Data in Chongqing, China."

EBioMedicine 47 (2019): 284-292.

[10]. Yang, W.; Cowling, B. J.; Lau, E. H.; Shaman, J., "Forecasting Influenza Epidemics in Hong Kong." *PLoS Comput Biol* 11.7 (2015): e1004383.

[11]. Pei, S.; Kandula, S.; Yang, W.; Shaman, J., Forecasting the Spatial Transmission of Influenza in the United States." *Proc Natl Acad Sci U S A* 115.11 (2018): 2752-2757.

[12]. Dugas, A. F.; Jalalpour, M.; Gel, Y.; Levin, S.; Torcaso, F.; Igusa, T.; Rothman, R. E., "Influenza forecasting with Google Flu Trends," *PLoS One* 8.2 (2013): e56176.

[13]. Li, R.; Bai, Y.; Heaney, A.; Kandula, S.; Cai, J.; Zhao, X.; Xu, B.; Shaman, J., "Inference and Forecast of H7N9 Influenza in China, 2013 to 2015." *Euro Surveill* 22.7 (2017).

[14]. Venna, S. R.; Tavanaei, A.; Gottumukkala, R. N.; Raghavan, V. V.; Maida, A. S.; Nichols, S., "A Novel Data-Driven Model for Real-Time Influenza Forecasting." *IEEE Access* 7 (2019): 7691-7701.

[15]. Volkova, S.; Ayton, E.; Porterfield, K.; Corley, C. D., "Forecasting Influenza-like Illness Dynamics for Military Populations Using Neural Networks and Social Media." *PLoS One* 12.12 (2017): e0188941.

[16]. Zhang, J.; Nawata, K., "Multi-step Prediction for Influenza Outbreak by an Adjusted Long Short-term Memory." *Epidemiol Infect* 146.7 (2018): 809-816.

[17]. Axelsen, J. B.; Yaari, R.; Grenfell, B. T.; Stone, L., "Multiannual Forecasting of Seasonal Influenza Dynamics Reveals Climatic and Evolutionary Drivers." *Proc Natl Acad Sci U S A* 111.26 (2014): 9538-42.

[18]. Osthus, D.; Moran, K. R., "Multiscale Influenza Forecasting." *Nat Commun* 12.1 (2021): 2991.

[19]. Dukic, V.; Lopes, H. F.; Polson, N. G., "Tracking Epidemics With Google Flu Trends Data and a State-Space SEIR Model." *Journal of the American Statistical Association* 107.500 (2012): 1410-1426.

[20]. Trawicki, M., "Deterministic Seirs Epidemic Model for Modeling Vital Dynamics, Vaccinations, and Temporary Immunity." *Mathematics* 5.1 (2017).

[21]. Shao, P.; Shateyi, S., "Stability Analysis of SEIRS Epidemic Model with Nonlinear

Incidence Rate Function." *Mathematics* 9.21 (2021).

[22]. Osthus, D.; Hickmann, K. S.; Caragea, P. C.; Higdon, D.; Del Valle, S. Y., "Forecasting Seasonal Influenza with a State-space SIR model." *Ann Appl Stat* 11.1 (2017): 202-224.

[23]. Cheng, H. Y.; Wu, Y. C.; Lin, M. H.; Liu, Y. L.; Tsai, Y. Y.; Wu, J. H.; Pan, K. H.; Ke, C. J.; Chen, C. M.; Liu, D. P.; Lin, I. F.; Chuang, J. H., "Applying Machine Learning Models with An Ensemble Approach for Accurate Real-Time Influenza Forecasting in Taiwan: Development and Validation Study." *J Med Internet Res* 22.8 (2020): e15394.

[24]. Venkatramanan, S.; Sadilek, A.; Fadikar, A.; Barrett, C. L.; Biggerstaff, M.; Chen, J.; Dotiwalla, X.; Eastham, P.; Gipson, B.; Higdon, D.; Kucuktunc, O.; Lieber, A.; Lewis, B. L.; Reynolds, Z.; Vullikanti, A. K.; Wang, L.; Marathe, M., "Forecasting Influenza Activity Using Machine-learned Mobility Map." *Nat Commun* 12.1 (2021): 726.

[25]. Zhang, Y.; Bambrick, H.; Mengersen, K.; Tong, S.; Hu, W., "Using Google Trends and Ambient Temperature to Predict Seasonal Influenza Outbreaks." *Environ Int* 117 (2018): 284-291.

[26]. Yang, L.; Li, G.; Yang, J.; Zhang, T.; Du, J.; Liu, T.; Zhang, X.; Han, X.; Li, W.; Ma, L.; Feng, L.; Yang, W., "Deep-Learning Model for Influenza Prediction From Multisource Heterogeneous Data in a Megacity: Model Development and Evaluation." *J Med Internet Res* 25 (2023): e44238.

[27]. Soliman, M.; Lyubchich, V.; Gel, Y. R., "Complementing the Power of Deep Learning with Statistical Model Fusion: Probabilistic Forecasting of Influenza in Dallas County, Texas, USA." *Epidemics* 28 (2019): 100345.

[28]. Yamana, T. K.; Kandula, S.; Shaman, J., "Individual Versus Superensemble Forecasts of Seasonal Influenza Outbreaks in the United States." *PLoS Comput Biol* 13.11 (2017): e1005801.

[29]. Athanasiou, M.; Fragkozidis, G.; Zarkogianni, K.; Nikita, K. S., "Long Short-term Memory-Based Prediction of the Spread of Influenza-Like Illness Leveraging Surveillance, Weather, and Twitter Data: Model Development and Validation." *J Med*

Internet Res 25 (2023): e42519.

[30]. Alessa, A.; Faezipour, M., "A Review of Influenza Detection and Prediction through Social Networking Sites." *Theor Biol Med Model* 15.1 (2018): 2.

[31]. Ertem, Z.; Raymond, D.; Meyers, L. A., "Optimal Multi-source Forecasting of Seasonal Influenza." *PLoS Comput Biol* 14.9 (2018): e1006236.

[32]. Scarpino, S. V.; Scott, J. G.; Eggo, R. M.; Clements, B.; Dimitrov, N. B.; Meyers, L. A., "Socioeconomic Bias in Influenza Surveillance." *PLoS Comput Biol* 16.7 (2020): e1007941.

[33]. Escuyer, K. L.; Waters, C. L.; Gowie, D. L.; Maxted, A. M.; Farrell, G. M.; Fuschino, M. E.; St George, K., "The Assessment of Data Sources for Influenza Virologic Surveillance in New York State." *Influenza Other Respir Viruses* 11.2 (2017): 138-147.

[34]. Baltrusaitis, K.; Santillana, M.; Crawley, A. W.; Chunara, R.; Smolinski, M.; Brownstein, J. S., "Determinants of Participants' Follow-Up and Characterization of Representativeness in Flu Near You, A Participatory Disease Surveillance System." *JMIR Public Health Surveill* 3.2 (2017): e18.

[35]. Kandula, S.; Shaman, J., "Reappraising the utility of Google Flu Trends." *PLoS Comput Biol* 15.8 (2019): e1007258.

[36]. Dong, X.; Boulton, M. L.; Carlson, B.; Montgomery, J. P.; Wells, E. V., "Syndromic Surveillance for Influenza in Tianjin, China: 2013-14." *J Public Health (Oxf)* 39.2 (2017): 274-281.

[37]. Polgreen, P. M.; Chen, Y.; Pennock, D. M.; Nelson, F. D., "Using Internet Searches for Influenza Surveillance." *Clin Infect Dis* 47.11 (2008): 1443-8.

[38]. Alessa, A.; Faezipour, M., "Flu Outbreak Prediction Using Twitter Posts Classification and Linear Regression With Historical Centers for Disease Control and Prevention Reports: Prediction Framework Study." *JMIR Public Health Surveill* 5.2 (2019): e12383.

[39]. Xue, H.; Bai, Y.; Hu, H.; Liang, H., "Influenza Activity Surveillance Based on Multiple Regression Model and Artificial Neural Network." *IEEE Access* 6 (2018):

563-575.

[40]. Xu, J.; Wu, Q., "Prediction on Influenza-like Virus Pathogen and Its Effects on Prognosis of Patients with Community Acquired Pneumonia under Long and Short Term Memory Neural Network Model." *Results in Physics* 24 (2021).

[41]. Du, L.; Pang, Y., "A Novel Data-driven Methodology for Influenza Outbreak Detection and Prediction." *Sci Rep* 11.1 (2021): 13275.

[42]. Xue, H.; Bai, Y.; Hu, H.; Liang, H., "Regional Level Influenza Study Based on Twitter and Machine Learning Method." *PLoS One* 14.4 (2019): e0215600.

[43]. Agor, J. K.; Ozaltin, O. Y., "Models for Predicting the Evolution of Influenza to Inform Vaccine Strain Selection." *Hum Vaccin Immunother* 14.3 (2018): 678-683.

[44]. Lampos, V.; Miller, A. C.; Crossan, S.; Stefansen, C., "Advances in Nowcasting Influenza-like Illness Rates Using Search Query Logs." *Sci Rep* 5 (2015): 12760.

[45]. Yang, C.-T.; Chen, Y.-A.; Chan, Y.-W.; Lee, C.-L.; Tsan, Y.-T.; Chan, W.-C.; Liu, P.-Y., "Influenza-like Illness Prediction Using a Long Short-term Memory Deep Learning Model with Multiple Open Data Sources." *The Journal of Supercomputing* 76.12 (2020): 9303-9329.

[46]. Miliou, I.; Xiong, X.; Rinzivillo, S.; Zhang, Q.; Rossetti, G.; Giannotti, F.; Pedreschi, D.; Vespignani, A., "Predicting Seasonal Influenza Using Supermarket Retail Records." *PLoS Comput Biol* 17.7 (2021): e1009087.

[47]. Santillana, M.; Nguyen, A. T.; Dredze, M.; Paul, M. J.; Nsoesie, E. O.; Brownstein, J. S., "Combining Search, Social Media, and Traditional Data Sources to Improve Influenza Surveillance." *PLoS Comput Biol* 11.10 (2015): e1004513.

[48]. Davidson, M. W.; Haim, D. A.; Radin, J. M., "Using Networks to Combine 'big data' and Traditional Surveillance to Improve Influenza Predictions." *Sci Rep* 5 (2015): 8154.

[49]. Poirier, C.; Hswen, Y.; Bouzille, G.; Cuggia, M.; Lavenu, A.; Brownstein, J. S.; Brewer, T.; Santillana, M., "Influenza Forecasting for French Regions Combining EHR, Web and Climatic Data Sources with a Machine Learning Ensemble Approach." *PLoS One* 16.5 (2021): e0250890.

[50]. Murayama, T.; Shimizu, N.; Fujita, S.; Wakamiya, S.; Aramaki, E., "Predicting Regional Influenza Epidemics with Uncertainty Estimation using Commuting Data in Japan." *PLoS One* 16.4 (2021): e0250417.

[51]. Moss, R.; Zarebski, A.; Dawson, P.; McCaw, J. M., "Forecasting Influenza Outbreak Dynamics in Melbourne from Internet Search Query Surveillance Data." *Influenza Other Respir Viruses* 10.4 (2016): 314-23.

[52]. Schneider, P. P.; van Gool, C. J.; Spreeuwenberg, P.; Hooiveld, M.; Donker, G. A.; Barnett, D. J.; Paget, J., "Using Web Search Queries to Monitor Influenza-like Illness: An Exploratory Retrospective Analysis, Netherlands, 2017/18 Influenza Season." *Euro Surveill* 25.21 (2020).

[53]. Liang, F.; Guan, P.; Wu, W.; Huang, D., "Forecasting Influenza Eepidemics by Integrating Internet Search Queries and Traditional Surveillance Data with the Support Vector Machine Regression Model in Liaoning, from 2011 to 2015." *PeerJ* 6 (2018): e5134.

[54]. Moss, R.; Zarebski, A.; Dawson, P.; Mc, C. J., "Retrospective Forecasting of the 2010-2014 Melbourne Influenza Seasons Using Multiple Surveillance Systems." *Epidemiol Infect* 145.1 (2017): 156-169.

[55]. Pawelek, K. A.; Oeldorf-Hirsch, A.; Rong, L., "Modeling the Impact of Twitter on Influenza Epidemics." *Math Biosci Eng* 11.6 (2014): 1337-56.

[56]. Edlund, S.; Kaufman, J.; Lessler, J.; Douglas, J.; Bromberg, M.; Kaufman, Z.; Bassal, R.; Chodick, G.; Marom, R.; Shalev, V.; Mesika, Y.; Ram, R.; Leventhal, A., "Comparing Three Basic Models for Seasonal Influenza." *Epidemics* 3.3-4 (2011): 135-42.

[57]. Azuma, T.; Nakada, N.; Yamashita, N.; Tanaka, H., "Prediction, Risk and Control of Anti-influenza Drugs in the Yodo River Basin, Japan during Seasonal and Pandemic Influenza using the Transmission Model for Infectious Disease." *Sci Total Environ* 521-522 (2015): 68-74.

[58]. Sebastiani, P.; Mandl, K. D.; Szolovits, P.; Kohane, I. S.; Ramoni, M. F., "A Bayesian Dynamic Model for Influenza Surveillance." *Stat Med* 25.11 (2006): 1803-16;

discussion: 1817-25.

[59]. Yin, R.; Tran, V. H.; Zhou, X.; Zheng, J.; Kwoh, C. K., "Predicting Antigenic Variants of H1N1 Influenza Virus Based on Epidemics and Pandemics using a Stacking Model." *PLoS One* 13.12 (2018): e0207777.

[60]. Hickmann, K. S.; Fairchild, G.; Priedhorsky, R.; Generous, N.; Hyman, J. M.; Deshpande, A.; Del Valle, S. Y., "Forecasting the 2013-2014 Influenza Season using Wikipedia." *PLoS Comput Biol* 11.5 (2015): e1004239.

[61]. Brooks, L. C.; Farrow, D. C.; Hyun, S.; Tibshirani, R. J.; Rosenfeld, R., "Flexible Modeling of Epidemics with an Empirical Bayes Framework." *PLoS Comput Biol* 11.8 (2015): e1004382.

[62]. Lu, F. S.; Hattab, M. W.; Clemente, C. L.; Biggerstaff, M.; Santillana, M., "Improved State-level Influenza Nowcasting in the United States Leveraging Internet-based Data and Network Approaches." *Nat Commun* 10.1 (2019): 147.

[63]. Zimmer, C.; Leuba, S. I.; Yaesoubi, R.; Cohen, T., "Use of Daily Internet Search Query Data Improves Real-time Projections of Influenza Epidemics." *J R Soc Interface* 15.147 (2018).

[64]. Li, J. H.; Wu, C. C.; Tseng, Y. J.; Han, S. T.; Pekosz, A.; Rothman, R.; Chen, K. F., "Applying Symptom Dynamics to Accurately Predict Influenza Virus Infection: An International Multicenter Influenza-like Illness Surveillance Study." *Influenza Other Respir Viruses*17.1 (2023): e13081.

[65]. Morita, H.; Kramer, S.; Heaney, A.; Gil, H.; Shaman, J., "Influenza Forecast Optimization when using Different Surveillance Data Types and Geographic Scale." *Influenza Other Respir Viruses* 12.6 (2018): 755-764.

[66]. Basile, L.; Oviedo de la Fuente, M.; Torner, N.; Martinez, A.; Jane, M., "Real-time Predictive Seasonal Influenza Model in Catalonia, Spain." *PLoS One* 13.3 (2018): e0193651.

[67]. Amorós, R.; Conesa, D.; López-Quílez, A.; Martinez-Beneito, M.-A., "A Spatio-temporal Hierarchical Markov Switching Model for the Early Detection of Influenza Outbreaks." *Stochastic Environmental Research and Risk Assessment* 34.2 (2020):

275-292.

[68]. Pollett, S.; Boscardin, W. J.; Azziz-Baumgartner, E.; Tinoco, Y. O.; Soto, G.; Romero, C.; Kok, J.; Biggerstaff, M.; Viboud, C.; Rutherford, G. W., "Evaluating Google Flu Trends in Latin America: Important Lessons for the Next Phase of Digital Disease Detection." *Clin Infect Dis* 64.1 (2017): 34-41.

[69]. Goldstein, E.; Cobey, S.; Takahashi, S.; Miller, J. C.; Lipsitch, M., "Predicting the Epidemic Sizes of Influenza A/H1N1, A/H3N2, and B: A Statistical Method." *PLoS Med* 8.7 (2011): e1001051.

[70]. Aldhyani, T. H. H.; Joshi, M. R.; AlMaaytah, S. A.; Alqarni, A. A.; Alsharif, N.; Uddin, M. I., "Using Sequence Mining to Predict Complex Systems: A Case Study in Influenza Epidemics." *Complexity* 2021 (2021): 1-16.

[71]. Hill, E. M.; Petrou, S.; de Lusignan, S.; Yonova, I.; Keeling, M. J., "Seasonal Influenza: Modelling Approaches to Capture Immunity Propagation." *PLoS Comput Biol* 15.10 (2019): e1007096.

[72]. Milinovich, G. J.; Williams, G. M.; Clements, A. C.; Hu, W., "Internet-based Surveillance Systems for Monitoring Emerging Infectious Diseases." *Lancet Infect Dis* 14.2 (2014): 160-8.

[73]. Guolo, A.; Varin, C., "Beta Regression for Time Series Analysis of Bounded Data, with Application to Canada Google® Flu Trends." *The Annals of Applied Statistics* 8.1 (2014).

[74]. Lu, J.; Meyer, S., "Forecasting Flu Activity in the United States: Benchmarking an Endemic-Epidemic Beta Model." *Int J Environ Res Public Health* 17.4 (2020).

[75]. Clemente, L.; Lu, F.; Santillana, M., "Improved Real-Time Influenza Surveillance: Using Internet Search Data in Eight Latin American Countries." *JMIR Public Health Surveill* 5.2 (2019): e12214.

[76]. Murayama, T.; Shimizu, N.; Fujita, S.; Wakamiya, S.; Aramaki, E., "Robust Two-stage Influenza Prediction Model Considering Regular and Irregular Trends." *PLoS One* 15.5 (2020): e0233126.

[77]. Chretien, J. P.; George, D.; Shaman, J.; Chitale, R. A.; McKenzie, F. E., "Influenza

Forecasting in Human Populations: A Scoping Review." *PLoS One* 9.4 (2014): e94130.

[78]. Nsoesie, E. O.; Brownstein, J. S.; Ramakrishnan, N.; Marathe, M. V., "A Systematic Review of Studies on Forecasting the Dynamics of Influenza Outbreaks." *Influenza Other Respir Viruses* 8.3 (2014): 309-16.

[79]. Ali, S. T.; Cowling, B. J., "Influenza Virus: Tracking, Predicting, and Forecasting." *Annu Rev Public Health* 42 (2021): 43-57.

[80]. Christaki, E., "New Technologies in Predicting, Preventing and Controlling Emerging Infectious Diseases." *Virulence* 6.6 (2015): 558-65.

[81]. Postnikov, E. B.; Tatarenkov, D. V., "Prediction of Flu Epidemic Activity with Dynamical Model Based on Weather Forecast." *Ecol. Complex.* 15 (2013): 109-113.

[82]. Bouzille, G.; Poirier, C.; Campillo-Gimenez, B.; Aubert, M. L.; Chabot, M.; Chazard, E.; Lavenu, A.; Cuggia, M., "Leveraging Hospital Big Data to Monitor Flu Epidemics. " *Comput. Meth. Programs Biomed.* 154 (2018): 153-160.

[83]. Aiken, E. L.; Nguyen, A. T.; Viboud, C.; Santillana, M., "Toward the Use of Neural Networks for Influenza Prediction at Multiple Spatial Resolutions." *Sci. Adv.* 7.25 (2021): 13.

[84]. Shaman, J.; Karspeck, A.; Yang, W.; Tamerius, J.; Lipsitch, M., "Real-time Influenza Forecasts during the 2012-2013 Season." *Nat. Commun.* 4 (2013): 10.

This study was funded in part by the following agencies:

National Key Research and Development Program of China (No. 2022YFC2600705);

Self-supporting Program of Guangzhou Laboratory，Grant No. SRPG22-007;

Science and Technology Development Fund of Macau SAR (005/2022/ALC);

Science and Technology Program of Guangzhou (No. 2022B01W0003);

Science and Technology Program of Guangzhou (Grant No. 202102100003);

Science and Technology Development Fund of Macau SAR (0045/2021/A);

Macau University of Science and Technology（FRG-20-021-MISE）.

Competing Interests Statement

We declare no competing interests.

*** Corresponding Authors information:**

Dr. Chitin Hon.

Department of Engineering Science, Faculty of Innovation Engineering, Macau University of Science and Technology, Taipa, Macau, China.

Email: cthon@must.edu.mo

Dr. Zifeng Yang.

National Clinical Research Center for Respiratory Disease, State Key Laboratory of Respiratory Disease, Guangzhou Institute of Respiratory Health, the First Affiliated Hospital of Guangzhou Medical University, 151 Yanjiang Road, Guangzhou, Guangdong, 510120, China.

Email: Jeffyah@163.com

Dr. Zhiqi Zeng.

State Key Laboratory of Respiratory Disease, National Clinical Research Center for Respiratory Disease, Guangzhou Institute of Respiratory Health, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China.

Email: zeng_zhiq@126.com