



Global infectious disease early warning models: An updated review and lessons from the COVID-19 pandemic



Wei-Hua Hu^a, Hui-Min Sun^a, Yong-Yue Wei^{a, b, c}, Yuan-Tao Hao^{a, b, c, *}

^a Department of Epidemiology and Biostatistics, School of Public Health, Peking University, 38 Xueyuan Road, Beijing, 100191, China

^b Peking University Center for Public Health and Epidemic Preparedness & Response, Peking University, 38 Xueyuan Road, Beijing, 100191, China

^c Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, China

ARTICLE INFO

Article history:

Received 14 March 2024

Received in revised form 29 August 2024

Accepted 1 December 2024

Available online 3 December 2024

Handling Editor: Dr Yijun Lou

Keywords:

Infectious disease

Early warning model

Combined model

Hawkes model

ABSTRACT

An early warning model for infectious diseases is a crucial tool for timely monitoring, prevention, and control of disease outbreaks. The integration of diverse multi-source data using big data and artificial intelligence techniques has emerged as a key approach in advancing these early warning models. This paper presents a comprehensive review of widely utilized early warning models for infectious diseases around the globe. Unlike previous review studies, this review encompasses newly developed approaches such as the combined model and Hawkes model after the COVID-19 pandemic, providing a thorough evaluation of their current application status and development prospects for the first time. These models not only rely on conventional surveillance data but also incorporate information from various sources. We aim to provide valuable insights for enhancing global infectious disease surveillance and early warning systems, as well as informing future research in this field, by summarizing the underlying modeling concepts, algorithms, and application scenarios of each model.

© 2024 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Infectious disease surveillance serves as a crucial epidemiological tool for monitoring the population's health. The goals of infectious disease surveillance are to describe the current burden and epidemiology of diseases, monitor trends, and identify outbreaks and new pathogens (Murray & Cohen, 2017). Internationally, a diverse array of mainstream infectious disease surveillance systems have been developed to monitor and signal potential outbreaks, including the traditional case-based surveillance system and non-traditional case-, event-, lab-, syndromic-, web search-based surveillance systems (W. Z. Yang, 2017).

Infectious disease surveillance data show a trend toward multiple sources. Traditionally, infectious disease surveillance data comprised mortality and morbidity, clinical data, laboratory reports, relevant field investigations, surveys, animal or vector studies, and demographic and environmental data (Declich & Carter, 1994). However, traditional surveillance methods based on the aforementioned data are limited in timeliness and sensitivity, attributable to factors such as the lengthy data

* Corresponding author. Peking University Center for Public Health and Epidemic Preparedness & Response, Peking University, 38 Xueyuan Road, Beijing, 100191, China.

E-mail address: haoyt@bjmu.edu.cn (Y.-T. Hao).

validation process (Y. T. Yang, Horneffer, & DiLisio, 2013). In recent years, new data sources have emerged to complement traditional infectious disease surveillance. These sources include sensors and mobile apps for collecting symptom and disease data, internet surveys and crowd-sourced data, scenario-based field experiment data, and artificial intelligence (AI) in conjunction with social media and web search data for outbreak prediction (Shausan, Nazarathy, & Dyda, 2023). Using these data sources in surveillance poses advantages regarding timeliness and resource efficiency.

The terms “surveillance data” and “early warning” are closely connected: the former provides the foundation for the latter, and the latter is an essential application of the former. The early warning of infectious diseases refers to the issuance of warning signals before and during the early stages of an outbreak or epidemic to alert the potential risk of expansion using specific analytical methods or models (W. Z. Yang, 2017). The past two decades have seen major breakthroughs in early warning systems and methods for infectious diseases. Particularly, the COVID-19 pandemic accelerated the research and application of big data, such as internet search, social media, and mobile location, alongside sophisticated analytical methods for predicting outbreaks and assessing transmission risks based on surveillance data. Hence, conducting a systematic analysis of the latest early warning models for infectious diseases is crucial, which can not only improve our understanding of the most advanced available methods but also provide valuable insights for advancing and enhancing early warning systems and methods to effectively combat infectious diseases.

The pandemic and long-term evolution of emerging infectious disease outbreaks represented by COVID-19 pose challenges to infectious disease modeling methods (Jacobsen et al., 2016). However, it also promote the enrichment and development of such methods, such as Hawkes process and combined models (Chiang, Liu, & Mohler, 2022; Kaplan, Park, Kresin, & Schoenberg, 2022; Kelly et al., 2019; Lamprinakou, Gandy, & McCoy, 2023; Park, Chaffee, Harrigan, & Schoenberg, 2022; X. L. Shi, Wei, & Chen, 2023; W. P. Zhao, Sun, Li, & Guan, 2022). While several scholars have previously reviewed early warning models for infectious diseases, many of these reviews were conducted before the occurrence of COVID-19 and are in need of updating (Kobres et al., 2019; Lai et al., 2021; X. X. Zhang et al., 2022). The reviews of infectious disease methods in the last 5 years have tended to focus mainly on COVID-19 and have not updated the current types of early warning models in the field of infectious diseases (Chen, Liu, Yu, & Li, 2021; Deng et al., 2022; H. H. Shi, Wei, & Chen, 2023; Yadav & Akhter, 2021). Therefore, the aim of this study is to provide a comprehensive review of the current variety of early warning models in the field of infectious diseases, building on previous reviews and complementing those emerging methods that have been enriched and developed during COVID-19.

For this review, we conducted searches on PubMed with the terms “infectious disease,” “warning,” and “model.” The searches returned 682 papers related to models for the control of infectious disease. A total of 53 studies met the criteria and were included in this review. We reclassified the models described in the included articles, outlined their applications to infectious diseases, and analyzed the pros and cons of the models to propose a new classification scheme and a selection basis for future modeling research to further promote the use of models for the prevention and control of infectious diseases.

2. Models

We categorized infectious disease early warning models into two types: models mainly based on traditional surveillance data and those mainly based on multi-source surveillance data. The former includes temporal and/or spatial warning models and dynamic models, while the latter encompasses AI techniques and combined models. The model classification framework was shown in Fig. 1. The timeline of various models being proposed and introduced into the field of public health was shown in Fig. 2. Table 1 summarized the key points of the definition or principle, features or applicability of various models.

2.1. Models mainly based on traditional surveillance data

There are three common approaches to building models, primarily based on traditional monitoring data. (1) spatial and/or temporal warning models, (2) dynamic models and (3) Hawkes models. The spatial and/or temporal warning models aim to detect potential outbreak hotspots by analyzing spatial and/or temporal patterns. Dynamic models focus on modeling the transmission dynamics within a population. We can improve the parameters and assumptions in the dynamic model, enhancing the accuracy of predictions, by comprehending the spatial patterns of disease occurrence. Hawkes processes model the unevenly spaced self-exciting arrivals of events in time. However, time series data is often collected on a fixed phase. Discrete-time Hawkes processes can model the evenly spaced arrivals of events in time and capturing the self-exciting character, which are relatively new and many theoretical properties are yet to be studied.

2.1.1. Temporal warning models

Temporal warning models analyze the time-varying characteristics of historical surveillance data and utilize baseline data from the same period to predict future occurrences and trends. These models extract pertinent patterns from single time-series surveillance data and establish early warning systems for infectious diseases. Temporal warning models can be categorized into control chart-based models, time series models, regression models, Markov chains, and gray models based on differences in modeling approaches.

A control chart determines indicator warning thresholds based on historical data and issue warnings when actual monitoring data is “out of control.” It is suitable for all infectious diseases and is effective for infectious diseases with seasonal epidemics or periodic epidemic law (Siettos & Russo, 2013). A cumulative sum control chart gradually accumulates the

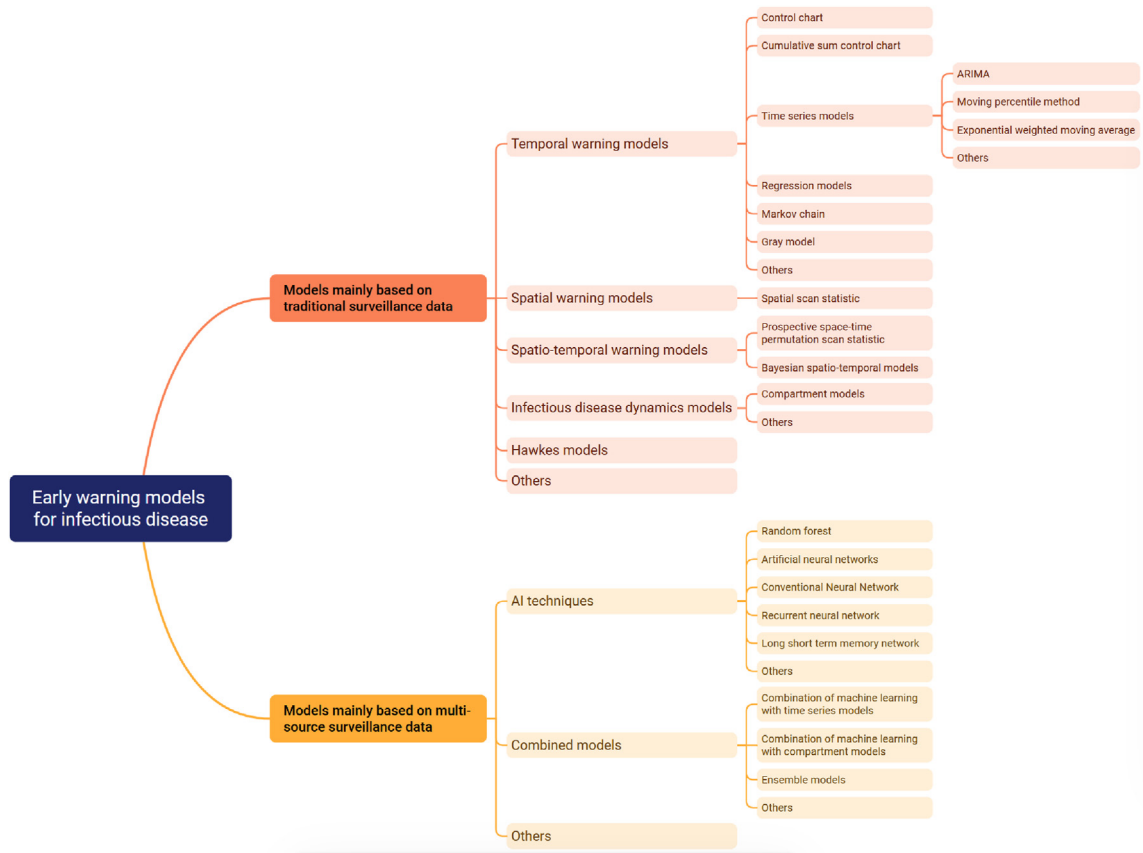


Fig. 1. Classification of early warning models for infectious diseases.

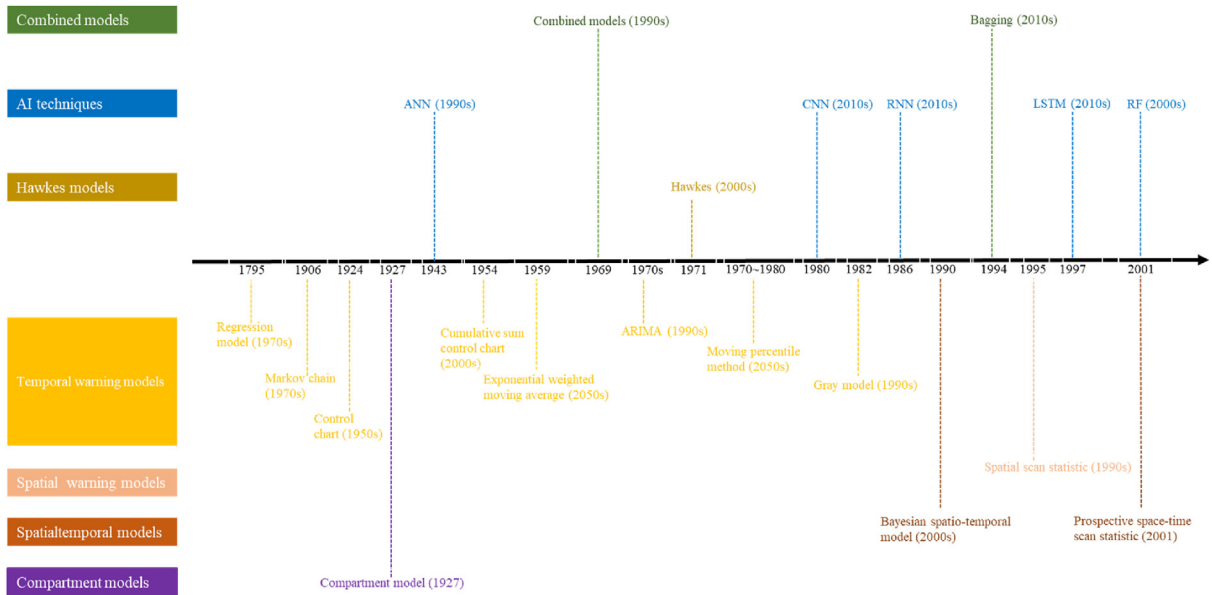


Fig. 2. Timeline of various models being proposed and introduced into the field of public health.

Table 1
Common early warning models for infectious diseases.

Category	Model	Definition or principle	Features or applicability
Models Mainly Based on Traditional Surveillance Data			
A. Temporal warning models			
	Control chart (Siettos & Russo, 2013)	Using historical data and considering the sensitivity, specificity, and positive predictive value of indicators, by establishing an upper or lower limit of indicator early warning as a threshold for identifying abnormal signals, an early warning can be issued when the actual monitored data "out of control".	It is suitable for all kinds of infectious diseases, and has better effect on infectious diseases with seasonal epidemic or periodic epidemic law.
	Cumulative sum control chart (Rakitzis et al., 2018)	Continuously accumulate the difference between the observed value and the expected value, gradually accumulate the fluctuations in the data, and quickly and sensitively detect the small abnormalities in the data.	Real-time surveillance of outbreaks with small changes in the number of cases and surveillance of relatively uncommon infectious diseases are effective. Do not rely on long-term historical baseline data.
	Time series analysis-ARIMA (Alegado & Tumibay, 2019)	The time sequence formed when the forecast target changes with time is taken as a random queue. There are mainly moving average method, exponential smoothing method and Box-Jenkins model.	It is mainly used to analyze and solve the stationarity, randomness and seasonality of the time series. It can also predict the development trend of the research object based on the analysis of the time series. Data preparation and operation are relatively simple and easy to implement, and the results are relatively accurate, and are usually used to predict short-term fluctuations of infectious diseases. Disadvantage: not suitable for rare and severe diseases simulation.
	Time series analysis-Moving percentile method (R. P. Wang et al., 2017)	Non-parametric test method. The median of historical monitoring data represents the average level of monitoring indicators, and the percentile of a certain confidence level is used as the control limit.	Early warning model based on time dimension.
	Time series analysis-Exponential weighted moving average (Alimohamadi et al., 2020; Burkom et al., 2021)	Sequence data processing method.	It is suitable for monitoring sequences with a relatively stable baseline, and is mostly used for short-term historical data. It can introduce weighting ideas based on the distance between historical data and current time, consider the correlation between past data and current data, and is highly sensitive to small changes, and can quickly identify outbreaks.
	Regression model (T. Q. Zhou, 2022)	A method of predicting how a variable change with other variables, in the absence of an outbreak, uses the distribution of residuals to determine the threshold. It is mainly divided into linear regression and nonlinear regression. The common ones are linear regression model, quadratic curve model, cubic curve model, and exponential curve model.	It can be used for outbreak detection based on laboratory results and cases reported by the surveillance system, and for symptom monitoring.
	Markov chain (Z. B. Xu et al., 2022)	Regarding time series as a random process, it is a way to predict the future through different initial states of things and apply probability theory to study the changing trend of random events.	It has simplicity and low data requirements, but has limitation in parameter estimation and assumption. It is suitable for simple models with stable data.
	Gray model (Duan & Nie, 2022)	After turning irregular raw data into more regularly generated data, model equations are established to predict future development trends.	It has no strict requirements on the sample size and probability distribution, the model is simple, and the prediction effect is good. Data and disease data that change in the form of an exponential function are more applicable. It is suitable for short-term prediction of diseases with stable epidemic factors.
B. Spatial warning models			
	Spatial scan statistic (Kulldorff, 1997)	It divides geographic space into many small areas as scanning windows, and uses Monte Carlo hypothesis test to detect the areas where the monitoring indicators are statistically different from the conventional level.	It can identify spatial clusters of infectious diseases and determine high-risk areas, but requires spatially stable data. It assumes that disease spread is uniform and does not account for factors such as population density and mobility that may influence transmission.
C. Spatio-temporal warning models			

(continued on next page)

Table 1 (continued)

Category	Model	Definition or principle	Features or applicability
	Prospective space-time permutation scan statistic (Kulldorff et al., 2005)	This method incorporates the temporal aspect into the spatial scan statistic. By identifying different regions at different time points, it effectively detects spatio-temporal clustering of diseases and enables early warnings.	It can identify spatial clustering of diseases and predict outbreaks. However, this model requires spatial clustering of infectious diseases and cannot issue warnings solely based on temporal clustering.
	Bayesian spatio-temporal models (Rotejanprasert et al., 2020)	It analyzes time effects, spatial effects, and spatio-temporal interaction effects under the Bayesian framework.	It is characterized by more accurate prediction for small sample information, allowing the existence of spatio-temporal interactions and over-discretization, but determining the a priori information is more complicated.
D. Infectious disease dynamics models			
	Compartment models (Kermack & McKendrick, 1991)	The basic idea is to divide the population into several classes (compartments), which represent populations in different disease states, and then use mathematical means to establish the kinetic equations of these variables, and then study the transmission dynamics of the disease process.	The parameters considered in model are difficult to be estimated comprehensively and the parameters may change dynamically at different stages of the epidemic, the prediction effect is often poor, but it has important application value for early warning and evaluation of prevention and control effects.
E. Hawkes models			
	Hawkes models (Lamprinakou et al., 2023; Schoenberg, 2023a)	The Hawkes process is a point process with self-exciting properties, used for modeling randomly occurring events. It assumes that the occurrence of an event elevates the likelihood of subsequent event.	It is particularly advantageous for analyzing small-scale infectious diseases and modeling the initial stage. It's simple yet flexible, allowing spatio-temporal covariates, and can be utilized alongside compartment models and neural networks.
Models Mainly Based on Multi-source Surveillance Data			
F. Artificial intelligence techniques			
	Random forest (Breiman, 2001; Kandula & Shaman, 2019)	A combinatorial prediction model based on decision trees.	Random forest is faster in operation, but its trained model efficacy is lower when the number of samples of a certain class in the dataset is significantly smaller.
	Artificial neural networks (Laureano-Rosario et al., 2018; J. F. Xu & Zhou, 2011)	It classifies and regresses complex information by simulating the process of processing information in the neural network of the human brain.	Good nonlinear fitting ability, but insufficient explanatory power.
	Conventional Neural Network (Dhillon & Verma, 2020; Zheng & Hu, 2021)	CNN is a class of feed-forward neural networks containing convolutional computation and deep structure, and the basic computing units include convolutional layer, activation function layer, pooling layer, and fully-connected layer.	Automatic extraction of target features, large memory consumption, high energy consumption.
	Recurrent neural network (Lu M, 2020)	The outputs of neurons in an RNN can be re-input to the neurons so that the data maintains a dependency.	Small number of parameters required, low transferability between gradients.
	Long short term memory network (P. Huang, 2019)	LSTM adds the controllability of memory function in RNN.	LSTM is suitable for longer interval and delayed time series in the epidemic, and it can adequately approximate complex nonlinear relationships, with long time memory function. However, the output results are difficult to interpret, and the accuracy is low when sample data were small.
G. Combined models			
	Combined models (Kogan et al., 2021; Santillana et al., 2015; Y. L. Shi, Cheng, Huang, & Ren, 2020; Z. F. Yang et al., 2020; Ying et al., 2020; S. Zhou, Shen, & Wang, 2020).	A model that combines multiple early warning modelling approaches to take full advantage of the strengths of each approach, thereby providing more comprehensive, accurate and flexible warnings. It is more common to see combinations of methods such as machine learning, temporal warning models, and Compartment models.	Combined models tend to outperform traditional models in terms of data source usage, model robustness, and model performances, but they are more complex to build and maintain.

difference between observed and expected values, quickly and sensitively detects small anomalies in the data, and does not rely on long-term historical baseline data (Rakitzis, Castagliola, & Maravelakis, 2018). Autoregressive integrated moving average (ARIMA) models systematize smooth time-series sampled data to estimate and infer the future state of a phenomenon by unveiling the underlying pattern between a target variable and time. These models are effective in capturing short-term disease fluctuations, although their applicability is limited when simulating rare and severe diseases (Alegado & Tumibay, 2019). In the moving percentile method, the median of long historical monitoring data represents the average level, and the percentile of a certain confidence level represents the control limit (R. P. Wang, Jiang, Michael, & Zhao, 2017).

Exponential weighted moving average (EWMA) is especially suitable for short-term historical data and sensitive to small changes, considering distance and correlation between past and current data (Alimohamadi et al., 2020; Burkom et al., 2021). A regression model can be used for outbreak detection based on laboratory results and cases reported by the surveillance system, as well as for symptom monitoring (T. Q. Zhou, 2022). The Markov chain predicts the future through different initial states of things and applies the probability theory to study the changing trend of random events. It has low data requirements but has limitations in parameter estimation (Z. B. Xu, Zhang, & Huang, 2022). The gray model is especially useful for small samples and short-term predictions (Duan & Nie, 2022).

In terms of applications, different international early warning systems use different time warning models. In the US, the electronic surveillance system for the early notification of community-based epidemics (ESSENCE) uses an exponentially weighted moving average model to compare and analyze the time-weighted average of current surveillance data with past surveillance data, which is highly sensitive to short-term fluctuations. Additionally, an adaptive multiple regression model was introduced to correct for the effect of holidays (the closer the time, the greater the weight) (Burkom et al., 2021). In Germany, SurvNet uses an automatic outbreak detection algorithm to analyze statutory infectious disease data weekly and generates a signal when detecting the number of incidence cases above a threshold (Straetemans, Altmann, Eckmanns, & Krause, 2008). China Infectious Disease Automated-Alert and Response System (CIDARS) uses the moving percentile method and cumulative and control chart method models to detect and warn different types of infectious diseases. If the incidence level of the number of cases in the current 7-day periods exceeds the level of the n th percentile (P_n) of the baseline data for the same period in the past 5 years (the threshold value varies for each type of disease), the system sends out one warning signal (S. Huang et al., 2022).

2.1.2. Spatial warning models

Spatial warning models analyze the spatial distribution characteristics of infectious diseases and provide insights into the spatial aggregation and outbreak risk by observing surveillance data within a specific period. The theoretical basis is that when an infectious disease outbreak occurs, the number of cases will increase abnormally in a short period and tend to cluster in a specific area, with obvious differences in incidence from other areas. The spatial scan statistic and SaTScan software, proposed and developed by Kulldorff, is one of the most influential methods and tools for identifying spatial aggregation in disease outbreaks (Kulldorff, 1997). This method divides geographic space into small scanning windows and utilizes Monte Carlo hypothesis tests to detect areas where monitoring indicators statistically deviate from the expected level and to identify spatial aggregation. However, it assumes a uniform spread of the disease and does not consider factors such as population density and mobility, which can influence transmission. The data should exhibit a certain degree of spatial stability to use this method effectively. Another approach, the flexible spatial scanning statistic, has been developed to identify “irregularly shaped” aggregation areas and has demonstrated better performance in detecting outbreak areas in practical applications (J. N. Zhou, Feng, Tan, & Li, 2010).

2.1.3. Spatiotemporal warning models

Spatiotemporal warning models simultaneously monitor and provide alerts from both temporal and spatial dimensions, allowing for more comprehensive information and higher timeliness and accuracy. Currently, the prospective space-time permutation scan statistic and the Bayesian space-time model are widely utilized.

The prospective space-time permutation scan statistic, proposed by Kulldorff, Heffernan, Hartman, Assunção, & Mostashari, (2005), incorporates the temporal aspect into the spatial scan statistic. It effectively detects spatiotemporal clustering of infectious diseases and enables early warnings by identifying different regions at different time points. Compared to purely temporal warning models, this approach not only predicts disease outbreaks but also identifies spatial clustering of infectious diseases, which is beneficial for relevant departments to implement timely countermeasures. For instance, the US employed this method for daily monitoring of COVID-19 outbreaks, facilitating informed decision-making and resource allocation (Hohl, Delmelle, Desjardins, & Lan, 2020). However, this model cannot issue warnings solely based on temporal clustering. Bayesian spatiotemporal models can analyze time effects, spatial effects, and spatiotemporal interaction effects under the Bayesian framework. It is characterized by a more accurate prediction for small sample information, allowing the existence of interactions and over-discretization, although determining the a priori information is more complicated (Rotejanaprasert, Ekpirat, Areechokchai, & Maude, 2020).

2.1.4. Dynamics models

Unlike spatiotemporal warning models, the dynamics model utilizes data to estimate system parameters and calculate key indicators, such as the basic regeneration number (R_0), the size of the diseased population, and the inflection point of the epidemic. This is achieved through mathematical analysis based on the transmission mechanism and development law of infectious diseases. These calculations provide a crucial foundation for epidemic research and early warning efforts.

The compartment model, initially proposed by Kennack and McKendrick in 1927, is the predominant dynamic model (Kermack & McKendrick, 1991). It encompasses two classic types of basic compartment models: the SIR model and the SIS model. The fundamental concept behind this model involves dividing the population into distinct classes or compartments that represent different disease states. Then, mathematical methods are employed to establish kinetic equations about the transmission dynamics of disease. However, the predictive efficacy of the model is often limited due to the complexities in the comprehensive estimation of the parameters and the potential dynamic changes of these parameters at various stages of an

epidemic. Nonetheless, it is extremely useful in facilitating early warning systems and evaluating the effectiveness of prevention and control measures (L. H. Huang, Wei, Shen, Zhu, & Chen, 2020). For instance, researchers conducted simulations of potential transmission scenarios employing various dynamic models during the COVID-19 epidemic. These models played a crucial role in forecasting epidemiological trends, evaluating healthcare resource requirements, analyzing vaccination rates, and formulating prevention and control strategies (Kissler, Tedijanto, Goldstein, Grad, & Lipsitch, 2020; R. Y. Li, Pei, et al., 2020).

2.1.5. Hawkes models

The Hawkes model, a type of point process, was introduced by the British statistician Alan G. Hawkes in 1971 (Hawkes, 1971). This model is prominently used in various fields such as finance, criminology, social networks, and more recently, epidemiology. The fundamental principle of the Hawkes model is that it represents events whose occurrence increases the probability of future events (Garetto, Leonardi, & Torrisi, 2021). This self-exciting property makes it particularly suitable for modeling contagious phenomena. For instance, some scholars used the Hawkes model to analyze the Ebola outbreak in Congo and West Africa, determining that the model offers specific advantages for examining small-scale infectious diseases and simulating the initial spreading process of such diseases (Kelly et al., 2019; Park et al., 2022). The COVID-19 pandemic has further accelerated the development and application of the Hawkes model and its variants. For example, Hawkes models have been adapted to incorporate spatial covariates, allowing for the real-time tracking and forecasting of infection hotspots (Chiang et al., 2022). Studies have shown that these models can outperform traditional compartmental models in certain scenarios, providing more granular and dynamic insights (Bertozzi, Franco, Mohler, Short, & Sledge, 2020). For example, a study compared the prediction results of the Hawkes model with those of the SEIR model and discovered that the Hawkes model provided more accurate predictions, with a root mean square error 71 times lower than that of the SEIR model (Park et al., 2022). In all, the basic Hawkes model is well-suited for analyzing and predicting small-scale epidemics or the initial stage of outbreaks. However, due to its simplicity and flexibility, the Hawkes model can be extended into various variants to better adapt to different pandemic stages and complex outbreak scenarios (Dukkipati, Gracious, & Gupta, 2021; Rizoiu, Mishra, Kong, Carman, & Xie, 2017; Unwin et al., 2021).

2.2. Models mainly based on multi-source surveillance data

As disease surveillance technology advances, there is a growing availability of multi-source data, including syndromic data, data from various hospital systems (e.g., radiology information systems, laboratory information systems, and image-archiving and transmission systems) (Durango-Chavez et al., 2022), pharmacy data, wastewater monitoring data (McMahan et al., 2021; Phan et al., 2023), and online public opinion data (e.g., keyword search indexes, social media heat, and news media reports) (Alsudias & Rayson, 2021; Daughton, Chunara, & Paul, 2020; Jing Wei Li et al., 2022; Kim & Ahn, 2021; Li, Xu, et al., 2020; Lim, Tucker, & Kumara, 2017). There is an increasing trend of diversification, heterogeneity, and big data in surveillance data, prompting the need for alternative approaches to early warning models. AI technologies offer the ability to swiftly collect, process, and analyze large-scale multi-source data, enabling timely and accurate warnings (Jordan & Mitchell, 2015). Moreover, these technologies possess the capability for independent learning and performance improvement (Bragazzi et al., 2020).

2.2.1. Artificial intelligence techniques

Machine learning algorithms are the key to the realization of AI technology. Examples include random Forest (RF), artificial neural network (ANN), convolutional neural network (CNN), recurrent neural network (RNN), and long short-term memory network (LSTM). Furthermore, machine learning methods are popular in infectious disease early warning.

The RF is a combinatorial prediction model originally proposed by Breiman in 2001 (Breiman, 2001). It is based on decision trees, with each decision tree constructed from a set of training samples obtained through a self-service sampling method. The data are recursively partitioned in the regression space by the decision tree until the subspace exhibits minimal variation. The final classification result is determined by aggregating the votes from all decision trees. While RF offers faster operation, its efficacy in modeling is diminished when the samples for a particular class are small. To address this limitation, some researchers integrated symptom monitoring and online public data into the model to improve the efficacy negatively impacted by scarce samples. For example, Kandula et al. demonstrated a 44% reduction in warning errors by incorporating the Google weekly influenza prediction trend as an additional data source (Kandula & Shaman, 2019).

The ANN was developed from the neuron model proposed in 1943, which classifies and regresses extensive complex information by simulating the information processing within the human brain's neural network. Neurons serve as the fundamental units for information processing in artificial neural networks and consist of three components: the input layer, hidden layer(s), and output layer. A great number of neurons form an artificial neural network according to a certain connection. These models exhibit excellent nonlinear fitting capabilities (Laureano-Rosario et al., 2018), making them robust against outliers and multicollinearity. However, the ANN cannot explicitly elucidate its inference and the underlying basis of its conclusions due to the intricate nature of the hidden layer (J. F. Xu & Zhou, 2011).

The CNN belongs to a class of feed-forward neural networks incorporating convolutional computation and deep structure. The fundamental computing units in the CNN include the convolutional layer, activation function layer, pooling layer, and fully-connected layer, in which the convolutional and pooling layers are interconnected to form the core component of the

network. The application of CNN gained popularity in 2006. The CNN can automatically extract target features, identify patterns in the sample set, and exhibit high classification accuracy and efficiency (Dhillon & Verma, 2020). However, memory consumption, energy usage, and production costs of the model escalate as the number of network layers increases. Zheng et al. found that a hybrid network of ANN and CNN was more accurate (95.1%) than ANN (89.1%) and CNN (90.1%) alone in the tuberculosis early warning model (Zheng & Hu, 2021).

The RNN is well-suited for processing time series data. The neuronal outputs in the RNN can be fed back as inputs to maintain data dependency and share parameters within the network structure, reducing the required parameters. Lu Min et al. established an RNN-based influenza prediction model for Beijing relying on public opinion data sources, showing good prediction results (Lu M, 2020). However, standard RNNs suffer from gradient vanishing or explosion and have limited gradient transferability during training, resulting in their inability to effectively utilize information from long sequences. To address these issues, Huang Peng et al. added the controllability of memory function in RNN to establish a prediction model of class B infectious diseases in Sichuan Province with LSTM, which was successfully implemented within the Sichuan Center for Disease Control and Prevention (P. Huang, 2019).

2.2.2. Combined model

Early warning of infectious diseases is challenging because the transmission dynamics of disease outbreaks are intricate, and models cannot fully account for the factors involved in disease development. Although various spatial and/or temporal models, dynamics models (mechanism-driven), and AI models (data-driven) have been proposed, some limitations remain in the existing studies. For example, many parameters will be introduced when more practical factors are considered in the mechanism-driven model. Additionally, determining proper values is difficult. Furthermore, data-driven methods are limited in elucidating the spread and persistence principles of diseases, and the use of low-quality data at the early stage of a disease outbreak can introduce uncertainties. Fortunately, the characteristics of different models are complementary to a certain extent because an increasing number of combined models are developed and provide better performance. Since the 21st century, the combination of machine learning and time series analysis has been the most popular combined model. Furthermore, the COVID-19 pandemic has also promoted the integration of machine learning with compartment models, Hawkes models, and the fusion of compartment models with time series analysis.

Recently, combining AI models (data-driven) with dynamics models (mechanism-driven) has become a research hotspot during the COVID-19 pandemic. Data-driven models can assist in reducing the reliance on assumptions and parameters in epidemic mechanisms by incorporating infectious disease features in data. Meanwhile, mechanism-driven models can enhance the interpretability of data-driven models and mitigate uncertainties caused by low-quality data. Consequently, integrating data-driven (especially deep neural networks) and mechanism-driven methods is promising for advancing the study of infectious diseases. For instance, Chen F et al. and Xiao YN et al. constructed a combined model including transmission-dynamics-informed neural networks (TDINNs) and epidemiological priors informed deep neural networks (Epi-DNNs) to determine the intensity of interventions during the COVID-19 pandemic, integrating a deep neural network with a dynamical model (He, Tang, & Xiao, 2023; Ning, Jia, Wei, Li, & Chen, 2023). Similarly, foreign scholar Marian Petrica employed a neural network to dynamically estimate parameters for SIRD models over seven consecutive days. Based on this model, future events can be predicted for the upcoming day based on the model (Petrica & Popescu, 2023). Furthermore, in addition to deep neural networks, the level-based learning swarm optimizer has been fused with SEIR models to forecast emerging infectious diseases, considering factors such as population migration (X. L. Shi, Wei, & Chen, 2023).

Meanwhile, the Hawkes process model was developed in the realm of infectious diseases due to the COVID-19, and its combined model with machine learning was used for the first time in this domain. For instance, Dukkipati et al. combined LSTM with Hawkes to estimate the dynamic reproduction number and forecast the daily cases of COVID-19, of India, laying the groundwork for the containment policy in the country (Dukkipati et al., 2021). Moreover, the predictive efficacy of compartment models may be constrained by their dependence on numerous assumptions and the lack of real-time or accurate data availability during outbreaks. Research has indicated that integrating time series analysis with a compartment model can address this limitation. Following the development of the compartment model, methodologies such as the ARIMA model can be employed to compute and forecast variances between the compartment model estimates and the actual data. These variances can be then utilized to refine model predictions, enhancing overall model reliability (Ala'raj, Majdalawieh, & Nizamuddin, 2021; W. P. Zhao et al., 2022).

Previously, numerous studies focused on combining AI (data-driven) with other models, with one of the commonly used models being the temporal warning model. However, the prediction performance of temporal warning models (e.g., ARIMA) is often insufficient for nonlinear parts and direct prediction of the original series. To address these issues, temporal warning models were combined with data-driven models. For example, Oliveira proposed the ARIMA-SVR model, combining ARIMA with support vector regression in 2016. This hybrid model significantly improved the prediction accuracy (de Oliveira & Ludermir, 2016). In 2021, Zhai et al. used the combination model of ARIMA-BPNN and ARIMA-ERNN to predict brucellosis in Shanxi Province, China (Zhai et al., 2021), demonstrating that combination models outperformed the standalone ARIMA model. Nevertheless, all the above-mentioned models predicted the original sequence. The accuracy of directly using the combined model for prediction is still insufficient when handling complex characteristics in the original sequences. To solve this problem, a decomposition- and reorganization-based strategy for constructing a combined model was proposed. Singular spectrum analysis (SSA) can decompose the complex original sequence into simpler and regular sub-sequences (Hassani,

2022). In 2023, Zhao et al. built an indirect combinatorial prediction model of SSA-SARIMA-LSTM for influenza prediction. This model improved the prediction accuracy by modeling and superimposing the sub-sequences (Z. Zhao et al., 2023).

In addition to time warning models, there are also combinations of AI with other models. For example, Santillana et al. proposed an influenza early warning model aggregating a network model with real-time multivariate linear regression to optimize the combination of multiple sources of data, such as Google search, social media data, hospital visit records, and influenza-like case surveillance, which is better than a single source of data for early warning (Santillana et al., 2015). Kogan et al. used a Bayesian model in the US to analyze a combination of information regarding COVID-19 on social media, internet search information, and fever surveillance data and captured an exponential rise in the number of confirmed cases 2–3 weeks in advance (Kogan et al., 2021). Madhav Marathe introduced a novel epidemic forecasting framework called deep learning-based epidemic forecasting with synthetic information (DEFISI) for short-term, high-resolution predictions of influenza. DEFISI combines multi-agent systems and deep neural networks, harnessing the benefits of both causal methods and AI (L. J. Wang, Chen, & Marathe, 2019). Moreover, stacking-based deterministic ensemble models were popular for infectious disease prediction (Mahajan et al., 2022; Sherratt et al., 2023).

Furthermore, studies combined methods other than AI. For instance, the Bayesian approach with the combination of SIR and spatial conditional autoregressive models was used to understand the transmission of the H1N1 pandemic in 2009 over time and space (X. D. Huang et al., 2016).

3. Discussion

Each model type has its own strengths, weaknesses, and applicability to different scientific questions. In this paper, infectious disease early warning models are divided into four main types: spatial and/or temporal warning model, dynamic model, AI techniques, and combined models. Spatiotemporal models can effectively detect disease outbreaks early in both time and space, facilitating the allocation of public health resources (Ting et al., 2021). However, this model requires a significant amount of geographic and migration data for a population. Dynamic models provide a mechanistic perspective on disease transmission. They can predict epidemiological trends, analyze influencing factors, and evaluate the effects of interventions, such as vaccination. However, determining model parameters and considering time-varying parameters can be challenging (Davarci, Yang, Viguerie, Yankeelov, & Lorenzo, 2023). Fortunately, AI techniques, such as neural networks, can be used to estimate the parameters of dynamic models and allow time-varying parameters to be considered, greatly improving the model prediction ability (He et al., 2023). However, AI algorithms are less interpretable, and their performance relies heavily on data quality. Among other combined models, the integration of time series and deep learning is widely used and improves predictions with fewer parameters (Pathan, Biswas, & Khandaker, 2020). Nevertheless, most models tend to perform better in the short term than in the long term. In conclusion, selecting a suitable model based on the specific situation and research objective is essential.

Several factors influence model selection (Deng et al., 2022). First, the stage of transmission plays an important role. In the early stages of an epidemic, understanding of the disease is limited, leading to the construction of only a simple SIR model. As disease progresses and essential insights are unveiled, building more complex models that can perform new or different functions becomes possible. Second, data availability is crucial. Insufficient real-world data limits researchers to basic data-driven models for predicting epidemic trajectories. However, models can be enhanced regarding types, accuracy, and complexity as data volume and diversity increase. The selected model should seamlessly handle various data types, including disease incidence, climate data, and socioenvironmental variables (Haque et al., 2024). Moreover, the choice of a statistical model should consider the specific characteristics of the infectious disease and influencing factors. For diseases with clear temporal patterns, ARIMA or SARIMA models may be suitable. In cases of complex interactions, machine learning techniques can provide valuable insights. They have been valuable for understanding non-linear relationships and complex interactions between environmental factors and diseases like COVID-19 and influenza. In addition, balancing model complexity with interpretability and accounting for uncertainty in model are vital (Mariotti, Alonso Moral, & Gatt, 2023). Bayesian methods and ensemble modelling can provide a more comprehensive understanding of uncertainty associated with various model choices (Abdar et al., 2021). Lastly, the purpose of the model must be clear. For instance, evaluating the impact of prevention and control measures requires consideration of the natural history of the disease, characteristics and settings of the intervention, and other factors, for which compartment models are applicable (Kong et al., 2022).

The outbreak of COVID-19 posed a serious challenge to the infectious disease surveillance system and early warning methods globally and locally in China (Meckawy, Stuckler, Mehta, Al-Ahdal, & Doebbeling, 2022). On the one hand, public health authorities can enhance the development of robust early warning systems by comprehensively understanding the strengths and limitations of different models and applying them to a group of diseases with similar transmission routes and characteristics (Haque et al., 2024; Semenza, 2015). The effectiveness of the early warning system can be improved through comprehensive application of various models and continuous evaluation and improvement (Chinese Academy of Medical Sciences and Peking Union Medical College Peking University Chinese Center for Disease Control and Prevention, 2024). On the other hand, as far as early warning models are concerned, their performance relies on two main aspects: the model and the data source. Research and practical applications of early warning models for infectious diseases using traditional surveillance data have currently reached an advanced stage, whereas AI models and combined models leveraging multi-source data are still in the developmental phase. The advent of big data has led to a diversification of data sources for infectious disease early warning, encompassing syndromic data from hospitals (such as fever, cough, and diarrhea), data from

different systems in hospitals, such as Radiography Information System (RIS) and Laboratory Information System (LIS), and data from public sentiment networks, such as internet search indices and social media trends (Gao et al., 2022; Hughes, Edeghere, O'Brien, Vivancos, & Elliot, 2020). Consequently, the integration of big data and AI techniques might play a pivotal role in the enhancement of early warning models (Jiao, Ji, Yan, & Qi, 2023). Further research and exploration are essential to take full advantage of multi-source data, optimize algorithm performance, and enhance the interpretability of these models. Hence, the combination of AI and other early warning models, especially dynamic models, will become a future development trend (He et al., 2023; Ning et al., 2023; Petrica & Popescu, 2023).

The COVID-19 pandemic has underscored the critical importance of surveilling unknown pathogens (“Disease X”) on a global scale (Cunningham & Hopkins, 2023). Given its enigmatic nature, Disease X presents unique challenges to existing models. How can infectious disease early warning models specifically contribute to Disease X surveillance and response? Firstly, in terms of pathogen identification and mutant strain prediction, researchers can leverage multimodal and multi-omics data of emerging infectious viruses. After constructing a virus mutation spectrum, they can use virus variant prediction models (e.g., LSTM, RF, explainable AI, combined models), mutant antigen and receptor affinity analysis (e.g., machine learning, molecular dynamics simulation), virus dynamic evolution and transmission models (e.g., compartment models, combined models), and variant monitoring and early warning (e.g., big data and artificial intelligence) to predict virus mutation and transmission, providing forward-looking technical support for epidemic prevention and control (Ren et al., 2023). Secondly, regarding disease transmission characteristics and trend prediction, researchers can enhance existing early warning models by incorporating subdivided compartments, dynamic parameters, agent-based models, combined models and interdisciplinary technologies. This approach enables a more in-depth identification of transmission characteristics and improved trend prediction. For instance, development of foundational or variant Hawkes model to promptly detect and forecast outbreaks of unknown pathogens based on its simple yet adaptable properties (Kelly et al., 2019; Park et al., 2022; Schoenberg, 2023b; Unwin et al., 2021); application of innovative combined models to better utilize multi-source monitoring data and combine big data technology to achieve multi-trigger intelligent monitoring (Jin, Dong, Yu, & Luo, 2022; Papageorgiou & Tsaklidis, 2023; Y. Y. Zhang, Tang, & Yu, 2023); development of novel dynamic GIS-based spatiotemporal models to link infectious diseases with internet-based data and social-environmental data (Nazia, Law, & Butt, 2023); and integration of internet-based models with social-environmental data to produce infectious disease surveillance systems capable of better identifying vulnerable/susceptible communities over space and time (McClymont et al., 2024). Finally, concerning the intervention effects of prevention and control measures, researchers can evaluate the impacts of various interventions, such as human mobility restrictions, non-pharmacological measures, healthcare resource availability, and vaccination strategies through diverse modeling methodologies like compartment models, agent-based models, system dynamics models, network models, and combined models (Xiang et al., 2021; L. H. Zhou et al., 2022).

4. Conclusion

The recent COVID-19 pandemic has significantly propelled the advancement of both existing and emerging early warning models for infectious diseases, while also presenting challenges and guiding future model development and application. Unlike previous review studies, this review encompasses newly emerging or recently developed approaches, such as the combined model and Hawkes model, providing a thorough evaluation of their current application status and development prospects for the first time. Additionally, this review not only summarizes early warning models for specific infectious disease types but also offers a comprehensive overview of those in the broader field of infectious diseases. This includes an analysis of the principles, characteristics, and applicability of these models, serving as a foundational resource for subsequent researchers to comprehend various models and select the most suitable ones. Each model possesses its unique set of strengths and weaknesses. This study reveals that scholars have been exploring various aspects, such as improving existing models, developing emerging methods, and implementing innovative model combinations, with significant results. This underscores the importance of fully utilizing multi-source surveillance data for various types of infectious diseases and employing appropriate models for prediction within the existing early warning system. Moreover, regular evaluation and updating of the system and models are essential to tackle unknown pathogens that are constantly evolving.

CRedit authorship contribution statement

Wei-Hua Hu: Writing – original draft, Methodology, Data curation. **Hui-Min Sun:** Data curation. **Yong-Yue Wei:** Writing – review & editing. **Yuan-Tao Hao:** Writing – review & editing, Supervision, Conceptualization.

Funding

Partly supported by the National Natural Science Foundation of China (81973150). Professor Yuan-Tao Hao gratefully acknowledges the support of the K.C.Wong Education Foundation.

References

- Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., ... Nahavandi, S. (2021). A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*, 76, 243–297. <https://doi.org/10.1016/j.inffus.2021.05.008>
- Ala'raj, M., Majdalawieh, M., & Nizamuddin, N. (2021). Modeling and forecasting of COVID-19 using a hybrid dynamic model based on SEIRD with ARIMA corrections. *Infectious Disease Modelling*, 6. <https://doi.org/10.1016/j.idm.2020.11.007>
- Alegado, R. T., & Tumibay, G. M. (2019). Forecasting measles immunization coverage using ARIMA model. *Journal of Computer and Communications*, 7(10), 157–168.
- Alimohamadi, Y., Zahraei, S. M., Karami, M., Yaseri, M., Lotfizad, M., & Holakouie-Naieni, K. (2020). *Timely detection of Pertussis outbreaks in Iran: The comparison performance of Wavelet-based outbreak detector, Exponential weighted moving average, and Poisson regression-based methods.*
- Alsudias, L., & Rayson, P. (2021). Social media monitoring of the COVID-19 pandemic and influenza epidemic with adaptation for informal language in Arabic twitter data: Qualitative study. *JMIR medical informatics*, 9(9), Article e27670. <https://doi.org/10.2196/27670>
- Bertozi, A. L., Franco, E., Mohler, G., Short, M. B., & Sledge, D. (2020). The challenges of modeling and forecasting the spread of COVID-19. *Proceedings of the National Academy of Sciences*, 117(29), 16732–16738.
- Bragazzi, N. L., Dai, H., Damiani, G., Behzadifar, M., Martini, M., & Wu, J. (2020). How big data and artificial intelligence can help better manage the COVID-19 pandemic. *International Journal of Environmental Research and Public Health*, 17(9), 3176.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.
- Burkom, H., Loschen, W., Wojcik, R., Holtry, R., Punjabi, M., Siwek, M., et al. (2021). Electronic surveillance system for the early notification of community-based epidemics (ESSENCE): Overview, components, and public health applications. *JMIR Public Health and Surveillance*, 7(6), Article e26303. <https://doi.org/10.2196/26303>
- Chen, Y. L., Liu, F. W., Yu, Q., & Li, T. Z. (2021). Review of fractional epidemic models. *Applied Mathematical Modelling*, 97, 281–307. <https://doi.org/10.1016/j.apm.2021.03.044>
- Chiang, W. H., Liu, X., & Mohler, G. (2022). Hawkes process modeling of COVID-19 with mobility leading indicators and spatial covariates. *International Journal of Forecasting*, 38(2), 505–520. <https://doi.org/10.1016/j.ijforecast.2021.07.001>
- Chinese Academy of Medical Sciences and Peking Union Medical College, Peking University, Chinese Center for Disease Control and Prevention. (2024). Sichuan university, Chinese Academy of Sciences, Beijing ditan hospital, & Chinese medical association. *Expert consensus on the key technologies for a multi-point trigger intelligent surveillance and early warning system for infectious diseases.* *Chinese Medical Journal*, 104(32), 2995–3009 (In Chinese).
- Cunningham, N., & Hopkins, S. (2023). Lessons identified for a future pandemic. *Journal of Antimicrobial Chemotherapy*, 78(Suppl 2), ii43–ii49. <https://doi.org/10.1093/jac/dkad310>
- Daughton, A. R., Chunara, R., & Paul, M. J. (2020). Comparison of social media, syndromic surveillance, and microbiologic acute respiratory infection data: Observational study. *JMIR Public Health and Surveillance*, 6(2), Article e14986. <https://doi.org/10.2196/14986>
- Davarci, O. O., Yang, E. Y., Viguier, A., Yankeelov, T. E., & Lorenzo, G. (2023). Dynamic parameterization of a modified SEIRD model to analyze and forecast the dynamics of COVID-19 outbreaks in the United States. *Engineering with computers*, 1–25. <https://doi.org/10.1007/s00366-023-01816-9>
- de Oliveira, J. F. L., & Ludermit, T. B. (2016). A hybrid evolutionary decomposition system for time series forecasting. *Neurocomputing*, 180, 27–34. <https://doi.org/10.1016/j.neucom.2015.07.113>
- Declich, S., & Carter, A. O. (1994). Public health surveillance: Historical origins, methods and evaluation. *Bulletin of the World Health Organization*, 72(2), 285.
- Deng, B., Niu, Y., Xu, J. W., Rui, J., Lin, S. N., Zhao, Z. Y., ... Chen, T. M. (2022). Mathematical models supporting control of COVID-19. *China CDC Wkly*, 4(40), 895.
- Dhillon, A., & Verma, G. K. (2020). Convolutional neural network: A review of models, methodologies and applications to object detection. *Progress in Artificial Intelligence*, 9(2), 85–112.
- Duan, H. M., & Nie, W. G. (2022). A novel grey model based on susceptible infected recovered model: A case study of COVID-19. *Physica A: Statistical Mechanics and Its Applications*, 602, Article 127622.
- Dukkipati, A., Gracious, T., & Gupta, S. (2021). *CoviHawkes: Temporal point process and deep learning based covid-19 forecasting for India.*
- Durango-Chavez, H. V., Toro-Huamanchumo, C. J., Silva-Caso, W., Martins-Luna, J., Aguilar-Luis, M. A., Del Valle-Mendoza, J., et al. (2022). Oropouche virus infection in patients with acute febrile syndrome: Is a predictive model based solely on signs and symptoms useful? *PLoS One*, 17(7), Article e0270294. <https://doi.org/10.1371/journal.pone.0270294>
- Gao, C. C., Zhang, R., Chen, X. C., Yao, T. H., Song, Q. Y., Ye, W., ... Wu, Y. Z. (2022). Integrating Internet multisource big data to predict the occurrence and development of COVID-19 cryptic transmission. *NPJ digital medicine*, 5(1), 161. <https://doi.org/10.1038/s41746-022-00704-8>
- Garetto, M., Leonardi, E., & Torrisi, G. L. (2021). A time-modulated Hawkes process to model the spread of COVID-19 and the impact of countermeasures. *Annual Reviews in Control*, 51, 551–563.
- Haque, S., Mengersen, K., Barr, I., Wang, L. P., Yang, W. Z., Vardoulakis, S., ... Hu, W. B. (2024). Towards development of functional climate-driven early warning systems for climate-sensitive infectious diseases: Statistical models and recommendations. *Environmental Research*, 249, Article 118568. <https://doi.org/10.1016/j.envres.2024.118568>
- Hassani, H. (2022). Singular spectrum analysis: Methodology and comparison. *Journal of Data Science*, 5(2), 239–257. [https://doi.org/10.6339/JDS.2007.05\(2\).396](https://doi.org/10.6339/JDS.2007.05(2).396)
- Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1), 83–90. <https://doi.org/10.2307/2334319>
- He, M. Q., Tang, S. Y., & Xiao, Y. N. (2023). Combining the dynamic model and deep neural networks to identify the intensity of interventions during COVID-19 pandemic. *PLoS Computational Biology*, 19(10), Article e1011535. <https://doi.org/10.1371/journal.pcbi.1011535>
- Hohl, A., Delmelle, E. M., Desjardins, M. R., & Lan, Y. (2020). Daily surveillance of COVID-19 using the prospective space-time scan statistic in the United States. *Spatial and Spatio-Temporal Epidemiology*, 34, Article 100354.
- Huang, P. (2019). *Research and implementation of machine learning-based prediction model for class B infectious diseases.* University of Electronic Science and Technology of China (In Chinese).
- Huang, X. D., Clements, A. C. A., Williams, G., Mengersen, K., Tong, S. L., & Hu, W. B. (2016). Bayesian estimation of the dynamics of pandemic (H1N1) 2009 influenza transmission in queensland: A space-time SIR-based model. *Environmental Research*, 146, 308–314. <https://doi.org/10.1016/j.envres.2016.01.013>
- Huang, S., Liu, C. X., Deng, Y., Zhang, C. H., Fan, S. M., Zheng, J. D., et al. (2022). Progress in the practice of surveillance and early warning of infectious diseases in major countries and regions. *Chinese Journal of Epidemiology*, 43(4), 591–597 (In Chinese).
- Huang, L. H., Wei, Y. Y., Shen, S. P., Zhu, C. W., & Chen, F. (2020). Evaluation of predictive models for novel coronavirus pneumonia. *Chinese Journal of Health Statistics*, 37(3), 322–326 (In Chinese).
- Hughes, H. E., Edeghere, O., O'Brien, S. J., Vivanco, R., & Elliot, A. J. (2020). Emergency department syndromic surveillance systems: A systematic review. *BMC Public Health*, 20(1), 1891. <https://doi.org/10.1186/s12889-020-09949-y>
- Jacobsen, K. H., Aguirre, A. A., Bailey, C. L., Baranova, A. V., Crooks, A. T., Croitoru, A., ... Agouris, P. (2016). Lessons from the Ebola outbreak: Action items for emerging infectious disease preparedness and response. *EcoHealth*, 13(1), 200–212. <https://doi.org/10.1007/s10393-016-1100-5>
- Jiao, Z. T., Ji, H. R., Yan, J., & Qi, X. P. (2023). Application of big data and artificial intelligence in epidemic surveillance and containment. *Intelligent Medicine*, 3(1), 36–43. <https://doi.org/10.1016/j.jimed.2022.10.003>
- Jin, W. Q., Dong, S. Q., Yu, C. Q., & Luo, Q. Q. (2022). A data-driven hybrid ensemble AI model for COVID-19 infection forecast using multiple neural networks and reinforced learning. *Computers in Biology and Medicine*, 146, Article 105560. <https://doi.org/10.1016/j.combiomed.2022.105560>
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255–260.
- Kandula, S., & Shaman, J. (2019). Reappraising the utility of Google flu trends. *PLoS Computational Biology*, 15(8), Article e1007258.

- Kaplan, A., Park, J., Kresin, C., & Schoenberg, F. (2022). Nonparametric estimation of recursive point processes with application to mumps in Pennsylvania. *Biomedical Journal*, 64(1), 20–32. <https://doi.org/10.1002/bimj.202000245>
- Kelly, J. D., Park, J., Harrigan, R. J., Hoff, N. A., Lee, S. D., Wannier, R., ... Schoenberg, F. P. (2019). Real-time predictions of the 2018–2019 Ebola virus disease outbreak in the Democratic Republic of the Congo using Hawkes point process models. *Epidemics*, 28, Article 100354. <https://doi.org/10.1016/j.epidem.2019.100354>
- Kermack, W. O., & McKendrick, A. G. (1991). Contributions to the mathematical theory of epidemics–I. 1927. *Bulletin of Mathematical Biology*, 53(1–2), 33–55.
- Kim, J., & Ahn, I. (2021). Infectious disease outbreak prediction using media articles with machine learning models. *Scientific Reports*, 11(1), 4413. <https://doi.org/10.1038/s41598-021-83926-2>
- Kissler, S. M., Tedijanto, C., Goldstein, E., Grad, Y. H., & Lipsitch, M. (2020). Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science*, 368(6493), 860–868.
- Kobres, P. Y., Chretien, J. P., Johansson, M. A., Morgan, J. J., Whung, P. Y., Mukundan, H., ... Pollett, S. (2019). A systematic review and evaluation of Zika virus forecasting and prediction research during a public health emergency of international concern. *PLoS Neglected Tropical Diseases*, 13(10), Article e0007451. <https://doi.org/10.1371/journal.pntd.0007451>
- Kogan, N. E., Clemente, L., Liautaud, P., Kaashoek, J., Link, N. B., Nguyen, A. T., ... Havas, C. (2021). An early warning approach to monitor COVID-19 activity with multiple digital traces in near real time. *Science Advances*, 7(10), Article eabd6989.
- Kong, L. C., Duan, M. W., Shi, J., Hong, J., Chang, Z. R., & Zhang, Z. J. (2022). Compartmental structures used in modeling COVID-19: A scoping review. *Infectious Diseases of Poverty*, 11(1), 72. <https://doi.org/10.1186/s40249-022-01001-y>
- Kulldorff, M. (1997). A spatial scan statistic. *Communications in Statistics - Theory and Methods*, 26(6), 1481–1496.
- Kulldorff, M., Heffernan, R., Hartman, J., Assunção, R., & Mostashari, F. (2005). A space–time permutation scan statistic for disease outbreak detection. *PLoS Medicine*, 2(3), Article e59.
- Lai, S. J., Feng, L. Z., Leng, Z. W., Lyu, X., Li, R. Y., Yin, L., ... Yang, W. Z. (2021). Summary and prospect of early warning models and systems for infectious disease outbreaks. *Chinese Journal of Epidemiology*, 42(8), 1330–1335 (In Chinese).
- Lamprinakou, S., Gandy, A., & McCoy, E. (2023). Using a latent Hawkes process for epidemiological modelling. *PLoS One*, 18(3), Article e0281370. <https://doi.org/10.1371/journal.pone.0281370>
- Laureano-Rosario, A. E., Duncan, A. P., Mendez-Lazaro, P. A., Garcia-Rejon, J. E., Gomez-Carro, S., Farfan-Ale, J., ... Muller-Karger, F. E. (2018). Application of artificial neural networks for dengue fever outbreak predictions in the northwest coast of Yucatan, Mexico and San Juan, Puerto Rico. *Tropical Medicine and Infectious Disease*, 3(1), 5.
- Li, J. W., Huang, W., Sia, C. L., Chen, Z., Wu, T. L., & Wang, Q. N. (2022). Enhancing COVID-19 epidemic forecasting accuracy by combining real-time and historical data from multiple internet-based sources: Analysis of social media data, online news articles, and search queries. *JMIR Public Health and Surveillance*, 8(6), Article e35266. <https://doi.org/10.2196/35266>
- Li, R. Y., Pei, S., Chen, B., Song, Y. M., Zhang, T., Yang, W., et al. (2020). Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science*, 368(6490), 489–493.
- Li, J. W., Xu, Q., Cuomo, R., Purushothaman, V., & Mackey, T. (2020). Data mining and content analysis of the Chinese social media platform weibo during the early COVID-19 outbreak: Retrospective observational infoveillance study. *JMIR Public Health and Surveillance*, 6(2), Article e18700. <https://doi.org/10.2196/18700>
- Lim, S., Tucker, C. S., & Kumara, S. (2017). An unsupervised machine learning model for discovering latent infectious diseases using social media data. *Journal of Biomedical Informatics*, 66, 82–94. <https://doi.org/10.1016/j.jbi.2016.12.007>
- Lu, M. D. J., & Ge, X. S. N. (2020). Neural network-based analysis of influenza big data. *China Journal of Medicine Libr InfScience*, 29(3), 26–31.
- Mahajan, A., Sharma, N., Aparicio-Oregon, S., Alyami, H., Alharbi, A., Anand, D., ... Goyal, N. (2022). A novel stacking-based deterministic ensemble model for infectious disease prediction. *Mathematics*, 10(10), 1714.
- Mariotti, E., Alonso Moral, J. M., & Gatt, A. (2023). Exploring the balance between interpretability and performance with carefully designed constrainable Neural Additive Models. *Information Fusion*, 99, Article 101882. <https://doi.org/10.1016/j.inffus.2023.101882>
- McClymont, H., Lambert, S. B., Barr, I., Vardoulakis, S., Bambrick, H., & Hu, W. (2024). Internet-based surveillance systems and infectious diseases prediction: An updated review of the last 10 Years and lessons from the COVID-19 pandemic. *Journal of Epidemiology and Global Health*. <https://doi.org/10.1007/s44197-024-00272-y>
- McMahan, C. S., Self, S., Rennett, L., Kalbaugh, C., Kriebel, D., Graves, D., ... Freedman, D. L. (2021). COVID-19 wastewater epidemiology: A model to estimate infected populations. *The Lancet Planetary Health*, 5(12), e874–e881. [https://doi.org/10.1016/S2542-5196\(21\)00230-8](https://doi.org/10.1016/S2542-5196(21)00230-8)
- Meckawan, R., Stuckler, D., Mehta, A., Al-Ahdal, T., & Doebbeling, B. N. (2022). Effectiveness of early warning systems in the detection of infectious diseases outbreaks: A systematic review. *BMC Public Health*, 22(1), 2216. <https://doi.org/10.1186/s12889-022-14625-4>
- Murray, J., & Cohen, A. L. (2017). Infectious disease surveillance. *International Encyclopedia of Public Health*, 222–229. <https://doi.org/10.1016/b978-0-12-803678-5.00517-8>
- Nazia, N., Law, J., & Butt, Z. A. (2023). Modelling the spatiotemporal spread of COVID-19 outbreaks and prioritization of the risk areas in Toronto, Canada. *Health & Place*, 80, Article 102988. <https://doi.org/10.1016/j.healthplace.2023.102988>
- Ning, X., Jia, L. L., Wei, Y. Y., Li, X. A., & Chen, F. (2023). Epi-DNNs: Epidemiological priors informed deep neural networks for modeling COVID-19 dynamics. *Computers in Biology and Medicine*, 158, Article 106693. <https://doi.org/10.1016/j.combiomed.2023.106693>
- Papageorgiou, V. E., & Tsaklidis, G. (2023). An improved epidemiological-unscented Kalman filter (hybrid SEIHCRDV-UKF) model for the prediction of COVID-19. Application on real-time data. *Chaos, Solitons & Fractals*, 166, Article 112914. <https://doi.org/10.1016/j.chaos.2022.112914>
- Park, J., Chaffee, A. W., Harrigan, R. J., & Schoenberg, F. P. (2022). A non-parametric Hawkes model of the spread of Ebola in west Africa. *Journal of Applied Statistics*, 49(3), 621–637. <https://doi.org/10.1080/02664763.2020.1825646>
- Pathan, R. K., Biswas, M., & Khandaker, M. U. (2020). Time series prediction of COVID-19 by mutation rate analysis using recurrent neural network-based LSTM model. *Chaos, Solitons & Fractals*, 138, Article 110018. <https://doi.org/10.1016/j.chaos.2020.110018>
- Petrica, M., & Popescu, I. (2023). Inverse problem for parameters identification in a modified SIRD epidemic model using ensemble neural networks. *BioData Mining*, 16(1), 22. <https://doi.org/10.1186/s13040-023-00337-x>
- Phan, T., Brozak, S., Pell, B., Oghuan, J., Gitter, A., Hu, T., ... Wu, F. (2023). Making waves: Integrating wastewater surveillance with dynamic modeling to track and predict viral outbreaks. *Water Research*, 243, Article 120372. <https://doi.org/10.1016/j.watres.2023.120372>
- Rakitzis, A. C., Castagiola, P., & Maravelakis, P. E. (2018). Cumulative sum control charts for monitoring geometrically inflated Poisson processes: An application to infectious disease counts data. *Statistical Methods in Medical Research*, 27(2), 622–641.
- Ren, H. T., Ling, Y. C., Cao, R. F., Wang, Z., Li, Y. X., & Huang, T. (2023). Early warning of emerging infectious diseases based on multimodal data. *Biosafety and Health*. <https://doi.org/10.1016/j.jsheal.2023.05.006>
- Rizoiu, M. A., Mishra, S., Kong, Q., Carman, M., & Xie, L. (2017). *SIR-Hawkes: Linking epidemic models and Hawkes processes to model diffusions in finite populations*.
- Rotejanaprasert, C., Ekapirat, N., Areechokchai, D., & Maude, R. J. (2020). Bayesian spatiotemporal modeling with sliding windows to correct reporting delays for real-time dengue surveillance in Thailand. *International Journal of Health Geographics*, 19, 1–13.
- Santillana, M., Nguyen, A. T., Dredze, M., Paul, M. J., Nsoesie, E. O., & Brownstein, J. S. (2015). Combining search, social media, and traditional data sources to improve influenza surveillance. *PLoS Computational Biology*, 11(10), Article e1004513.
- Schoenberg, F. (2023a). Estimating Covid-19 transmission time using Hawkes point processes. *Annals of Applied Statistics*, 17(4), 3349–3362.
- Schoenberg, F. (2023b). Some statistical problems involved in forecasting and estimating the spread of SARS-CoV-2 using Hawkes point processes and SEIR models. *Environmental and Ecological Statistics*, 30(4), 851–862.

- Semenza, J. C. (2015). Prototype early warning systems for vector-borne diseases in Europe. *International Journal of Environmental Research and Public Health*, 12(6), 6333–6351. <https://doi.org/10.3390/ijerph120606333>
- Shausan, A., Nazarathy, Y., & Dyda, A. (2023). Emerging data inputs for infectious diseases surveillance and decision making. *Frontiers in Digital Health*, 5, Article 1131731. <https://doi.org/10.3389/fgdth.2023.1131731>
- Sherratt, K., Gruson, H., Grah, R., Johnson, H., Niehus, R., Prasse, B., ... Funk, S. (2023). Predictive performance of multi-model ensemble forecasts of COVID-19 across European nations. *Elife*, 12. <https://doi.org/10.7554/eLife.81916>
- Shi, Y. L., Cheng, H. H., Huang, L. Y., & Ren, T. X. (2020). Using a discrete stochastic model to study the epidemic dynamics of COVID-19 in Hubei, China (Vol. 2, pp. 145–154). *Journal of University of Chinese Academy of Sciences* (In Chinese).
- Shi, H. H., Wang, J. Y., Cheng, J. W., Qi, X. P., Ji, H. R., Struchiner, C. J., ... Turgiev, A. S. (2023). Big data technology in infectious diseases modeling, simulation, and prediction after the COVID-19 outbreak. *Intelligent Medicine*, 3(2), 85–96. <https://doi.org/10.1016/j.imed.2023.01.002>
- Shi, X. L., Wei, F. F., & Chen, W. N. (2023). A swarm-optimizer-assisted simulation and prediction model for emerging infectious diseases based on SEIR. *Complex & intelligent systems*, 9(2), 2189–2204. <https://doi.org/10.1007/s40747-022-00908-1>
- Siettos, C. I., & Russo, L. (2013). Mathematical modeling of infectious disease dynamics. *Virulence*, 4(4), 295–306.
- Straetmans, M., Altmann, D., Eckmanns, T., & Krause, G. (2008). Automatic outbreak detection algorithm versus electronic reporting system. *Emerging Infectious Diseases*, 14(10), 1610–1612. <https://doi.org/10.3201/eid1410.071354>
- Ting, C. Y., Zakariah, H., Kamaludin, F., Cheng, D. L. W., Tan, N. Y. Z., & Yee, H. J. (2021). Geospatial analytics for COVID-19 active case detection. *Computers, Materials & Continua*, 67(1), 835–848.
- Unwin, H. J. T., Routledge, I., Flaxman, S., Rizoiu, M.-A., Lai, S., Cohen, J., ... Bhatt, S. (2021). Using Hawkes Processes to model imported and local malaria cases in near-elimination settings. *PLoS Computational Biology*, 17(4), Article e1008830.
- Wang, L. J., Chen, J. Z., & Marathe, M. V. (2019). Defsi: Deep learning based epidemic forecasting with synthetic information. *Paper presented at the AAAI conference on artificial intelligence*.
- Wang, R. P., Jiang, Y. G., Michael, E., & Zhao, G. M. (2017). How to select a proper early warning threshold to detect infectious disease outbreaks based on the China infectious disease automated alert and response system (CIDARS). *BMC Public Health*, 17(1), 1–10.
- Xiang, Y., Jia, Y. H., Chen, L. L., Guo, L., Shu, B. Z., & Long, E. S. (2021). COVID-19 epidemic prediction and the impact of public health interventions: A review of COVID-19 epidemic models. *Infectious Disease Modelling*, 6, 324–342. <https://doi.org/10.1016/j.idm.2021.01.001>
- Xu, Z. B., Zhang, H. M., & Huang, Z. Y. (2022). A continuous markov-chain model for the simulation of COVID-19 epidemic dynamics. *Biology*, 11(2), 190.
- Xu, J. F., & Zhou, X. N. (2011). Application of artificial neural networks in infectious diseases. *Chinese Journal of Parasitology & Parasitic Diseases*, 29(1), 49–54.
- Yadav, S. K., & Akhter, Y. (2021). Statistical modeling for the prediction of infectious disease dissemination with special reference to COVID-19 spread. *Frontiers in Public Health*, 9, Article 645405. <https://doi.org/10.3389/fpubh.2021.645405>
- Yang, W. Z. (2017). *Early warning for infectious disease outbreak*. Cambridge: Academic Press.
- Yang, Y. T., Horneffer, M., & DiLisio, N. (2013). Mining social media and web searches for disease detection. *Journal of Public Health Research*, 2(1), Article e2014. <https://doi.org/10.1186/1547-5949-2-1>
- Yang, Z. F., Zeng, Z. Q., Wang, K., Wong, S.-S., Liang, W. H., Zanin, M., ... Mai, Z. T. (2020). Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of Thoracic Disease*, 12(3), 165.
- Ying, S., Xu, Y. J., Dou, X. Y., Chen, X. Y., Zhao, J., & Guo, H. (2020). Spatial-temporal analysis of COVID-19 transmission based on geo-location linked data. *Geomatics and Information Science of Wuhan University*, 45(6), 798–807 (In Chinese).
- Zhai, M. M., Li, W. H., Tie, P., Wang, X. C., Xie, T., Ren, H., ... Qiu, L. X. (2021). Research on the predictive effect of a combined model of ARIMA and neural networks on human brucellosis in Shanxi Province, China: A time series predictive analysis. *BMC Infectious Diseases*, 21(1), 280. <https://doi.org/10.1186/s12879-021-05973-4>
- Zhang, X. X., Feng, L. Z., Lai, S. J., Ma, L. B., Zhang, T., Yang, J., ... Yang, W. Z. (2022). Research progress on early warning model of influenza. *Chinese Journal of Preventive Medicine*, 56(11), 1576–1583 (In Chinese).
- Zhang, Y. Y., Tang, S., & Yu, G. (2023). An interpretable hybrid predictive model of COVID-19 cases using autoregressive model and LSTM. *Scientific Reports*, 13(1), 6708. <https://doi.org/10.1038/s41598-023-33685-z>
- Zhao, W. P., Sun, Y. P., Li, Y., & Guan, W. M. (2022). Prediction of COVID-19 data using hybrid modeling approaches. *Frontiers in Public Health*, 10, Article 923978. <https://doi.org/10.3389/fpubh.2022.923978>
- Zhao, Z., Zhai, M. M., Li, G. H., Gao, X. F., Song, W. Z., Wang, X. C., ... Qiu, L. X. (2023). Study on the prediction effect of a combined model of SARIMA and LSTM based on SSA for influenza in Shanxi Province, China. *BMC Infectious Diseases*, 23(1), 71. <https://doi.org/10.1186/s12879-023-08025-1>
- Zheng, S., & Hu, X. M. (2021). Early warning method for public health emergency under artificial neural network in the context of deep learning. *Frontiers in Psychology*, 12, Article 594031.
- Zhou, T. Q. (2022). A linear regression prediction model of infectious disease spread based on baidu migration and effective distance. *Computational and Mathematical Methods in Medicine*, 2022.
- Zhou, J. N., Feng, Z. J., Tan, K., & Li, X. S. (2010). Study on the application of flexible spatial scan statistic to spatial aggregation of infection disease. *Chinese Journal of Disease Control & Prevention*, (6), 475–478 (In Chinese).
- Zhou, S., Shen, L., & Wang, W. (2020). An analysis and evaluation of the development trend of COVID-19 in Jiangsu Province from the perspective of spatiotemporal big data. *Modern Surveying and Mapping*, 43(3), 5–10 (In Chinese).
- Zhou, L. H., Yan, W. X., Li, S., Yang, H. X., Zhang, X. Y., Lu, W. L., ... Wang, Y. G. (2022). Cost-effectiveness of interventions for the prevention and control of COVID-19: Systematic review of 85 modelling studies. *Journal of Global Health*, 12, Article 05022. <https://doi.org/10.7189/jogh.12.05022>