RESEARCH NOTE

Open Access

Outbreak detector: a web application to boost disease surveillance systems and timely detection of infectious disease epidemics

Bushra Zareie[1,5], Jalal Poorolajal[1], Amin Roshani[2], Ahmed Menbari[3] and Manoochehr Karami[4*]

Abstract

Objective Digital technologies have improved the performance of surveillance systems through early detection of outbreaks and epidemic control. The aim of this study is to introduce an outbreak detection web application called OBDETECTOR (Outbreak Detector), which as a professional web application has the ability to process weekly or daily reported data from disease surveillance systems and facilitates the early detection of disease outbreaks.

Results OBDETECTOR generates a histogram that exhibits the trend of infection within a time range selected by the user. The output comprises red triangles and plus signs, where the former denotes outbreak days determined by the algorithm applied to the data, and the latter represents days identified as outbreaks by the researcher. The graph also displays threshold values and its symbols enable researchers to compute evaluation criteria for outbreak detection algorithms, including sensitivity and specificity. OBDETECTOR allows users to modify algorithm parameters based on their research objectives immediately after loading data. The implementation of automatic web applications results in immediate reporting, precise analysis, and prompt alert notification. Moreover, Public Health authorities and other stakeholders of surveillance can benefit from the widespread accessibility and user-friendliness of these tools, enhancing their knowledge and skills for better engagement in surveillance programs.

Keywords Automatic outbreak detection, Early warning systems, Outbreaks, Syndromic surveillance systems, Web application

Introduction

Disease surveillance is the process used to collect, manage, analyze, interpret and report reliable information about patients in the community. Public health surveillance is a continuous process of collecting health data in order to monitor the health status of communities and provide or revise the required services [1, 2]. Identifying outbreaks of infectious diseases is one of the controversial goals of public health surveillance. Early detection of outbreaks has always been a concern for public health. Government agencies and researchers define the successful performance of surveillance systems by their ability to timely identify public health threats, such as outbreaks,

*Correspondence:
Manoochehr Karami
man.karami@yahoo.com
[1]Department of Epidemiology, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran
[2]Department of Statistics, Lorestan University, Khorramabad, Iran
[3]Department of Computer Engineering, Sannandaj Branch, Islamic Azad University, Sanandaj, Iran
[4]Department of Epidemiology, School of Public Health and Safety, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[5]Bone and Joint Reconstruction Research Center, Department of Orthopedics, School of Medicine, Iran University of Medical Sciences, Tehran, Iran

because a faster investigation enables the prevention and containment of outbreaks and is very important for the implementation of control measures. In recent years, the surveillance system in relation to this goal has seen a very rapid and significant growth, which is due to two factors; new concerns about large-scale bioterrorism attacks and increased public awareness of emerging and re-emerging infections. These advances have led to the introduction of syndromic surveillance, increased databases, and the creation of automated outbreak detection systems to process data on large numbers of infections [3].

There is a wide list of methods for the early detection of epidemic signals through routine surveillance and syndromic surveillance. A very general classification of these methods is based on temporal clusters, which include dozens of outbreak detection algorithms in each of these categories. Therefore, the first main issue is that it is difficult for managers and experts in the field of health and public health to have a clear picture of all the available methods for continuous monitoring of the data of the surveillance system, or even to recognize the right time to use each of them. Another problem is that the algorithms used in biological surveillance have often been compared using real data, and there are only a few studies that have been reported using fully simulated data sets that can be reproduced by other researchers. On the other hand, real data sets are rarely and freely available, and this has reduced the possibility of conducting valid, real and meaningful studies around the world; Therefore, the second main problem is the difficulty of comparing algorithms based on the results of different studies and their evaluation [4–6]. Another important issue in the field of outbreak detection algorithms is the implementation capability of the algorithms so far. The ease of implementation of each method depends to some extent on the specialized knowledge of people. Some methods are quite complex and require high knowledge, expertise and experience, while others require basic knowledge. In general, control charts and the Shewhart method are easy to perform, but some other methods are more complex in implementation and interpretation [6, 7]. In summary, it can be said that although there are continuous advances in computer science and high processing power in data analysis, the development of user-friendly tools to implement and monitor the performance of surveillance systems is an important issue that has not been seriously addressed [8].

We will address the purpose of the study. According to the available evidence, there is no user-friendly web application (web-app) for implementing such aberration detection algorithms, and because of the importance of transferring existing knowledge related to outbreak detection algorithms in the context of the health system as a practical tool and the lack of such a set of detection tools and Epidemic management of communicable diseases in Iran, this study is proposed with the aim of setting up a specialized web-app. With this study, we intend to provide the guidelines for the use of algorithms to implement them as easily as possible in surveillance system databases. The role of surveillance systems, especially with the syndromic approach and outbreak detection algorithms, is very important in the field of infectious diseases for the timely detection of diseases such as the Covid-19 pandemic.

Material & methods

Web-app name, stated purpose of web-app and domain

The name of the web-app is OBDETECTOR, which stands for outbreak detector, aiming to automatically detect outbreaks. The domain of the outbreak detection web-app is available at [9] (Fig. 1).

Mechanisms for detecting outbreaks

The OBDETECTOR utilizes an array of outbreak detection algorithms. Since temporal algorithms are the most common type of algorithms in the syndromic surveillance system, in OBDETECTOR we have focused on temporal and spatio-temporal methods and six algorithms were selected to analyze the data of the surveillance system. Features to support outbreak response include Automatic notification to public health staff of outbreak signals for further follow-up.

Technical items of OBDETECTOR and web-app language

The web-app consists of a host for introducing the web-app and a server for running algorithms. Each algorithm has been programmed using R software and several packages, including surveillance, lubridate, msm, tidyverse, qcc; In addition to the coding of each algorithm, the design of the data entry panel and the setting of parameters have been done using R and Rshiny software. The hardware specifications of the server include 8GB of RAM and a Core i4 processor. The operating system is Linux with Ubuntu distribution and version 20.04. The programming languages used are R, Php, Html, Js, and Css. The database used is Mysql. Certain sections, such as the algorithm introduction, homepage, and data structure, are in Persian and english to facilitate communication with surveillance staff. However, the main sections, including arguments, data input/output are in English.

Algorithms

Farrington

Because a lot of surveillance data has significant scattering, a quasi-Poisson regression model was introduced by Farrington et al. (1996) and applied to early detection of outbreaks from reports received at the Communicable Disease Surveillance Center. Let y_i be the number

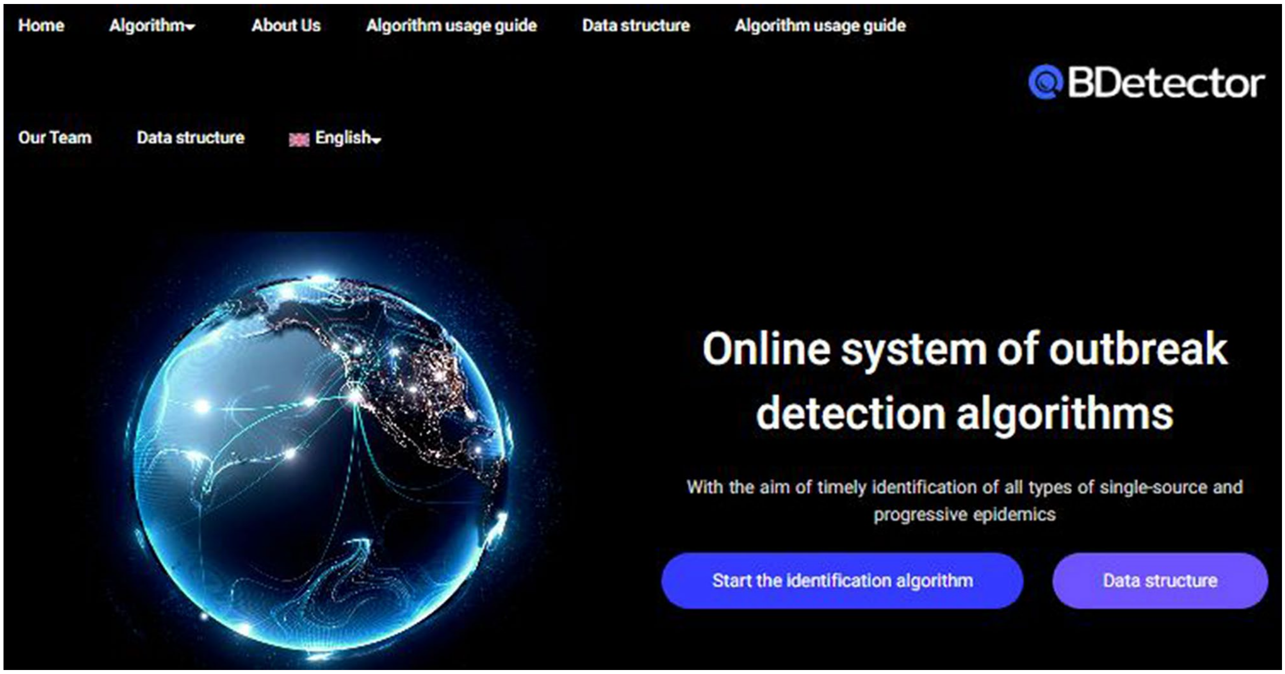

Fig. 1 The home page of the OBDETECTOR detection web-app

of reported cases of a disease under surveillance corresponding to week t_i, which independently have mean μ_i and variance $\varnothing\mu_i$. Considering a linear time trend in the disease frequency report, the regression model was defined by Farrington as

$$log\mu_i = \alpha + \beta t_i \quad (1)$$

,where t_i measures time on a weekly scale. Estimates of model parameters are calculated using the pseudo-likelihood method [10].

CUSUM

One of the control charts for detecting small changes is the Cumulative Sum (CUSUM), which was introduced by Page (1954) as a method for controlling continuous data. The CUSUM chart uses cumulative sums of deviations from the sample values to the target value, directly utilizing all the information in the sequence of sample values. Suppose the sample size is $n \geq 1$ and the mean of the first j samples is $\bar{x}_j$. If the target value for the mean of this process is μ_0, then the CUSUM chart is created by plotting against the sample number i. The quantity c_i is called the cumulative sum, which also includes the i-th sample [11].

$$c_i = \sum{}_{j=1}^{i} (x_j - \mu) \quad (2)$$

Farrington flexible

The Farrington Flexible algorithm is one of the algorithms based on GLM and the improved method of the Farrington algorithm, which was developed by Nuofaily et al. (2013). This model estimates the number of infections in the last week and includes a linear trend and a ten-level annual factor. This factor includes a seven-week reference period (one recent week, three past weeks, and three future weeks;$t_0 \pm 3$) and nine five-week periods each year. Considering b years in the past, this model also includes the number of comparable weeks in previous years. The corresponding linear log model is

$$log\mu_i = \theta + \beta t_i + \delta_{j(ti)} \quad (3)$$

where $j(t_i)$ is the seasonal factor corresponding to week t_i; With the assumption that $j(t_0) = 0$, $\delta_0 = 0$. In this model, a trend is always is fitted, except in special infections where the data are sparse [12].

EWMA

Exponentially weighted moving average (EWMA) was introduced by Roberts (1959) [43]. EWMA at time t is defined based on Y_t statistic as

$$Y_t = \lambda X_t + (1 - \lambda) Y_{t-1}, \ t = 1, 2, 3, \cdots \quad (4)$$

where X_t is the number of patients (cases) and λ is the weighting parameter or smoothing constant value and includes different values between $0 \leq \lambda \leq 1$. In Eq. (3), the starting value of EWMA, Y_0, is considered equal to the in control value μ_0. Sometimes the average primary data is used as the in control value, therefore

$$Y_0 = \bar{X} \tag{5}$$

.Suppose σ is the standard deviation of x_t, if Y_t is greater than the warning threshold level, the outbreak warning will be announced [13]. The warning threshold level or upper control limit of EWMA is

$$UCL = \mu_0 + L\sigma\sqrt{\frac{\lambda}{(2-\lambda)}\left[1-(1-\lambda)^{2t}\right]} \tag{6}$$

EARS

The Early Aberration Reporting System (EARS) is designed to provide advanced surveillance for short periods around events such as the Olympic Games, for which there is generally little or no prior information. EARS was first introduced by the centers for disease control and prevention, and since September 11, 2001, has been used as a standard surveillance system in many US local health departments; EARS also applies to New Zealand notifiable disease surveillance data and is updated weekly. The primary purpose of EARS is to provide multiple aberration detection methods to national, state, and local health departments and allow users to change sensitivity and specificity thresholds to values considered important to public health by state and local health departments after selecting valid aberration detection methods. In this algorithm, three early detection methods named C1-MILD, C2-MEDIUM and C3-HIGH have been implemented. The terms mild, medium and ultra refer to the level of sensitivity of the three statistical methods. C1-MILD and C2-MEDIUM are actually types of Shewhart charts that use the moving average and sample standard deviation to standardize each observation, and C3 is the result of combining information based on C2. Threshold values in all three methods, C1-MILD, C2-MEDIUM and C3-HIGH, are obtained using the one-way CUSUM method [14].

HMM

Strat and Karat (1999) proposed the use of Hidden Markov models (HMM) for monitoring epidemiological data. HMM have previously been used in many fields, including electrocardiographic signal analysis, seizure frequency analysis in epilepsy, and meteorology. The main idea of this method is that it divides the time series of registered diseases into two parts, the epidemic period and the non-epidemic period. Assume that y_t for $t = 1, 2, \cdots, n$ is an observed value of the random process $Y = (Y_t; t = 1, 2, \cdots, n)$ and is associated with a hidden variable such as S_t that defines the conditional distribution of Y. If $S_t = j$, the conditional distribution of Y_t has density

$$Y_t \mid \mathrm{j} \sim f_{jt}(y_t; \theta_j), j = 1, 2, \cdots, \mathrm{m} \tag{7}$$

so that f_{jt} is a predetermined density such as Poisson or Gaussian distribution and θ_j is a parameter to be estimated. It is assumed that the hidden sequence S_t for $t = 1, 2, \cdots, n$ follows a two-state homogeneous Markov chain of order 1 with the following fixed transition probabilities

$$p_{kl} = P(S_{t+1} = l \mid S_t = k);\ k.l \in \{0.1\} \tag{8}$$

For example, suppose that y_t is the observed incidence rate of Influenza-like Illness (ILI) in week t and there are two distributions corresponding to the incidence rate of ILI in the epidemic and non-epidemic periods;p_{01} for $j = 0, 1$ is the probability of changing from the non-epidemic period to the epidemic period [15].

Algorithm selection guide and data structure

In OBDETECTOR, a section called "Algorithm Selection Guide" is designed. In this section, it is possible for users to choose the appropriate algorithm to run on the data based on the characteristics of the data; This classification is provided based on the review of the literature and the development and expansion of the latest results in the field of temporal and single variable algorithms [4]. It should be noted that based on the different states of the incidence level and the existence of trends in the data, a set of algorithms and not just one specific algorithm is suggested; failure to propose unique algorithms is due to the different performance of algorithms based on different evaluation criteria (Fig. 2). Let us explain this part a little better with some examples. One of the challenges for researchers in the outbreak detection algorithm field is that it is difficult to definitively determine which algorithm performs better. Faverjon's article discusses the factors that influence the performance of outbreak detection algorithms in detail. These factors include the type of data being monitored, the available historical baseline, the presence of trends in the data, the presence of auxiliary variables, and the level of disease incidence.

Figure 2 in the article has outlined the recommended use of 6 different algorithms based on the level of incidence and the presence of trends, as per Faverjon's guidance. For instance, for a disease like brucellosis, which has a trend in endemic areas, it is preferable to use the Farrington Flexible algorithm and the Farrington algorithm (GLM-based methods) for outbreak detection. On the other hand, for a disease like COVID-19 that has a

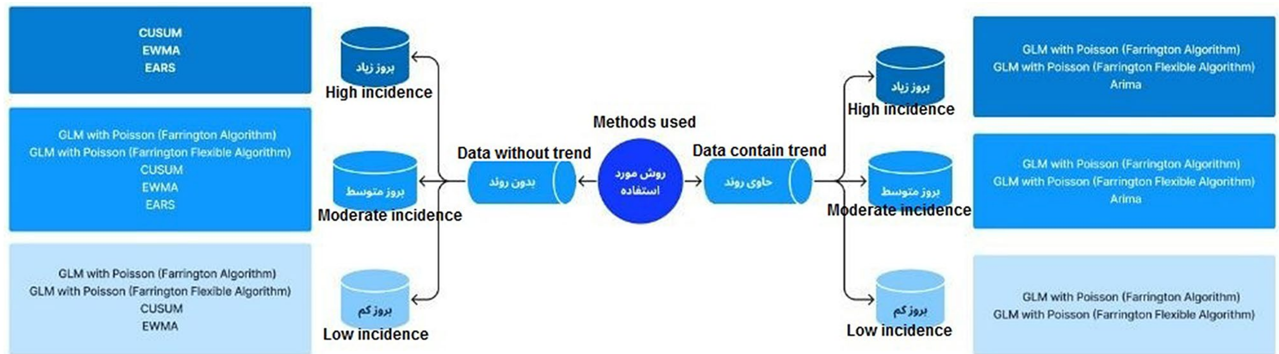

Fig. 2 Algorithm selection guide based on the incidence level and trend

high incidence level but no clear trend, algorithms such as EWMA, EARS, and CUSUM are more suitable.

By providing these concrete examples, we can better illustrate the nuances involved in selecting the appropriate outbreak detection algorithm based on the characteristics of the disease and data under consideration.

OBDETECTOR web-app has the possibility to read data in Excel (.xlsx) format. The data should contain two columns; count and state. In the count column, there is the disease/syndrome frequency data based on time (daily/weekly/monthly) and in the state column, there is data related to the outbreak status (presence of outbreak=1, absence of outbreak=0). In the section related to setting the parameters of each algorithm, information about the start time of the data is received from the user so that time periods can be built automatically based on the daily/weekly/monthly frequency of the data. So there is no need to enter the time of the data in the uploaded file, the date of the data is automatically created based on entering the time of the first data in the panel of each algorithm.

In situations where the researcher's goal is to evaluate the performance of an algorithm or compare the performance of two algorithms, the data related to the outbreak situation is recorded and used as a gold standard based on previous evidence, the researcher's experience, or the results of a valid algorithm; If the researcher's goal is only to identify the outbreak using an algorithm, the variable values of the outbreak state are assumed to be zero and entered.

Results

Main page

The home page includes the logo, title, menu, call buttons, text and various visual elements. The web-app logo is located on the right side and at the top of the page. In the outbreak detection web-app, the title is inserted below the logo and the purpose of the web-app and the type of service provided are shown. The web-app menu is located on the main page and includes icons such as algorithms, data structure, algorithm selection guide, about web-app and our team (Fig. 1). The features and capabilities of the web-app and the introduction of algorithms are briefly listed on the main page.

Algorithms

The pages related to each algorithm are organized in two sections. In the "Description" section, a guide table of the arguments of each algorithm is included, and below the table, a brief introduction of each algorithm is discussed along with the introduction of the main sources for further study. On the left side of the " implementation" section, the arguments panel of each algorithm is located in order to receive inputs and data files, and the results of the implementation of each algorithm appear on the right side. The guide table of arguments for each algorithm is shown in (Table 1). These arguments were selected through the surveillance package [16]. The user panel in the implementation section of each algorithm is also shown in supplementary file (S1-S6).

Output OBDETECTOR

The outputs of running each algorithm include a histogram of the infection trend, threshold values, and a vector of alarm values for every time point in the range. Some examples of the implementation of outbreak detection algorithms on real surveillance data are shown in Sect. 3.5. The data was from the surveillance system of the Center for Disease Control and Prevention in Iran.

Data structure, algorithm selection guide, about the web-app and our team

In the "Data structure" section, it is dedicated to the data loading guide including data format, data type, required variables. In the "Algorithm Selection Guide" section, Fig. 2 is placed for users to help them make a better choice. In the section of "About the web-app", the definition of the surveillance system and its importance in the health system, the successful performance of surveillance systems and the importance of early detection

Table 1 Guide table of algorithms arguments

Arguments	Algorithms that use this argument	Description
Start Year	All algorithms[1]	Enter the start year of the data
Start Month	All algorithms[1]	Enter the start month of the data
Start Day	Farrington, Farrington Flexible, EWMA, EARS, HMM	Enter the data start date
Frequency	Farrington, CUSUM, Farrington Flexible, EWMA, HMM	If the data scale is weekly, select the number 52 and if the data scale is monthly, select the number 12
Range of Data(min)	All algorithms[1]	Enter the beginning of range under review
Range of Data(max)	All algorithms[1]	Enter the end of the range under review
Choose xlsx File	All algorithms[1]	In this section, upload the data file. The data format should be according to the data structure
b	Farrington, Farrington Flexible	How many years back in time to include when forming the base counts.
w)windows size(	Farrington, Farrington Flexible.	Windows size, i.e. number of weeks to include before and after the current week
Alpha	Farrington, Farrington Flexible, EARS	An approximate (two-sided) $(1-\alpha)$ prediction interval is calculated
Reference Value (k)	CUSUM	The reference value
Decision Interval (H)	CUSUM	The decision boundary
weightsThreshold	Farrington Flexible	A scalar indicating when observations are seen as outlier. In the original Farrington proposal the value was 1 (default value), in the improved version this value is suggested to be 2.58
Target Value	EWMA,	Specify the in control value μ_0
Standard Deviation	EWMA	Specify the standard deviation
Lambda	EWMA	Specify the weighting parameter or smoothing constant value
L	EWMA	Enter the width of the control limit
Method	EARS	String indicating which method to use: "C1" for EARS C1-MILD method (Default), "C2" for EARS C2-MEDIUM method, "C3" for EARS C3-HIGH method
Baseline	EARS	How many time points to use for calculating the baseline,
Min Sigma	EARS	By default 0. If minSigma is higher than 0, for C1 and C2, the quantity zAlpha * minSigma is then the alerting threshold if the baseline is zero. Howard Burkom suggests using a value of 0.5
Mtilde	HMM	Number of observations back in time to use for fitting the HMM (including the current observation)
Trend	HMM	Boolean stating whether a linear time trend exists

1: All algorithms include Farrington, CUSUM, Farrington Flexible, EWMA, EARS, HMM

of outbreaks, the purpose of the outbreak detection web-app and its most important capabilities, along with the introduction of the algorithms loaded in the web-app, have been briefly discussed. In the "our team" section, the developers and investors of the online outbreak detection web-app have also been introduced.

A practical example with OBDETECTOR web-app

Comparing the performance of EWMA and EARS algorithms in detecting the COVID-19 epidemic

In this part, we want to make a comparison between two different algorithms implemented with the web application using the data of COVID-19.

To compare the performance of the EWMA and EARS algorithms in detecting the COVID-19 epidemic, we need a reliable reference or "golden standard" against which to evaluate the results. This reference could be the output of another algorithm, the findings of an alternative surveillance system (such as routine surveillance), or the expert opinion of an epidemiologist. Some of the input parameters were kept consistent across the implementation of both algorithms. These included the start date of the data (2020-02-20) and the evaluation period (last one year, with a range of 200 to 320 days). For the EARS algorithm, the other user-defined parameters were: Baseline = 7 days, Alpha = 0.05, and Min sigma = 0. To better visualize the algorithm performance, a shorter time frame (last one year) was selected for the analysis (Fig. 3). In the case of the EWMA method, the daily COVID-19 data was used, with the following user panel parameters: Frequency = Daily, Target Value = 1360.77 (based on the average of the recorded days), Standard Deviation = 1107.69, Weighting Parameter (Lambda) = 0.25, and Control Limit Width (L) = 2.5. By keeping the input parameters consistent between the two algorithms, we can make a more meaningful comparison of their performance in detecting the COVID-19 epidemic. Different metrics can be used to compare the performance of the

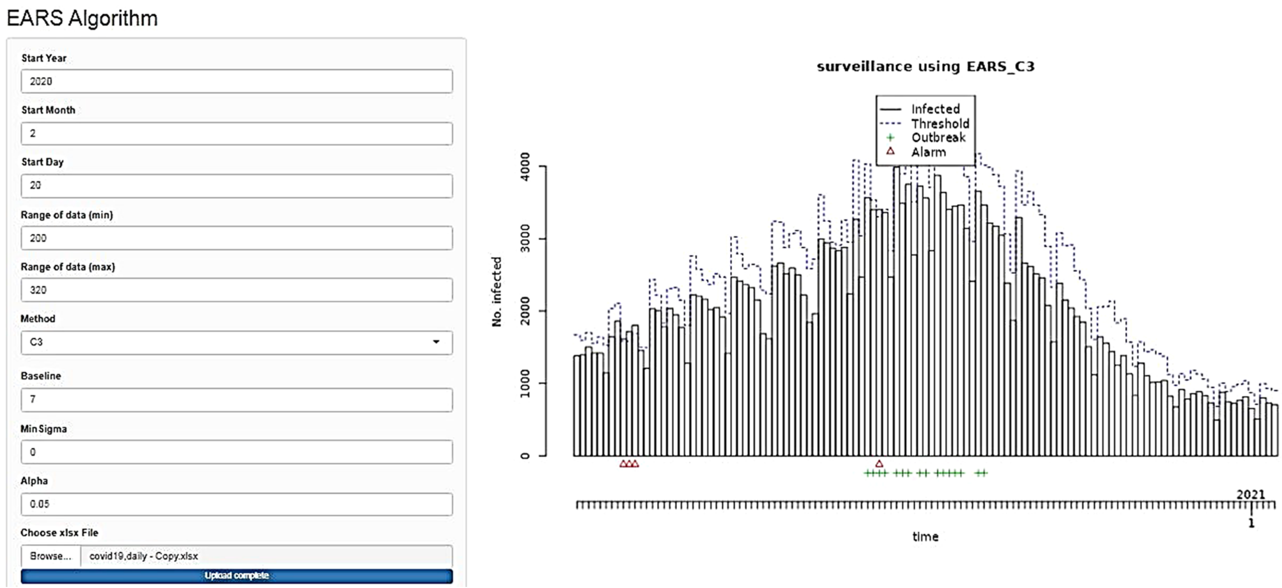

Fig. 3 Applying the EARS algorithm to COVID-19 data from 2020 to 2021 in the the OBDETECTOR web application

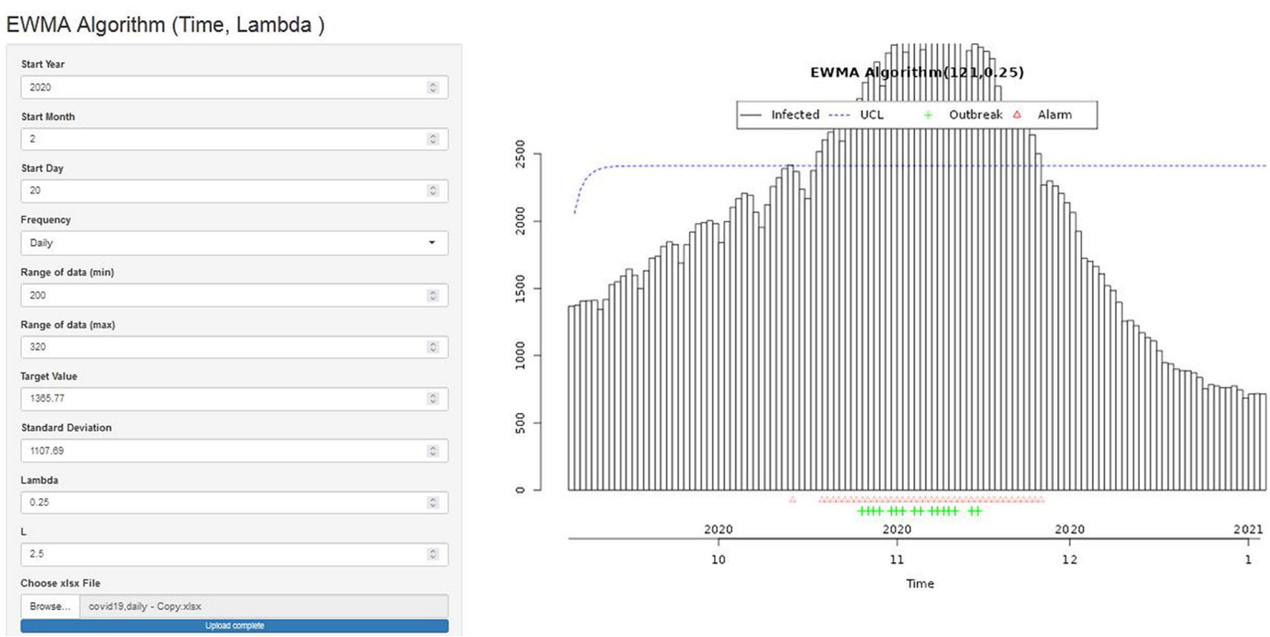

Fig. 4 Applying the EWMA algorithm to COVID-19 data from 2020 to 2021 in the the OBDETECTOR web application

Table 2 Comparing the performance of EWMA and EARS algorithms in detecting the COVID-19 epidemic

Disease	Algorithm	Sensitivity	Specificity	False positive rate	Timeliness	Figure
Covid-19	EARS-C3	6%	97%	3%	3	Figure3
Covid-19	EWMA	100%	77%	23%	0	Figure4

algorithms, such as sensitivity, specificity, false positive rate, and timeliness (Fig. 4). The values for these metrics were calculated based on the output of each algorithm and are presented in Table 2.

Discussion

Principal results

The online web-app of outbreak detection algorithms has been launched in the Iranian domain, with the aim of timely detection of all types of point-source and progressive epidemics. The most important capabilities of this web-app is the introduction of an accepted classification system for detecting changes in the data of the surveillance system based on the trend and level of incidence, implementation of various outbreak detection algorithms, the possibility of changing the parameters related to each algorithm based on the characteristics of the data or the goals of the researcher, facilitating the selection of outbreak detection algorithms from a wide list of available algorithms, the possibility of receiving the results of applying various algorithms immediately after uploading the data of the surveillance system, providing outputs with the same structure in order to compare the performance of outbreak detection algorithms, maintaining data confidentiality, the possibility of continuous monitoring of known diseases/syndromes under the surveillance system. It is important to recall that this question may arise: why don't we use threshold definitions to estimate the start of an outbreak, and what is the need for outbreak detection algorithms? There are clear definitions of the start of outbreaks for many diseases such as measles and Ebola. In response, we must say that the primary place for outbreak detection algorithms is in syndromic surveillance systems, where there are no clear definitions of different thresholds for various syndromes. We can analyze various non-clinical data sources such as school absenteeism, over-the-counter medication sales, and peak disease or syndrome searches on Google, using outbreak detection algorithms. Failing to use outbreak detection algorithms would make syndromic surveillance systems ineffective. However, for many diseases under surveillance, the definition of the start and even the end of an outbreak may not be clear, and outbreak detection algorithms can still be helpful for these diseases in the routine surveillance system. Establishing a timely warning system and quick response to disease outbreaks and events with unknown causes affecting people's health and matters related to international health through additional tools in the surveillance system, compilation of specialized tools in the management of infectious diseases to meet the educational and skill needs of working employees and related to disease control at different levels has always been the main goals of disease control and prevention centers and organizations. The current study is a technological study in order to achieve the above goals; digital health technologies, such as web-based and smartphone applications, offer unique and useful opportunities for timely and effective responses to infectious disease outbreaks. This is certainly confirmed by the rapid innovations and design of digital tools in response to the current COVID-19 pandemic, the most visible of which are mobile phone-based contact tracing applications [17]. This web application has the capability for individuals to independently upload data. It is also interoperable and can be integrated into healthcare surveillance systems for use. Stakeholders can implement these algorithms on their data alongside the daily recording of various syndromes, so that the monitoring system can be fully executed.

Six algorithms are included in the OBDETECTOR web-app. The final selection of algorithms was based on three criteria: algorithm validity, variety of algorithms, need for automatic processing; which will be briefly explained for each criterion. The most valid algorithms have been discussed and compared in review articles and systematic review articles. In some articles, frameworks have been designed based on the positive and negative features of each algorithm. Six valid algorithms were selected in this way. Algorithms were selected based on various classes of statistical methods, such as Shewhart charts, generalized linear models (GLM), time series, statistical process control methods, etc. In some methods, there are more than a dozen types of algorithms [3, 4]; for the final selection of algorithms in some categories such as GLM, different types of algorithms were programmed and implemented one by one, and finally, two better algorithms were selected based on the evaluation criteria. The creation of automatic analysis systems for those algorithms that lead to the significant facilitation of their implementation was prioritized.

Each outbreak detection algorithm includes a statistical method that requires inputs under the name of arguments or parameters for its implementation. Arguments can be defined in two forms: dynamic or static. Dynamic arguments can be viewed and set in the algorithm implementation panel; static arguments are initialized in the coding process of each algorithm and are not visible during algorithm implementation. For example, in Farrington's algorithm, frequency is a dynamic argument in which the data is defined as daily, weekly or monthly, and powertrans is a fixed argument related to data transformation, which is set to the value of 2.3 among several choices, and when the algorithm is run, this transformation is applied to the data without the intervention of the researcher.

In the case of the implemented example, it is necessary to draw your attention to the technical point of how this web app can help in early detection of outbreaks. The example we provided for COVID-19 is a real-world example of data recorded by the routine surveillance system. In this example, the classic definition of an outbreak, i.e., values exceeding the expected range, was used as the gold standard. Based on the Timeliness[1] and POD-1week[2] indicators, we compared the performance of these two algorithms against the gold standard.

[1] Detection on the first outbreak day is equivalent to a timeliness of 0 days.

[2] Probability of Detection during the first week (POD-1week): which makes it possible to evaluate the methods' ability to enable early control measures.

Let's assume we used the EWMA algorithm to monitor COVID-19 data. This algorithm, based on the Timeliness indicator, identified the start of the outbreak without any delay. Based on the POD-1week indicator, this algorithm detected the outbreak in the first week. In the EARs algorithm, the Timeliness value is 3, and the outbreak was detected in the first week. You can see that different algorithms have diverse results. If the policymaker's goal is to identify the outbreak early, they can use the EWMA algorithm to achieve this goal, as this algorithm has also demonstrated the ability to detect the outbreak early in this example.

Limitation

Our aim is to support knowledge exchange among researchers and public health authorities in the area of outbreak detection algorithms. As a first step, we acknowledge that there may be some limitations in our web-app. However, through enhanced communication with our users, we have decided to address this limitation by providing a Persian language option on the system's homepage, as there is currently no such application available in our country. Another limitation of our study was the lack of access to syndromic data. We suggest that researchers use the system to identify outbreaks by applying algorithms to the data from the syndromic surveillance system. We welcome feedback from health managers and users to improve the number and types of algorithms and other languages offered in the future.

Comparison with prior work

So far, 15 web-based programs have been introduced, with the aim of setting up an outbreak early warning web-app, improving reporting, data sharing, rapid response to epidemics and setting thresholds. In two of the studies, such as the OBDETECTOR web-app, there is the ability to automatically detect outbreaks. In other programs, there is a need to load data into data analysis programs [18, 19]. Like OBDETECTOR, most of the introduced web-based apps can be used at the national and regional levels, and some of them are also used in emergency situations and epidemiology field activities [20]. The number and variety of algorithms presented are the clear differences between OBDETECTOR and the introduced web-based apps. In most of them, only one or two outbreak detection methods have been used. The methods of Farrington, Sa TScan, Stroup and fixed thresholds have been the most used methods. The Smi Net web-app had the largest number of used algorithms, including Sa TScan, Farrington, Simple threshold and Outbreak [21], and in many programs, the detection mechanism is not clearly mentioned [22]. One of the similar web- based apps with OBDETECTOR is the response approach to the epidemic. In all of them, a warning notification is automatically displayed to the user for further follow-up. In the OBDETECTOR web-app, it is not possible to identify the type of syndrome and disease and the desired location for the owner of the web-app by uploading data to the user. The principles of maintaining confidentiality exist only in two of the similar web-based apps [20, 23]. Ease and flexibility of use in web-based applications have been mentioned in 10 cases [18, 19, 21–28]. In the OBDETECTOR, a brief introduction of each algorithm, algorithm selection guide and deployment of an arguments panel with the dynamic parameters have been discussed in order to increase users' knowledge and interaction with the web-app argument; In exciting web-based apps for outbreak detection, increasing the skill and ability of users and the possibility of interacting with the program are available in two programs each [19, 24–26].

Conclusion

Technological advances in outbreak detection algorithms allow for continuous monitoring of known diseases or syndromes, facilitating early detection and implementation of protective measures to limit the impact of outbreaks. Automated web-apps modernize paper-based surveillance systems, providing immediate reporting, accurate analytics, and real-time alerts. These tools offer wide access and ease of use for health managers and workers, increasing their knowledge and skills to participate in higher quality surveillance activities. This tool enables better planning and resource allocation, preventing outbreak progression and providing optimal infrastructure for surveillance systems to face epidemics. This benefits public health authorities, surveillance staff, and stakeholders of surveillance.

Abbreviations

CUSUM	Cumulative Sum
EARS	Early Aberration Reporting System
EWMA	Exponentially Weighted Moving Average
GLM	Generalized Linear Models
HMM	Hidden Markov Models
ILI	Influenza-like Illness
Web-app	Web Application

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13104-024-06892-8.

Supplementary Material 1

Acknowledgements

These results were obtained as a part of a Ph.D. thesis in Epidemiology. The authors would like to appreciate the Vice-Chancellor for Research and Technology of the Hamadan University of Medical Sciences, Hamadan, Iran, for approval and financial support of this study (No. 140009237703).

Author contributions
Conceptualization: Bushra Zareie: Investigation: Bushra Zareie, Jalal Poorolajal, Amin Roshani, Manoochehr Karami: Data curation: Amin Roshani: Formal analysis: Bushra Zareie, Amin Roshani, Ahmed Menbari: Project administration: Manoochehr Karami.:Supervision: Jalal Poorolajal, Manoochehr Karami. Writing – original draft: BushraZareie, Jalal Poorolajal, Amin Roshani, Manoochehr Karami, Ahmed Menbari. Writing – review &editing: Bushra Zareie, Jalal Poorolajal, Amin Roshani, Ahmed Menbari, Manoochehr Karami.

Funding
This work was supported by the Vice-Chancellor of Research and Technology, Hamadan University of Medical Sciences (No. 140009237703).

Data availability
The datasets used during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 9 March 2024 / Accepted: 6 August 2024
Published online: 20 August 2024

References

1. Organization WH. International health regulations (2005). World Health Organization; 2008.
2. Shokoohi M, et al. Drug use patterns and associated factors among female sex workers in Iran. Addict Behav. 2019;90:40–7.
3. Unkel S, et al. Statistical methods for the prospective detection of infectious disease outbreaks: a review. J Royal Stat Society: Ser (Statistics Society). 2012;175(1):49–82.
4. Faverjon C, Berezowski J. Choosing the best algorithm for event detection based on the intended application: a conceptual framework for syndromic surveillance. J Biomed Inform. 2018;85:126–35.
5. Buckeridge DL, et al. Understanding detection performance in public health surveillance: modeling aberrancy-detection algorithms. J Am Med Inform Assoc. 2008;15(6):760–9.
6. De Vries A, Reneau J. Application of statistical process control charts to monitor changes in animal production systems. J Anim Sci. 2010;88(suppl13):E11–24.
7. O'Brien S, Christie P. Do CuSums have a role in routine communicable disease surveillance? Public Health. 1997;111(4):255–8.
8. Dorea FC, Lindberg A, Elvander M. Veterinary syndromic surveillance in practice: costs and benefits for governmental organizations. Infect Ecol Epidemiol. 2015;5(1):29973.
9. Zareie B, Poorolajal KM, Roshani J. A. OBDETECTOR. 2023 June 16,2023]; https://obdetector.ir/
10. Farrington C, et al. A statistical algorithm for the early detection of outbreaks of infectious disease. J Royal Stat Society: Ser (Statistics Society). 1996;159(3):547–63.
11. Page ES. Continuous inspection schemes. Biometrika. 1954;41(1/2):100–15.
12. Noufaily A, et al. An improved algorithm for outbreak detection in multiple surveillance systems. Stat Med. 2013;32(7):1206–22.
13. Roberts S. Control chart tests based on geometric moving averages. Technometrics. 2000;42(1):97–101.
14. Hutwagner L, et al. A simulation model for assessing aberration detection methods used in public health surveillance for systems with limited baselines. Stat Med. 2005;24(4):543–50.
15. Le Strat Y, Carrat F. Monitoring epidemiologic surveillance data using hidden Markov models. Stat Med. 1999;18(24):3463–78.
16. Höhle M et al. Temporal and spatio-temporal modeling and monitoring of Epidemic Phenomena. Surveillance R Package Version, 2017. 1(0).
17. Quinn E, et al. Web-based apps for responding to acute infectious disease outbreaks in the community: systematic review. JMIR Public Health Surveillance. 2021;7(4):e24330.
18. Widdowson M-A, et al. Automated, laboratory-based system using the internet for disease outbreak detection, the Netherlands. Emerg Infect Dis. 2003;9(9):1046.
19. Hulth A, et al. Practical usage of computer-supported outbreak detection in five European countries. Eurosurveillance. 2010;15(36):19658.
20. Li Y-p, et al. Decision support system for the response to infectious disease emergencies based on WebGIS and mobile services in China. PLoS ONE. 2013;8(1):e54842.
21. Rolfhamre P, et al. SmiNet-2: description of an internet-based surveillance system for communicable diseases in Sweden. Eurosurveillance. 2006;11(5):15–6.
22. El-Khatib Z, et al. SMS-based smartphone application for disease surveillance has doubled completeness and timeliness in a limited-resource setting–evaluation of a 15-week pilot program in Central African Republic (CAR). Confl Health. 2018;12:1–11.
23. Groeneveld GH, et al. ICARES: a real-time automated detection tool for clusters of infectious diseases in the Netherlands. BMC Infect Dis. 2017;17:1–9.
24. Krause G, et al. SurvNet electronic surveillance system for infectious disease outbreaks. Ger Emerg Infect Dis. 2007;13(10):1548.
25. Salmon M, et al. A system for automated outbreak detection of communicable diseases in Germany. Eurosurveillance. 2016;21(13):30180.
26. Straetemans M, et al. Automatic outbreak detection algorithm versus electronic reporting system. Emerg Infect Dis. 2008;14(10):1610.
27. Yang W, et al. A nationwide web-based automated system for outbreak early detection and rapid response in China. Western Pac Surveillance Response Journal: WPSAR. 2011;2(1):10.
28. Zhang H, et al. Evaluation of the performance of a dengue outbreak detection tool for China. PLoS ONE. 2014;9(8):e106144.

Publisher's Note