

# 文献汇报

Advances in Artificial Intelligence for Infectious-Disease  
Surveillance

2023-12-12

李康国

Press Space for next page →



# 目录

1. [目录](#)
2. [基本介绍](#)
3. [文章结构](#)
4. [Abstract](#)
  1. [Part 2/3](#)
5. [AI applications in disease surveillance](#)
  1. [实现](#)
  2. [Pathogen classification](#)
  3. [Source identification](#)
  4. [Risk assessment](#)
6. [Extended applications](#)
7. [Surveillance roadblocks and future directions](#)
  1. [Data source representation](#)
  2. [The limits of AI](#)
8. [Learn More](#)

# 基本介绍

《Advances in Artificial Intelligence for Infectious-Disease Surveillance》2023年4月27日以Review形式发表在《The New England Journal of Medicine》。作者分别来Harvard Medical School, Beijing Normal University.

THE NEW ENGLAND JOURNAL OF MEDICINE

REVIEW ARTICLE

AI IN MEDICINE

Jeffrey M. Drazen, M.D., Editor, Isaac S. Kohane, M.D., Ph.D., Guest Editor,  
and Tze-Yun Leong, Ph.D., Guest Editor

## Advances in Artificial Intelligence for Infectious-Disease Surveillance

John S. Brownstein, Ph.D., Benjamin Rader, M.P.H.,  
Christina M. Astley, M.D., Sc.D., and Huaiyu Tian, Ph.D.

FLORENCE NIGHTINGALE'S INNOVATIVE "ROSE DIAGRAM" OF PREVENTABLE deaths revolutionized data-driven disease surveillance.<sup>1</sup> Raw hospital mortality data collected during the Crimean War were transformed into a compelling, visual insight — poor sanitary conditions killed more people than battle wounds did. This act of synthesizing noisy, complex data into an elegant, effective message was the foundation for a royal commission to track morbidity and mortality and thus launched a new era in which analytic methods were used to better monitor and manage infectious disease. In the more than 160 years since the first publication of Nightingale's rose diagram, tools and technology for translating high-density data and uncovering hidden patterns to provide public health solutions have continued to evolve. Manual techniques are now complemented by machine-learning algorithms. Artificial intelligence (AI) tools can now identify intricate,

From the Computational Epidemiology Laboratory (J.S.B., B.R., C.M.A.) and the Division of Endocrinology (C.M.A.), Boston Children's Hospital, Harvard Medical School (J.S.B., C.M.A.), and Boston University School of Public Health (B.R.), Boston, and the Broad Institute of MIT and Harvard, Cambridge (C.M.A.) — all in Massachusetts; and the State Key Laboratory of Remote Sensing Science and Center for Global Change and Public Health, Beijing Normal University, Beijing (H.T.). Dr. Brownstein can be contacted at [john.brownstein@childrens.harvard.edu](mailto:john.brownstein@childrens.harvard.edu) or at Boston Children's Hospital, 300 Longwood Ave., BCH3125 Bldg., Boston, MA

# 文章结构

- **Abstract** - 从南丁格尔玫瑰图到AI，总结COVID-19中AI技术应用
- **AI application in disease surveillance**
  - **Early warning**
  - **Pathogen classification**
  - **Risk assessment**
  - **Source identification**
  - **Hotspot detection**
  - **Tracking and forecasting**
- **Extended applications**
- **Surveillance roadblocks and future directions**
  - **Data volume and quality**
  - **Data source representation**
  - **Privacy**
  - **The limits of AI**

# Abstract

## Part 1/3

**Florence Nightingale’s innovative “rose diagram” of preventable deaths revolutionized data-driven disease surveillance.** Raw hospital mortality data collected during the Crimean War were transformed into a compelling, visual insight — poor sanitary conditions killed more people than battle wounds did. This act of synthesizing noisy, complex data into an elegant, effective message was the foundation for a royal commission to track morbidity and mortality and thus launched a new era in which analytic methods were used to better monitor and manage infectious disease. In the more than 160 years since the first publication of Nightingale’s rose diagram, tools and technology for translating high-density data and uncovering hidden patterns to provide public health solutions have continued to evolve. **Manual techniques are now complemented by machine learning algorithms.** Artificial intelligence (AI) tools can now identify intricate, previously invisible data structures, providing innovative solutions to old problems. Together, these advances are propelling infectious-disease surveillance forward.

□ 回溯历史，将监测方式的转变: noisy, complex data→elegant, effective diagram→manual techniques→machine learning algorithms

□ 这个是摘要，为啥要这样子写？感觉更像是背景，但是这个论文没有背景部分。

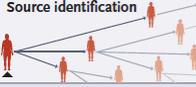
## Part 2/3

The coronavirus disease 2019 (Covid-19) pandemic has highlighted the speed with which infections can spread and devastate the world — and the extreme importance of an equally nimble, expeditious, and clever armamentarium of public health tools to counter those effects. Throughout this crisis, we have witnessed a multitude of AI solutions deployed to play this role — some much more successful than others. As new pathogens emerge or old challenges return to command our attention, the incorporation of the lessons learned into our public health playbook is a priority. In this review article, we reflect on the effects of new and longstanding AI solutions for infectious-disease surveillance. AI applications have been shown to be successful for a diverse set of functions, including **early-warning systems,2,3 hotspot detection,4,5 epidemiologic tracking and forecasting,6,7 and resource allocation8** (Fig. 1). We discuss a few recent examples.<sup>9,11,12</sup> We begin with how AI and machine learning can power early-warning tools and help distinguish among various circulating pathogens (e.g., severe acute respiratory syndrome coronavirus 2 SARS-CoV-2 vs. influenza virus). We then discuss AI and machine-learning tools that can backtrack epidemics to their source and an algorithmic method that can direct an efficient response to an ongoing epidemic. Finally, we emphasize the critical limitations of AI and machine learning for public health surveillance and discuss salient considerations to improve implementation in the future.

□ 介绍文章的结构

□ 这个综述与系统综述不一样。

# Part 3/3

Function	Examples
<b>Early warning</b> 	<ul style="list-style-type: none"> <li>Natural-language processing of news sources to identify outbreaks (Freifeld et al., <i>JAMIA</i> 2008)</li> <li>Unsupervised machine learning of social media data to detect unknown infections (Lim, Tucker, and Kumara, <i>J Biomed Inform</i> 2017)</li> </ul>
<b>Pathogen classification</b> 	<ul style="list-style-type: none"> <li>Convolutional neural network model for reading antibiograms (Pascucci et al., <i>Nat Commun</i> 2021)</li> <li>Convolutional neural network model to automate malaria microscopy and diagnosis (Liang et al., <i>IEEE</i> 2016)</li> </ul>
<b>Risk assessment</b> 	<ul style="list-style-type: none"> <li>Reinforcement learning of Covid-19 positivity rates to target limited testing in Greece (Bastani et al., <i>Nature</i> 2021)</li> <li>Machine-learning models including random forest and extreme gradient boosting to use syndromic surveillance for Covid-19 risk prediction (Dantas, <i>PLoS One</i> 2021)</li> </ul>
<b>Source identification</b> 	<ul style="list-style-type: none"> <li>Automated data mining of electronic medical records to uncover hidden routes of infection transmission (Sundermann et al., <i>Clin Infect Dis</i> 2021)</li> <li>Supervised machine learning in combination with digital signal processing for genomic tracing of Covid-19 (Randhawa et al., <i>PLoS One</i> 2020)</li> </ul>
<b>Hotspot detection</b> 	<ul style="list-style-type: none"> <li>Neural computing engine to correlate sound from hospital waiting rooms with influenza spikes (Al Hossain et al., <i>Proc ACM Interact Mob Wearable Ubiquitous Technol</i> 2020)</li> <li>Multilayer perceptron artificial neural network model to detect spatial clustering of tuberculosis (Mollalo et al., <i>Int J Environ Res Public Health</i> 2019)</li> </ul>
<b>Tracking and forecasting</b> 	<ul style="list-style-type: none"> <li>Real-time stacking of multiple models to improve forecasts of seasonal influenza (Reich et al., <i>PLoS Comput Biol</i> 2019)</li> <li>Machine learning to combine new data sources for monitoring Covid-19 (Liu et al., <i>J Med Internet Res</i> 2020)</li> </ul>
<p><b>Figure 1. Various Functions of Artificial Intelligence (AI) for Infectious-Disease Surveillance.</b>            Shown is a nonexhaustive list of functions of AI-aided infectious-disease surveillance and representative examples from the published literature.<sup>2,13</sup> Each example includes the type of AI algorithm, a brief description of its purpose, and the associated citation. Covid-19 denotes coronavirus disease 2019.</p>	

# AI applications in disease surveillance

## Early warning

### 背景

At any given moment, the Web is **flooded** with disease reports in the form of news articles, press releases, professional discussion boards, and other curated fragments of information. ... However, the volume and distributed nature of these reports constitute much more information than can be made sense of promptly by even highly trained persons, making early warning of emerging viruses **nearly impossible**.

### 解决办法

AI-trained algorithms that can **parse, filter, classify, and aggregate text** for signals of infectious-disease events with high accuracy at unprecedented speeds.

### 举例

- **HealthMap**
  - **H1N1 in Mexico**
  - **vaping-induced pulmonary disease in US**

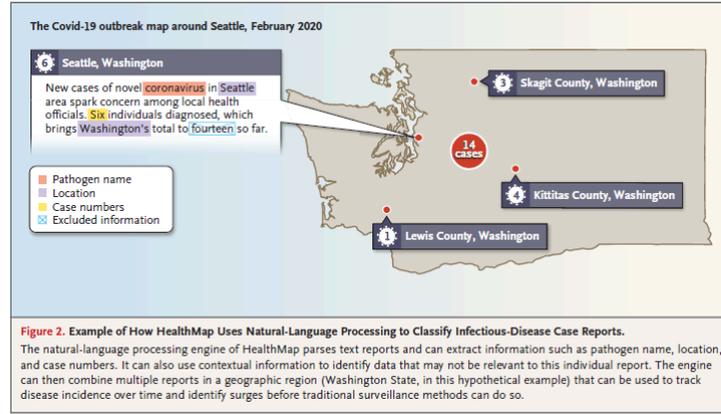
# 实现

natural-language processing to **search through text posted** across the Web for signals of infectious-disease events in real time by **comparing the text with a dictionary** of known pathogens and geographic areas.

ignore noise and parse relevant reports by **identifying disease-related text** such as the name of a pathogen and incidence numbers.

HealthMap also automatically extracts geographic information that can be used to tie multiple reports together and identify disease clusters that **cross-jurisdictional public health authorities may have missed**.

Continuously expanding dictionary with text in more than nine languages. ... the ability to simultaneously provide **worldwide coverage and hyperlocal situational awareness**.



# Pathogen classification

## 背景

AI has led to **advances in diagnostic classification** in a variety of fields, including neuroimaging (e.g., improving diagnostic tests for Alzheimer's disease) and oncology (e.g., detecting breast cancer).

## 实现

### 症状监测

By defining the **pathologic characteristics of an outbreak**, public health authorities are able to respond accordingly (e.g., by ensuring an adequate supply of oseltamivir when influenza cases are increasing in a region).

reliance on simple syndromic definitions can result in misidentification of an outbreak, particularly when pathogens share symptoms and routes of transmission. For example, a “Covidlike illness” syndrome suggested a false wave of **Covid-19 in Canada**, whereas pathogen data instead pointed to circulating seasonal viruses such as **enterovirus or rhinovirus**.

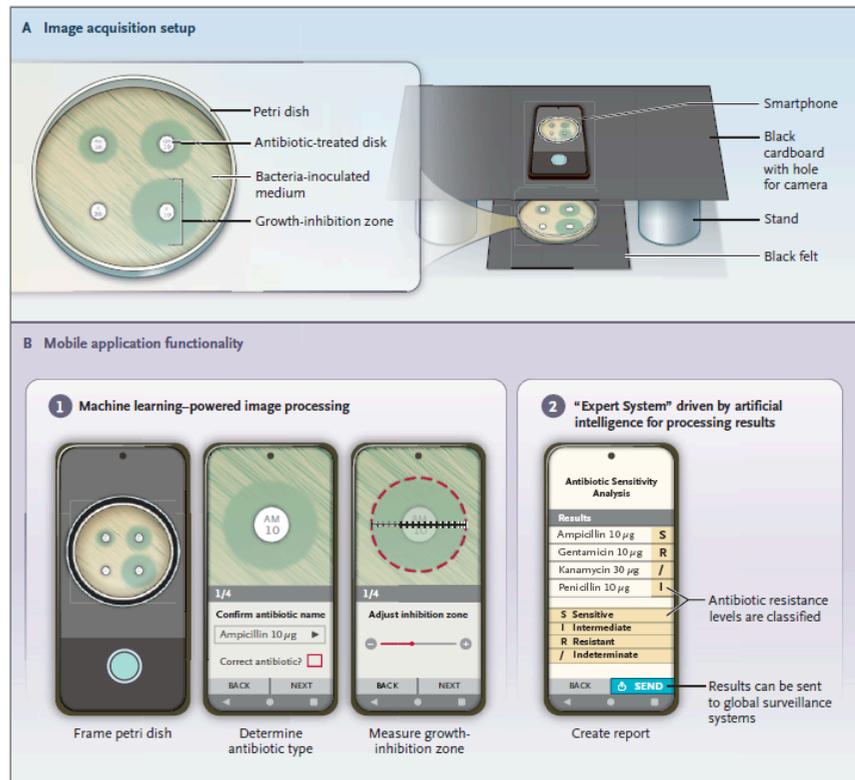
# 实现

## 耐药菌识别

Kirby-Bauer disk-diffusion test is a simple, low-cost technique for determining bacterial susceptibility to drugs from the diameter of the area in which growth of the bacteria is inhibited around an antibiotic-treated disk in a petri dish.

measurement quality is user-dependent and can result in misclassification of bacterial as susceptible or resistance.

- **图像识别**：量化抑菌圈直径
- **深度学习**：专家系统给出结果
- **自动化**：自动上传至WHO数据库



**Figure 3.** Example of Mobile Application to Measure Antibiotic Susceptibility with AI.

A mobile phone application developed by Pascucci and colleagues<sup>9</sup> uses machine learning and AI to classify bacterial susceptibility to various antibiotics. Panel A shows the image acquisition setup, and Panel B shows the mobile phone application. The application is designed to read a Kirby-Bauer disk-diffusion test, first by using machine-learning and image-processing techniques and then by organizing the results with the use of an AI-driven "expert system." The mobile application supports the ability to make high-quality reads in resource-limited settings and to forward the results to global antimicrobial resistance surveillance systems.

# Source identification

## 背景

When an outbreak has been identified, the next step is to stop the outbreak by first tracing and then cutting off routes of transmission. For hospital-based outbreak detection, tracking of infections with the use of **spatiotemporal clustering** and **contact tracing** can be performed by hand to identify targets for intervention.

**Genetic similarities of whole-genome surveillance sequences** can also be used to tie clinical cases together. However, this method cannot be used to identify sources of infection, and even when used in conjunction with traditional hospital-based outbreak detection, it may fail to identify complex transmission patterns, knowledge of which is required to direct interventions.

## 实现

### EDS-HAT: 结合基因组监测数据和机器学习

The algorithm was trained by means of a **case-control method** that parsed the EMR data from patients known to have infections from the same outbreak (cases) and EMR data from other patients in the hospital (controls used to establish baseline levels of exposure relatedness). ... real-time machine learning based on EMRs in combination with whole-genome sequencing could prevent up to 40% of hospital-borne infections in the nine locations studied and potentially save money.

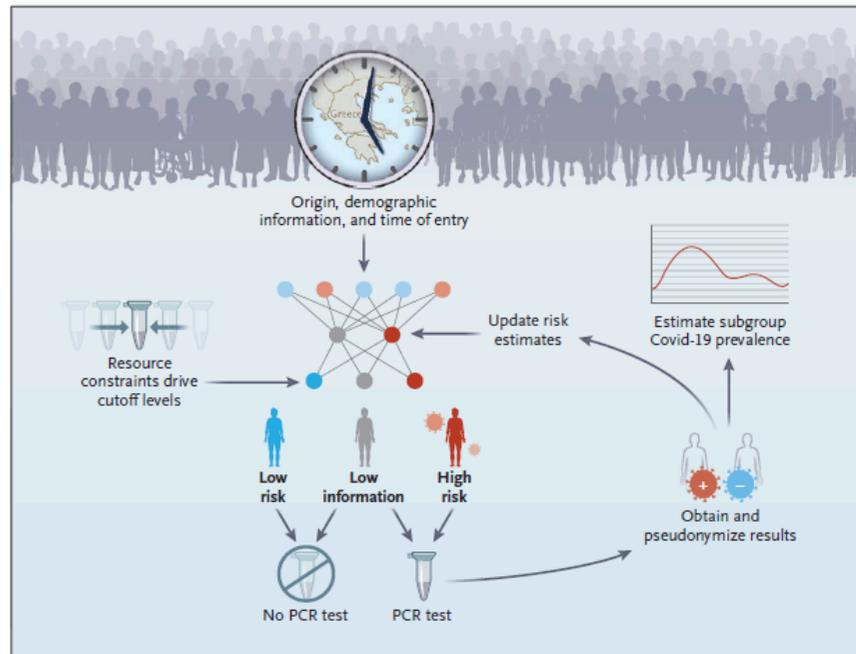
# Risk assessment

## 背景

For widespread infections such as those that occur in pandemics, complete elimination of infection at a single source is unlikely. In these scenarios, vaccination, contact tracing, and non-pharmaceutical interventions such as movement restrictions and mask wearing can be used to reduce transmission.

## 举例

- **QR codes** in China
- **Medical Chatbots**
- **Eva** in Greece



**Figure 4. Example of Reinforcement Learning for Covid-19 Border Surveillance.**

Eva is a reinforcement learning system used in Greece to allocate a limited supply of Covid-19 tests at the border of the country.<sup>11</sup> The algorithm uses information about the travelers in order to assign them to risk categories, with polymerase-chain-reaction (PCR) tests allocated accordingly. The risk estimate for each category is regularly updated to incorporate new information from the most recent batch of test results. Eva also sets testing cutoff levels, based on both risk and the available supply of tests, and makes Covid-19 prevalence estimates for each risk category. Pseudonymization refers to a deidentification procedure in which personally identifiable information is replaced by other identifiers.

# Extended applications

Individual event	Example of signal-generating method	Algorithm category	Signal of possible infectious disease in a population	Surveillance output
Biosignals passively measured by smartwatch	Gradient-boosting decision tree	Supervised classification	 <b>A Change in biosignals</b>	Early indication of possible outbreak
<b>Method advantages</b> <ul style="list-style-type: none"> <li>• Early warning can direct treatment and prevent spread</li> <li>• Continuously measured without requiring intervention</li> </ul>		<b>Method disadvantages</b> <ul style="list-style-type: none"> <li>• Disease signal is nonspecific</li> <li>• Requires deployment of device before outbreak</li> </ul>		
Cough detected by smart listening device	Regional proposal network	Artificial neural network	 <b>B Cough begins</b>	Spike in persons whose symptoms are detected early
<b>Method advantages</b> <ul style="list-style-type: none"> <li>• Passively monitor with already adopted devices</li> <li>• Can be used in homes or larger settings (e.g., waiting rooms)</li> </ul>		<b>Method disadvantages</b> <ul style="list-style-type: none"> <li>• Requires advanced privacy protection schemes</li> <li>• Symptomatic person (i.e., who coughed) may be unknown</li> </ul>		
Internet search query for viral testing site	Support vector regression	Supervised classification	 <b>C Search query for testing</b>	Hotspot of care-seeking behavior
<b>Method advantages</b> <ul style="list-style-type: none"> <li>• Can be inexpensive and centrally monitored</li> <li>• Captures behavior without requiring explicit participation</li> </ul>		<b>Method disadvantages</b> <ul style="list-style-type: none"> <li>• Testing possibly unrelated to symptom status (e.g., for travel)</li> <li>• Searches may not lead to testing (e.g., resource constraints)</li> </ul>		
Symptoms entered into website	Participatory surveillance	Human curated	 <b>D Enters symptoms online</b>	Real-time prevalence of possible cases
<b>Method advantages</b> <ul style="list-style-type: none"> <li>• Information can be disseminated without bureaucratic delay</li> <li>• Captures mild cases that may not formally test across settings</li> </ul>		<b>Method disadvantages</b> <ul style="list-style-type: none"> <li>• Participants skew toward persons with high health literacy</li> <li>• Relies on syndromic definitions that may describe many causes</li> </ul>		

Individual event	Example of signal-generating method	Algorithm category	Signal of possible infectious disease in a population	Surveillance output
Test result positive for virus	Traditional public health surveillance	Human curated	 <b>E Positive test result returned</b>	Official case counts
<b>Method advantages</b> <ul style="list-style-type: none"> <li>• Standard diagnostic accuracy</li> <li>• Mandatory reporting can capture rare and dangerous pathogens</li> </ul>		<b>Method disadvantages</b> <ul style="list-style-type: none"> <li>• Verification can be slow and expensive</li> <li>• Requires resources that may not be available in certain settings</li> </ul>		
Post on social media about diagnosis	Natural-language processing	Supervised classification	 <b>F Post diagnosis on social media</b>	Real-time prevalence of confirmed cases
<b>Method advantages</b> <ul style="list-style-type: none"> <li>• Rapid collection and dissemination of results</li> <li>• Wide array of users who may be missed by most other systems</li> </ul>		<b>Method disadvantages</b> <ul style="list-style-type: none"> <li>• Computationally expensive and difficult to parse signal from noise</li> <li>• Symptoms nonverified and can be vulnerable to Internet trolls</li> </ul>		
Mask wearing captured by CCTV	Convolutional neural network	Artificial neural network	 <b>G Mask wearing starts</b>	Nonpharmaceutical intervention levels
<b>Method advantages</b> <ul style="list-style-type: none"> <li>• Not vulnerable to desirability bias (i.e., captures true behavior)</li> <li>• High level of geographic specificity</li> </ul>		<b>Method disadvantages</b> <ul style="list-style-type: none"> <li>• Highly invasive and susceptible to privacy abuse</li> <li>• Resource intensive, especially outside urban locales</li> </ul>		

智能手表心率变化 -> 手机识别症状 -> 网络检索检测机构 -> 网络检索相关症状 -> 检测阳性率上升 -> 社交媒体相关诊断增加 -> 社交场所口罩佩戴人群增加

These AI-driven approaches complement the human-curated ones, including traditional public health surveillance, which is highly accurate but has a longer latency, and participatory surveillance, which can produce insights in real time but

# Surveillance roadblocks and future directions

## Data volume and quality

**"Big data hubris"** reminds us that even the most accurate AI-trained infectious-disease surveillance systems can lead to over-fitting (i.e., predictions that are **not generalizable** because they are too tailored to specific data) and should complement rather than replace high-quality traditional surveillance.

Disease-tracking systems that are not supplemented by **molecular testing** may not be able to disentangle co-circulating infections that have similar clinical manifestations.

AI algorithms designed for surveillance of diseases such as Covid-19 will require frequent recalibration as new pathogen variants emerge and exogenous variables.

machine-learning algorithms trained on low-quality data will not add value, and in some circumstances they may even be harmful.

# Data source representation

- 年龄？
- 性别？
- 种族？

# Privacy

- 用户数据上传至服务器
- 用户数据过度收集

# The limits of AI

AI can not replace the cross-jurisdictional and cross-functional coordination.

# Learn More



[PDF download](#)  
[GitHub Repository](#)  
[My Resume](#)