

阎文蓉, 宋 阳, 陈邦华, 等. 多源数据在流感预测中的应用研究[J]. 中华医学图书情报杂志, 2022, 31(9): 12-19.

DOI: 10.3969/j.issn.1671-3982.2022.09.002

· 研究与探讨 ·

多源数据在流感预测中的应用研究

阎文蓉¹, 宋 阳¹, 陈邦华², 马敬东¹

[摘要]目的: 梳理目前流感预测研究中多源数据的应用, 为流感等传染病监测研究提供启示和参考。方法: 系统检索 Web of Science 数据库、Scopus 数据库、PubMed 数据库、中国知网 (CNKI) 数据库、万方数据知识服务平台和维普中文期刊服务平台, 检索时间为 2009 年 1 月至 2022 年 4 月。纳入基于多源数据进行流感预测的相关研究文献, 并对符合纳入和排除标准的文献进行评述。结果: 共纳入文献 115 篇, 根据多源数据使用频率, 从高至低依次为互联网数据、环境数据、症状监测数据和组合数据, 基于多源数据的流感预测模型包括传染病模型、时空模型、机器学习模型和集成模型。结论: 基于传统监测系统和新型数据源的预测模型在提升流感预测预警能力上已取得成效, 但仍需从数据质量提升、预测模型优化、预警系统平台和工具改进、信息技术集成等方面进一步提高预测的准确性和及时性。

[关键词]流感预测; 预警; 监测; 机器学习

[中图分类号]R511.704; R-058 **[文献标志码]**A **[文章编号]**1671-3982(2022)09-0012-08

Application of multi-source data in influenza prediction

YAN Wen-rong¹, SONG Yang¹, CHEN Bang-hua², MA Jing-dong¹

(1.School of Medicine and Health Management, Tongji Medical College of Huazhong University of Science and Technology, Wuhan 430030, Hubei Province, China; 2.Wuhan Center For Disease Control and Prevention, Wuhan 430022, Hubei Province, China)

Corresponding author: MA Jing-dong

[Abstract] Objective The application studies of multi-source data in current influenza prediction were sorted out to provide insights and reference for influenza and other infectious disease surveillance studies. **Methods** Literature in Web of Science, Scopus, PubMed, Chinese National Knowledge Infrastructure (CNKI), Wanfang Data Knowledge Service Platform and VIP Database for Chinese Technical Periodicals was systematically searched from January 2009 to April 2022. Literature on studies related to influenza prediction using multi-source of data was included, and literature that met the inclusion and exclusion criteria was reviewed. **Results** A total of 115 papers were included and were internet data, environmental data, symptom surveillance data, and combined data according to the frequency of multi-source data use in descending order. Influenza prediction models based on

[基金项目]武汉市预防医学科研专项重大项目“突发公共卫生事件应急决策支持系统和应对处置能力建设研究”(WY22M03)

[作者单位]1.华中科技大学同济医学院医药卫生管理学院, 湖北武汉 430030; 2.武汉市疾病预防控制中心, 湖北武汉 430022

[作者简介]阎文蓉 (1998-), 女, 硕士研究生, 研究方向为卫生信息管理。

[通讯作者]马敬东 (1978-), 男, 博士, 教授, 研究方向为健康大数据分析。E-mail: jdma@hust.edu.cn

multi-source data included infectious disease models, spatio-temporal models, machine learning models, and integrated models. **Conclusion** Prediction models based on traditional surveillance systems and new data sources have been effective in improving influenza prediction and early warning capabilities, but further improvements in prediction accuracy and timeliness are still needed in terms of data quality improvement, prediction model optimization, early

warning system platform and tool improvement, and information technology integration.

[Key words]Influenza prediction; Early warning; Surveillance; Machine learning

流行性感冒简称“流感”，是由流感病毒引起的一种急性呼吸道疾病，属于丙类传染病。流感每年影响全球 5%~10% 的成年人和超过 20% 的儿童，导致数百万病例和至少 29 万人死亡^[1-2]。对流感进行及时可靠的预测和预警可以为政府卫生部门制定防控策略提供科学依据，降低流感发病率和死亡率。2019 年，世界卫生组织发布了《2019-2030 年全球流感战略》，强调开展建模方法的创新研究，使用新数据源，提高对流感发生、持续时间和风险等级的预测水平^[3]。传统流感监测依赖于系统上报的流感信息，上报周期较长，且高峰季节信息上报效率低，存在一定的诊断报告不足和漏报等情况，导致决策者不能准确有效地采取防控措施^[4]。为了解决这一问题，流感及流感样疾病（influenza like illness, ILI）预测研究使用了不同因素监测和预测流感的方法，互联网搜索数据、气候数据、非处方药销售数据等成为监测数据的重要组成部分^[5]。为了加速流感预测预警方法的研究进展，本文对基于多源数据的流感预测数据源指标体系和建模方法进行综述，为流感等传染病监测的研究提供启示和参考。

1 资料与方法

1.1 文献检索策略

按照《系统综述和荟萃分析优先报告的条目》（Preferred Reporting Items for Systematic Reviews and Meta-Analyses）（即 PRISMA 指南），分别在中英文数据库中检索，中文数据库包括中国知网（CNKI）数据库、万方数据知识服务平台和维普中文期刊服务平台，英文数据库包括 Web of

Science（WoS）数据库、Scopus 数据库、PubMed 数据库。由于距今最近的一次规模较大的流感大流行发生在 2009 年，因此将检索时间限定为 2009 年 1 月至 2022 年 4 月。中文数据库的检索式为“流感 AND（预测 OR 预警）”，来源类别选择“核心期刊、CSSCI 和 CSCD”。英文数据库的检索式为“（influenza OR Flu OR ILI）AND（predict OR forecast OR early warning）”。

1.2 纳入与排除标准

按照预先确定的纳入和排除标准对文献进行筛选。纳入标准：数据源采纳互联网数据、环境数据、症状监测数据等数据源中的一类或多类，研究内容为流感预测和预警，研究类型为定量研究。排除标准：数据源中仅采纳流感或流感样疾病的数据，研究内容为动物流感、流行病学调查与分析、病原学分析，文献类型为综述、会议摘要等。

1.3 数据提取与质量控制

对检索到的文献信息进行收集整理，内容包括第一作者、发表时间、数据源、建模方法、预测结果等。2 名研究者对文献信息进行筛选，当意见不统一时引入第 3 名研究者共同决定，通过审查文献的标题和摘要确定需要进行全文分析的文献，精读后由 2 名研究者通过讨论确定最终纳入研究的文献。

2 基于多源数据的流感预测研究现状

2.1 相关研究文献

文献检索步骤如图 1 所示，剔除重复文献后，通过审查标题和摘要筛选出 262 篇文献进行全文分析，最终纳入 115 篇文献，其中中文文献 18 篇，英文文献 97 篇。

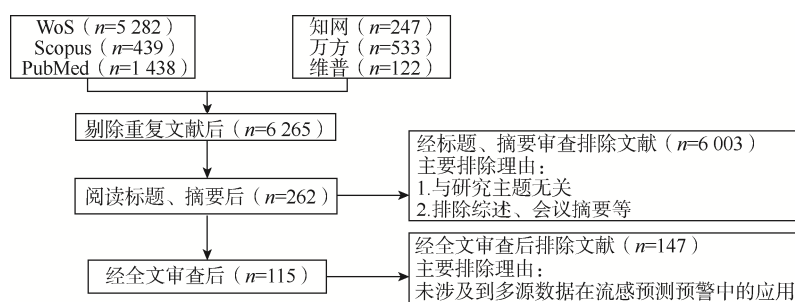


图 1 基于多源数据的流感预测文献检索流程

分析发现, 相关文献数量在 2017 年之后出现较快增长, 2018–2022 年相关文献数量为 67 篇 (占 58%)。对纳入文献进行归纳和总结后发现, 数据源主要包括互联网数据、环境数据、症状监测数据和组合数据等类型。其中涉及搜索引擎、社交网站、新闻网站等互联网数据的文献 54 篇

(占 47%); 涉及气候因素、空气质量、人口密度、假日效应等环境数据的文献 24 篇 (占 21%); 涉及工作/学校缺勤记录、药房/超市销售记录、实验室数据、通勤数据、物联网设备等症状监测数据的文献 20 篇 (占 17%); 涉及上述 3 种数据源的文献 17 篇 (占 15%) (图 2)。

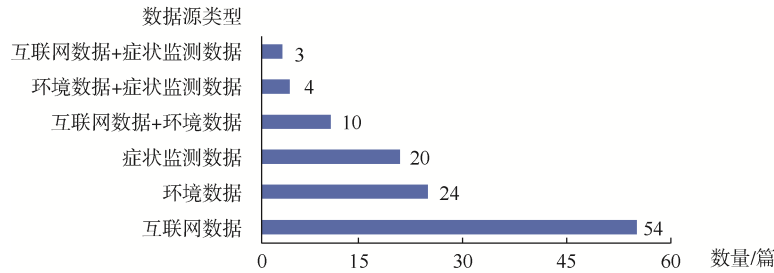


图 2 流感预测的多源数据分类

2.2 各类数据源的流感预测

数据的有效识别和纳入是预测流感的关键步骤。为了提高流感预测的及时性和准确性, 多源数据的应用已成为公共卫生监测体系中不可或缺的

一部分。多样化的数据有助于筛选各种与流感相关的预测因子, 从而提高预测模型的准确性。因此, 本文对多源数据的来源进行了整理并分析了各类数据源的优缺点 (表 1)。

表 1 多源数据的来源及其优缺点

数据源	数据来源	描述	优点	缺点
互联网数据	搜索引擎: 谷歌、百度、维基百科等	用户在互联网的搜索行为	实时性强, 低成本, 高时间和高空间分辨率	容易受到与流感无关的事件影响, 相对较低的可靠性和可用性, 存在媒体偏见 ^[6-9]
	社交网站: 推特、微博等	在社交网站发布的信息		
	新闻网站	来自官方和非官方的新闻文章		
	疾病预防控制中心网站	疾控中心网站发布的内容		
环境数据	空气质量	PM _{2.5} 、SO ₂ 、NO ₂ 、O ₃ 等	准确性高	混杂因素多, 受地区限制
	气候因素	湿度、温度、降水等		
	人口密度	官方统计的人口指标数据 ^[10]		
	假日效应	节假日前后, 时间序列走势会有明显的区别		
症状监测数据	急诊科或门诊的主诉	患者报告的疾病症状 ^[11]	可靠性高, 可以提供详细的疾病信息, 准确性高	获取难度大, 数据处理难度大, 数据结构复杂
	电子健康档案	患者就诊的数据		
	国家健康保险服务索赔数据库	包括所有人口的医疗和处方药索赔记录 ^[12]		
	药房/超市销售记录	药房或超市销售的特殊物品记录		
	学校/工作缺勤记录	从学校或工作场所收集的请假记录 ^[13]		
	通勤数据	跨地区的人员流动 ^[14]		
	实验室数据	实验室检查结果		
物联网设备	通过物联网的设备监测流感活动, 如智能温度计、移动 App、传感器等 ^[15-17]			
组合数据	—	互联网数据、环境数据、症状监测数据的组合应用	可以提高预测的准确性	数据源的选择和组合有局限性

2.2.1 基于互联网数据的流感预测

随着网络信息技术的飞速发展,社交网站和搜索引擎成为重要的信息集成平台,互联网中的用户生成信息(users generate content, UGC)和搜索查询数据是识别流行病开始的重要信号,可以用于预测包括流感在内的各种传染病^[14]。

搜索引擎的特点是将用户的搜索查询转化为全球各地流感活动的预报^[18],谷歌的“流感趋势”就是其中之一,但由于“流感趋势”严重高估了 2011-2013 年美国流感的发病率,失去了公众信任,其失败的原因可能在于“大数据傲慢”和算法的变化。因此,许多研究人员在“流感趋势”的基础上提出了改进策略^[18-19]。如有学者利用 ARGO Net 方法整合了与流感相关的谷歌搜索频率、电子健康记录及流感样疾病数据,通过数据挖掘工具 Google Correlate 将高度相关的搜索关键词输入到时间序列中,可以实时捕捉在线搜索行为与流感流行的动态关联^[20]。

社交网站上发布的内容可以作为评估人群疾病发病率的可靠且快速的来源。如有学者证明了推特数据与流感数据互补,推文中包含流感数据没有的潜在信息,而这些信息通常反映了流感流行前期 1~2 周的变化趋势^[21]。研究人员通过推文还可以获得具有高地理分辨率的在线实时健康相关数据,根据用户发布推文的地理位置、用户启用 GPS 跟踪的位置、推文中提及的地理位置等能够实时监测和建模不同地理位置的流感趋势^[22]。除了社交媒体外,有学者在香港卫生署卫生防护中心网站收集到的新闻文章中提取与流感高度相关的“H1N1”“H3N2”“Swine”“Flu”等关键词,通过统计每周包含每个关键词的新闻文章数量来生成时间序列数据,利用支持向量机(support vector machine, SVM)预测每周流感样病例的比例是否会增加或减少,平均准确率达到 86.7%^[23]。

互联网数据在一定程度上填补了传统监测手段预测流感发展趋势的空白,但易受媒体和舆论导向、群众关注度、虚假信息、个人上网习惯和地域因素的影响,无法及时准确反映流感流行动态,容易导致预测结果出现偏差。同时,目前的研究多是基于先验知识确定的搜索词进行分析,反映的是疫

情的历史特征,难以用来预测突发和新发疫情^[24]。

2.2.2 基于环境数据的流感预测

流感的流行和暴发受生物、季节、气候环境等因素影响^[25]。其中气候和空气质量是影响某一地区流感发病率的关键因素,通过机器学习和数据分析,挖掘气候和空气质量对流感发病的影响和作用是目前的研究热点。

流感高发于温带地区的冬季,较低的空气温度和湿度对流感的传播影响较大。有学者指出当温度每升高 1 °C,感染率就会降低 1.1%,相对湿度每降低 10%,感染率就会增加约 1.6%^[26]。同时,流感的暴发是由多种综合因素叠加导致的,可以利用患者暴露于不利环境和疾病发作之间的时间延迟,实现流感病例的早期发现^[27-28]。如有研究发现,平均气温、最高气温、最低气温、蒸汽压和相对湿度与流感活动的滞后时间分别为 2、2、1、1 和 0 周^[29]。考虑到气候的滞后效应,有学者在中国香港建立了基于气候的流感影响因素指标体系,包括大气压、绝对最高温度、平均温度、绝对最低温度、平均露点温度和季节性流感病毒的阳性检测次数等,对流感暴发趋势的预测准确率高达 96.47%^[5]。有学者通过互相关函数分析气候、空气污染对流感发病率的滞后相关性,发现 NO₂ 浓度与流感发病率呈正相关,PM_{2.5}、PM₁₀、SO₂、CO、O₃ 等与流感发病率呈负相关,其中 NO₂ 的多元差分自回归移动平均模型(autoregressive integrated moving average model with exogenous input variables, ARIMAX)预测性能最好^[30]。

气候、空气质量等指标对流感活动的影响提示,在流感预测实际应用中应多考虑重要的环境指标,而不仅是流感活动的历史数据。分析全球各地流感监测资料,发现影响流感季节性变化的主要影响因素包括动力共振、太阳辐射、露点温度变化、行为和干预效应、人群对优势流行毒株的免疫力等^[25]。也有研究证实通过假日效应、人口密度等因素可有效预测流感暴发^[31-32]。

2.2.3 基于症状监测数据的流感预测

症状监测是指通过临床症候群的发生频率数据,以及与疾病相关的数据,如学校或单位的缺勤记录、药房或超市的特定物品销售记录、物联网设

备数据等,发现疾病在时间和空间上的异常分布(聚集性),从而对疾病暴发进行早期预警和快速处置^[33-34]。

人群密集场所容易造成传染病传播,学校或单位的缺勤记录成为监测传染病疫情和突发公共卫生事件发生的重要数据源^[35]。研究人员监测了天津市 4 所县级学校 2012 年 9 月到 2014 年 12 月学生因病缺勤记录,通过累积控制图(cumulative sum control chart, CUSUM)检测异常信号,当 1 周内连续发出 2 次警报即确认流感暴发^[36]。此外,超市特定物品和药房非处方药的销售记录可以反映客户的行为变化,从而为预测季节性流感的发生提供有价值的数据库。有学者通过 Apriori 算法识别在超市购买的可以反映季节性流感传播的物品,如橙子、胡萝卜、鱼、肉类、卫生纸、矿泉水和洗洁精等,将上述物品纳入预测模型后可以比常规监测系统提前 4 周预测流感发病率^[37]。有学者发现儿童流感相关的非处方药销售量与滞后 4 天的儿童流感样病例数的相关性较好,可以利用该数据库实现对儿童流感疫情暴发的早期预警^[38]。

随着人工智能、物联网、云计算技术的进步,通过利用智能手机^[39]、移动 App^[15]和其他连接传感器^[40]的数据,有可能实现疾病从个体层面到群体层面的监测和预警。有学者将智能温度计收集的数据与流感样疾病监测数据进行比较,确定智能温度计数据和流感活动存在高度相关,准确反映了当前地区潜在的疾病传播^[15]。症状监测数据来源广泛,除上述非常规数据库外,医院诊断、实验室检查、疾控监测等也可提供及时、可靠的数据支撑,但在收集和分析数据过程中仍存在数据噪声、数据库资源整合不足、研究机构分散等问题亟待解决。

2.2.4 基于组合数据的流感预测

了解不同数据库的特点有助于选择最具预测性的数据库组合^[41]。有研究比较了不同数据库的预测性能,其中流感样疾病数据对预测的贡献最大,其次是气候因素,而基于互联网的百度指数和微博数据对模型的影响最小^[42];有关研究结果显示,气候因素、谷歌搜索数据、电子健康档案等数据库都有助于改善流感预测性能,但不同数据库的

组合会产生不同的性能,电子健康档案和谷歌搜索数据的组合数据库多用于实时和长期预测,预测效果最好^[43];有研究比较了传统的公共卫生监测数据、电子健康档案、维基百科搜索数据和社交网站发布的内容,发现即使公共卫生监测数据产生了最准确的预测,但组合数据库仍比单独数据库的预测性能更稳定和准确^[41]。因此,互联网数据、环境数据、症状监测数据等数据库的组合可以增强数据库间的关联,通过多数据库的互补进一步提高流感预测的准确性。

已有的研究证明了利用多数据库进行流感预测的应用前景,但多数据库的应用仍存在局限。一方面数据库的选择和组合取决于研究人员的意向,数据库收集成本、数据库来源的可靠性和数据库的可用性等因素限制了研究人员的选择范围;另一方面,数据库收集过程中通常包含大量与任务无关的信息,数据库处理和整合的效率也会影响多数据库在流感预测应用中的实践效果。同时,数据库标准不统一、信息冗余分散、隐私泄露等问题是实现多数据库实时共享与组合的关键挑战。

2.3 基于多数据库的流感预测模型

流感预测模型主要根据与流感相关的数据,运用适合的方法和模型对流感传播过程进行定性、定量分析和计算机模拟,分析流感暴发和流行的原因和关键因素,预测未来一定时期流感流行的规律和发展趋势^[44]。多数据库的应用为流感预测提供了新的思路和方法,本文通过梳理不同的数据库,总结了监测流感活动和预测流感发展趋势的不同算法。

2.3.1 传染病模型

在流行病学的背景下,利用“易感-感染(SI)”、“易感-感染-恢复(SIR)”、“易感-暴露-感染-恢复(SEIR)”和“易感-感染-易感(SIS)”等传染病模型^[45],根据疾病状态将人群划分为多个仓室并估计疾病在其中的传播速率,在分析疾病的发展过程、流行规律、变化趋势中发挥了重要作用。这些模型通常与其他方法结合使用以进行流感预测或通过构建附加状态和外源变量进行扩展^[31]。如基于湿度的“易感-感染-恢复-易感(SIRS)”流感模型,除考虑人群的易感性外,

将影响流感病毒生存与传播的绝对湿度也纳入到模型中以预测流感活动趋势^[46]。但此类模型在完全混合、同质的人群中无法捕捉到不同年龄组和环境差异^[47]。

2.3.2 时空模型

时空分析是从人群、病原、时间和地区等维度探测疾病的异常变化、分析时空聚集特征、确定高风险地区 and 传染病的时空分布。回顾性时空扫描统计 (retrospective spatiotemporal permutation scanning statistics) 基于时空聚类分析, 整合空间与时间特征, 通过不断调整窗口大小来识别疾病的高危区域和时间集群^[48-49]。有学者从时空角度分析互联网数据, 通过回顾性时空扫描统计发现了与流感相关的推文中高概率暴发流感的主要时空集群, 该集群包括大规模聚集和通勤的地点, 有助于进一步确定流感传播的重点区域^[50]。

2.3.3 机器学习模型

机器学习模型将流感预测看作分类回归或时间序列分析问题, 基于统计和数据驱动的差分自回归移动平均模型 (autoregressive integrated moving average, ARIMA)^[51]、长短期记忆神经网络 (long short-term memory, LSTM)^[52]、SVM^[53] 等模型经过实践证实要优于传统的统计模型并被广泛应用于流感预测。ARIMA 模型及其变体是时间序列预测模型中最为经典也是应用最为广泛的模型之一。有学者比较了 ARIMA 模型与包含外源温度计数据的扩展 ARIMA (ARIMAX) 模型对流感活动预测的准确性, 发现 ARIMAX 模型可以更有效地捕捉外部因素的影响^[54]。有学者通过机器学习回归方法中的分类回归树 (classification and regression tree, CART) 制定基于气象因素预警的阈值, 通过分支的走向可以基本确定判定方式, 提供了一种数据可视化的流感暴发预警阈值方法^[55]。

2.3.4 集成模型

集成学习以构建并组合多个模型为目标任务, 因为结合不同模型的输出可以获得比单个模型更好的性能。堆叠法是常用的方法之一。有学者选择循环神经网络 (recurrent neural network, RNN)、LSTM、门控循环单元 (gated recurrent unit, GRU)

等模型作为基学习器, 通过交叉验证得到下一阶段模型的输入数据进行训练, 可以有效防止模型的过拟合, 提高模型的性能^[27]。此外, 研究人员还根据不同模型的优势开发新的模型, 如自适应人工智能模型 (self-adaptive AI model, SAAIM)^[42]、自回归分布滞后模型和广义回归神经网络的集成模型 (autoregressive distributed lag-generalized regression neural network, ARDL-GRNN)^[56] 及多种机器学习方法组成的联合预测模型^[57] 等, 在处理结构化和非结构化的多源数据中实现了更好的性能。但集成学习建立和训练的模型也存在不足, 成本更高, 并且不能帮助研究人员更好了解样本和总体之间的未知差异^[58]。

3 总结和展望

在全球疫情的影响和冲击下, 如何快速预防和应对流感仍是一项严峻的挑战。本文通过总结多源数据的类型及其建模方法, 梳理国内外基于多源数据的流感预测研究现状, 发现互联网数据、环境数据、症状监测数据是流感预测研究中除传统流感监测数据外的关键数据源, 机器学习模型、传染病模型、时空模型和集成模型是流感预测的常用方法和工具。多源数据扩大了传统流感监测数据的范围和深度, 同时互联网数据、环境数据、症状监测数据等数据源的组合和集成模型正逐步应用于流感预测研究, 为公共卫生机构和研究人员提供了实时监测全球传染病动态、发现新发和突发传染病疫情的研究方向。

目前, 基于多源数据的流感预测模型在准确性和灵敏性上有各自的优势和不足。如何在信息技术的支持下, 基于多源数据集成不同预测模型提高流感预测预警的准确性和及时性是下一步研究的重点。具体包括以下几方面。一是数据质量的提升, 建立数据来源更丰富、指标体系更完善的流感预测因子数据集, 规避数据噪声、数据冗余及隐私泄露等风险, 提高数据的质量和可用性; 二是预测模型的优化, 通过统计方法集成多模型的优势是当前流感预测研究的趋势, 同时应建立统一的评估指标体系衡量预测结果在实践中的意义; 三是预警系统平台和工具的改进, 搭建公共卫生机构、医院、气候、环境等多部门、多机构协同沟通的综合性数据平

台, 建立更加准确的流行态势感知和风险预警工具, 推动流感在时空上分布和变化的可视化分析; 四是云计算、大数据、物联网等技术的集成, 为多源数据的采集、传输、存储和使用提供了强大的技术支持, 实时、在线的数据可以实现对人群疾病的客观测量, 信息技术的集成将推进以流感为主的常见传染病的预防控制与干预体系的完善。

本文也存在不足之处, 如未对建模方法进行详细描述, 不同预测模型有各自的优缺点和适用范围, 随着数据源类型日益丰富, 今后将进一步深入研究适用多源数据的建模方法。

【参考文献】

- [1] Iuliano A D, Roguski K M, Chang H H, *et al.* Estimates of global seasonal influenza-associated respiratory mortality: a modelling study[J]. *The Lancet*, 2018, 391(10127): 1285-1300.
- [2] World Health Organization. Influenza(Seasonal) [EB/OL]. [2022-05-10]. [https://www.who.int/news-room/fact-sheets/detail/influenza-\(seasonal\)](https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal)).
- [3] World Health Organization. Global influenza strategy 2019-2030[M/OL]. [2022-05-10]. <https://apps.who.int/iris/handle/10665/311184>.
- [4] 余 昭, 孙琬琬, 刘社兰, 等. 呼吸道传染病监测体系运行状况分析[J]. *预防医学*, 2021, 33(1): 101-103.
- [5] Wang Z X, Ntambara J, Lu Y, *et al.* Construction of influenza early warning model based on combinatorial judgment classifier: a case study of seasonal influenza in Hong Kong[J]. *Current Medical Science*, 2022, 42(1): 226-236.
- [6] Zhang Y, Bambrick H, Mengersen K, *et al.* Using Google Trends and ambient temperature to predict seasonal influenza outbreaks[J]. *Environment International*, 2018, 117: 284-291.
- [7] Samaras L, García-Barriocanal E, Sicilia M-A. Comparing social media and google to detect and predict severe epidemics[J]. *Scientific Reports*, 2020, 10(1): 4747.
- [8] Zhang Y Z, Yakob L, Bonsall M B, *et al.* Predicting seasonal influenza epidemics using cross-hemisphere influenza surveillance data and local Internet query data[J]. *Scientific Reports*, 2019, 9(1): 3262.
- [9] Lu F S, Hou S Q, Baltrusaitis K, *et al.* Accurate influenza monitoring and forecasting using novel internet data streams: a case study in the Boston metropolis[J]. *JMIR Public Health and Surveillance*, 2018, 4(1): e4.
- [10] Wei Y C, Ou Y L, Li J Q, *et al.* Forecasting the potential number of influenza-like illness cases by fusing internet public opinion[J]. *Sustainability*, 2022, 14(5): 2803.
- [11] Paul R, Han D, DeDoncker E, *et al.* Dynamic downscaling and daily nowcasting from influenza surveillance data[J]. *Statistics in Medicine*, 2022, 41(21): 4159-4175.
- [12] Son W S, Park J E, Kwon O. Early detection of influenza outbreak using time derivative of incidence[J]. *EPJ Data Science*, 2020, 9(1): 28.
- [13] Quandelacy T M, Zimmer S, Lessler J, *et al.* Predicting virologically confirmed influenza using school absences in Allegheny County, Pennsylvania, USA during the 2007-2015 influenza seasons[J]. *Influenza and Other Respiratory Viruses*, 2021, 15(6): 757-766.
- [14] Murayama T, Shimizu N, Fujita S, *et al.* Predicting regional influenza epidemics with uncertainty estimation using commuting data in Japan[J]. *PLoS One*, 2021, 16(4): e0250417.
- [15] Ackley S F, Pilewski S, Petrovic V S, *et al.* Assessing the utility of a smart thermometer and mobile application as a surveillance tool for influenza and influenza-like illness[J]. *Health Informatics Journal*, 2020, 26(3): 2148-2158.
- [16] Radin J M, Wineinger N E, Topol E J, *et al.* Harnessing wearable device data to improve state-level real-time surveillance of influenza-like illness in the USA: a population-based study[J]. *Lancet Digital Health*, 2020, 2(2): e85-e93.
- [17] Miller A C, Peterson R A, Singh I, *et al.* Improving state-level influenza surveillance by incorporating real-time smartphone-connected thermometer readings across different geographic domains[J]. *Open Forum Infectious Diseases*, 2019, 6(11): ofz455.
- [18] Lamos V, Miller A C, Crossan S, *et al.* Advances in nowcasting influenza-like illness rates using search query logs[J]. *Scientific Reports*, 2015, 5: 12760.
- [19] Lazer D, Kennedy R, King G, *et al.* The parable of Google Flu: traps in big data analysis[J]. *Science*, 2014, 343(6176): 1203-1205.
- [20] Lu F S, Hattab M W, Clemente C L, *et al.* Improved state-level influenza nowcasting in the united states leveraging internet-based data and network approaches[J]. *Nature Communications*, 2019, 10(1): 147.
- [21] Xue H X, Bai Y, Hu H P, *et al.* Regional level influenza study based on Twitter and machine learning method[J]. *PLoS One*, 2019, 14(4): e0215600.
- [22] Wang F, Wang H Y, Xu K, *et al.* Regional level influenza study with geo-tagged twitter data[J]. *Journal of Medical Systems*, 2016, 40(8): 189.
- [23] Kim J, Ahn I. Weekly ILI patient ratio change prediction using news articles with support vector machine[J]. *BMC Bioinformatics*, 2019, 20(1): 259.
- [24] 李秀婷, 刘 凡, 董纪昌, 等. 基于互联网搜索数据的中国流感监测[J]. *系统工程理论与实践*, 2013, 33(12): 3028-3034.
- [25] Zhou Steven Zhixiang. 季节性流感的多元激发理论及其包含气象和社会行为因素的数学模型[J]. *热带气象学报*, 2009, 25(6): 706-716.
- [26] Singh D E, Marinescu M-C, Carretero J, *et al.* Evaluating the impact of the weather conditions on the influenza propagation. [J]. *BMC Infectious Diseases*, 2020, 20(1): 265.
- [27] Kristiani E, Chen Y-A, Yang C-T, *et al.* Using deep ensemble for influenza-like illness consultation rate prediction[J]. *Future Generation Computer Systems*, 2021, 117: 369-386.
- [28] 查文婷, 李渭通, 何嘉慧, 等. 湖南省流行性感冒与气象因素关系及预测[J]. *中国公共卫生*, 2021, 37(3): 537-541.
- [29] Bai Y L, Huang D S, Liu J, *et al.* Effect of meteorological factors on influenza-like illness from 2012 to 2015 in Huludao, a northeastern city in China[J]. *Peer J*, 2019, 7: e6919.

- [30] Zheng Y, Wang K, Zhang L, *et al.* Study on the relationship between the incidence of influenza and climate indicators and the prediction of influenza incidence[J]. *Environmental Science and Pollution Research*, 2021, 28(1): 473-481.
- [31] Brooks L C, Farrow D C, Hyun S, *et al.* Flexible modeling of epidemics with an empirical Bayes framework[J]. *PLoS Computational Biology*, 2015, 11(8): e1004382.
- [32] Pei S, Kandula S, Yang W, *et al.* Forecasting the spatial transmission of influenza in the United States[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2018, 115(11): 2752-2757.
- [33] Chen H, Zeng D, Yan P. Syndromic surveillance data sources and collection strategies[M] // *Integrated Series in Information Systems*. New York, NY: Springer US, 2010: 33-48.
- [34] Buehler J W, Hopkins R S, Overhage J M, *et al.* Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC Working Group[J]. *MMWR Recommendations and Reports*, 2004, 53(RR-5): 1-11.
- [35] Fatiregun A A, Sangowawa A O, Abubakar O A. Outcome of training on yellow fever surveillance in a South-Western State, Nigeria: implications for improved field implementation[J]. *Nigerian Journal of Clinical Practice*, 2010, 13(4): 449-452.
- [36] Xu W T, Chen T M, Dong X C, *et al.* Outbreak detection and evaluation of a school-based influenza-like-illness syndromic surveillance in Tianjin, China[J]. *PLoS One*, 2017, 12(9): e0184527.
- [37] Miliou I, Xiong X Y, Rinzivillo S, *et al.* Predicting seasonal influenza using supermarket retail records[J]. *PLoS Computational Biology*, 2021, 17(7): e1009087.
- [38] 田丽红, 谭莉, 范允舟, 等. 综合症状监测系统 in 儿童流行性感 冒疫情监测中的应用[J]. *中华预防医学杂志*, 2013, 47(12): 1095-1099.
- [39] Venkatramanan S, Sadilek A, Fadikar A, *et al.* Forecasting influenza activity using machine-learned mobility map[J]. *Nature Communications*, 2021, 12(1): 726.
- [40] Grzesiak E, Bent B, McClain M T, *et al.* Assessment of the feasibility of using noninvasive wearable biometric monitoring sensors to detect influenza and the common cold before symptom onset[J]. *JAMA Network Open*, 2021, 4(9): e2128534.
- [41] Ertem Z, Raymond D, Meyers L A. Optimal multi-source forecasting of seasonal influenza[J]. *PLoS Computational Biology*, 2018, 14(9): e1006236.
- [42] Su K, Xu L, Li G Q, *et al.* Forecasting influenza activity using self-adaptive AI model and multi-source data in Chongqing, China[J]. *EBioMedicine*, 2019, 47: 284-292.
- [43] Poirier C, Hswen Y, Bouzillé G, *et al.* Influenza forecasting for French regions combining EHR, web and climatic data sources with a machine learning ensemble approach[J]. *PLoS One*, 2021, 16(5): e0250890.
- [44] 李琼芬, 黄甜, 王荣华, 等. 传染病疫情预测预警模型研究进展[J]. *中国公共卫生*, 2013, 29(11): 1695-1697.
- [45] Younsi F Z, Bounekar A, Hamdadou D, *et al.* Integration of multiple regression model in an epidemiological decision support system[J]. *International Journal of Information Technology & Decision Making*, 2019, 18(6): 1755-1783.
- [46] Zimmer C, Leuba S I, Yaesoubi R, *et al.* Use of daily Internet search query data improves real-time projections of influenza epidemics[J]. *Journal of the Royal Society Interface*, 2018, 15(147): 20180220.
- [47] Nsoesie E O, Brownstein J S, Ramakrishnan N, *et al.* A systematic review of studies on forecasting the dynamics of influenza outbreaks[J]. *Influenza and Other Respiratory Viruses*, 2014, 8(3): 309-316.
- [48] He D H, Lui R, Wang L, *et al.* Global spatio-temporal patterns of influenza in the post-pandemic era[J]. *Scientific Reports*, 2015, 5: 11013.
- [49] Dong W, Yang K, Xu Q L, *et al.* A predictive risk model for a(H7N9) human infections based on spatial-temporal autocorrelation and risk factors: China, 2013-2014[J]. *International Journal of Environmental Research and Public Health*, 2015, 12(12): 15204-15221.
- [50] Nagar R, Yuan Q Y, Freifeld C C, *et al.* A case study of the New York City 2012-2013 influenza season with daily geocoded Twitter data from temporal and spatiotemporal perspectives[J]. *Journal of Medical Internet Research*, 2014, 16(10): e236.
- [51] Monamele G C, Vernet M-A, Nsaibirni R F J, *et al.* Associations between meteorological parameters and influenza activity in a subtropical country: case of five sentinel sites in Yaoundé-Cameroon[J]. *PLoS One*, 2017, 12(10): e0186914.
- [52] Venna S R, Tavanaei A, Gottumukkala R N, *et al.* A novel data-driven model for real-time influenza forecasting[J]. *IEEE Access*, 2019, 7: 7691-7701.
- [53] Santos J C, Matos S. Analysing Twitter and web queries for flu trend prediction. [J]. *Theoretical Biology & Medical Modelling*, 2014, 11(Suppl 1): S6.
- [54] Miller A C, Singh I, Koehler E, *et al.* A smartphone-driven thermometer application for real-time population- and individual-level influenza surveillance[J]. *Clinical Infectious Diseases*, 2018, 67(3): 388-397.
- [55] Ge H, Fan D B, Wan M, *et al.* How to determine the early warning threshold value of meteorological factors on influenza through big data analysis and machine learning[J]. *Computational and Mathematical Methods in Medicine*, 2020: 8845459.
- [56] He F, Hu Z J, Zhang W C, *et al.* Construction and evaluation of two computational models for predicting the incidence of influenza in Nagasaki prefecture, Japan[J]. *Scientific Reports*, 2017, 7(1): 7192.
- [57] 柴国荣, 王斌, 沙勇忠. 基于多机器学习方法联合的公共卫生风险预测研究: 以兰州市流感预测为例[J]. *数据分析与知识发现*, 2021, 5(1): 90-98.
- [58] He Y, Zhao Y, Chen Y, *et al.* Nowcasting influenza-like illness (ILI) via a deep learning approach using google search data: an empirical study on Taiwan ILI[J]. *International Journal of Intelligent Systems*, 2022, 37(3): 2648-2674.

[收稿日期: 2022-08-10]

[本文编辑: 黄思敏]