·专题报道——人工智能在公共卫生中的应用·

# 大语言模型在广州市法定传染病趋势预测中的应用：基于 2024 年监测资料

吴宇骞，董淑文，向欣荣，李宗欣，甘建哲，秦鹏哲

广州市疾病预防控制中心（广州市卫生监督所），广州 510440
通信作者：秦鹏哲，E-mail：396508@qq.com

**【摘　要】目的** 比较多种大语言模型与人工判断在传染病趋势预测中的准确率。**方法** 利用广州市 2024 年 1—12 月 17 种传染病的月报数据，用 5 种大语言模型（LLM）及公卫医师专家判断进行预测，并与实际趋势比较。**结果** 基于共 17 种传染病进行 99 次预测，其中公卫医师预测准确率为 84.85%，AI 集成预测准确率为 73.74%，5 种 AI 预测准确率依次为 Kimi（70.71%）、豆包（68.69%）、灵犀（68.69%）、文心一言（65.66%）、Deep Seek（63.64%）。基于 99 条参考的数据资料，在与登革热、诺如病毒感染、手足口病、百日咳、流行性感冒相关的在疫情拐点识别中，公卫医师的准确率达 50.0%，其次是 Kimi（30.0%）、AI 集成（20.0%）、豆包（20.0%）；公卫医师的 F1 分数为 0.56，其次是 Kimi（0.40）、豆包（0.38）、AI 集成（0.27）。**结论** 当前公卫医师在传染病趋势预测中准确性高于 AI，尤其在识别复杂社会因素影响方面表现更优。
**【关键词】** 人工智能；公共卫生；传染病；预测；广州市

# Application of large language models in predicting trends of notifiable infectious diseases in Guangzhou: based on 2024 surveillance data

WU Yuqian, DONG Shuwen, XIANG Xinrong, LI Zongxin, GAN Jianzhe, QIN Pengzhe (*Guangzhou Center for Disease Control and Prevention (Guangzhou Health Supervision Institute)*, *Guangzhou* 510440, *China*)
Corresponding author: QIN Pengzhe, E-mail: 396508@qq.com

**【Abstract】 Objective** To compare the accuracy of multiple large language models (LLM) and manual judgment in predicting infectious disease trends. **Methods** Using monthly report data of 17 infectious diseases in Guangzhou from January to December 2024, we used five LLM and expert judgments from public health physicians for prediction and compared the predicted results with actual trends. **Results** A total of 99 predictions were conducted for 17 infectious diseases. The accuracy rate of prediction by public health physicians was 84.85%, and that of the AI-integrated prediction was 73.74%. The prediction accuracy of AI products followed the trend of Kimi (70.71%), Doubao (68.69%), Lingxi (68.69%, parallel), ERNIE Bot (65.66%), and Deep Seek (63.64%). On the basis of 99 referred data items related to dengue fever, norovirus infection, hand-foot-mouth disease, pertussis, and influenza, in the identification of epidemic inflection points, public health physicians showed the accuracy of 50.0%, followed by Kimi (30.0%), AI integration (20.0%), and Doubao (20.0%, parallel); public health physicians showed the F1 score of 0.56, followed by Kimi (0.40), Doubao (0.38), and AI integration (0.27). **Conclusions** Currently, public health physicians have higher accuracy in predicting infectious disease trends than AI, especially in identifying the impacts of complex social factors.
**【Keywords】** artificial intelligence; public health; infectious disease; prediction; Guangzhou city

传染病是增加全球疾病负担和降低人均预期寿命的重要因素[1]，其暴发和流行常带来较大的健康与社会经济影响[2]。我国已建立全球最大的传染病直报系统，依托哨点监测实现快速响应，为传染病防控和公共卫生决策提供了有力支撑[3]。传染病趋势预测对制定防控策略、优化资源配置具有重要意义[4]。随着人工智能（artificial intelligence, AI）技术快速发展，国家卫生健康委员会、国家疾病预防控制局于 2024 年联合印发《卫生健康行业人工智能应用场景参考指引》，积极推进“人工智能+”在疾控领域的应用。AI 在整合多源数据[5]、构建预测模型[6]、识别高风险区域[7]等方面展现出优势。自 2023 年以来，大语言模型（large language model, LLM）在医学领域展现了巨

开放获取：CC BY-NC-ND 4.0　DOI：10.11847/zgggws1148796
基金项目：广州市卫生健康科技青年人才培育项目（20261A031045）
第一作者：吴宇骞（1992—），博士，研究方向：传染病病原体系统发育、数字化在传染病监测中的应用。
收稿日期：2025-09-02　修回日期：2025-12-19　录用日期：2026-03-09
利益冲突：不存在　伦理审查：免除审查　出版授权：全体作者已与编辑部签署作者贡献声明及版权转让协议

大潜力，但其在流行病趋势预测的应用仍显不足[8–9]。已有文献指出，LLM 在面对公共卫生挑战，尤其是流感等季节性疾病的流行预测方面的效果仍待探究[10]。本研究基于 2024 年广州市每月传染病月报数据，选取 WPS 灵犀、文心一言、Deep Seek、豆包和 Kimi 五种 AI 工具，对 2024 年广州市 17 种传染病趋势进行预测，重点识别百日咳、登革热等 5 种高发疾病疫情拐点，并与公卫医师的判断结果进行比较，以评估 LLM 在传染病预测中的实际效果，为推进 AI 辅助预警提供依据。

## 1 资料与方法

1.1 资料来源与预测内容 本研究采用 2024 年 1 月—2025 年 2 月收录于广州市疾病预防控制中心疫情汇编《每月疾病监测动态》中的法定传染病监测数据。数据来源包括中国疾病预防控制信息系统(“大疫情网”)、广州海关、流感哨点医院、市场采样数据等，包括新型冠状病毒感染、流行性感冒、百日咳、流行性脑膜炎、诺如病毒感染、肝吸虫病、猩红热、手足口病、出血热、布鲁氏菌病、流行性腮腺炎、登革热、麻疹、恙虫病、猴痘、水痘、人感动物源性流感等 17 种国家法定传染病。为聚集分析，本研究根据广州市的地理气候特征及 2024 年华南地区疫情关注重点，选取关注月份较多的百日咳(关注 12 个月)、登革热(关注 11 个月)、流行性感冒(关注 10 个月)、诺如病毒感染(关注 7 个月)、手足口病(关注 5 个月)共 5 种传染病作为重点分析疾病，共计对 84 条月度数据进行预测。核心研究内容为：评估不同方法对上述疾病次月发病趋势(上升、平稳、下降)的预测准确性，并探究其对疫情拐点(趋势反转点)的识别能力。分析所依据的主要数据指标包括：各传染病种当月确诊病例数、累计确诊病例数、与 2023 年同期对比确诊病例数据等。流行性感冒指标还包括“流行株系”“哨点医院门诊情况”等专项指标。其余 12 种传染病因涉及关注月份较少或本地报告病例数长期维持较低水平，进行趋势与拐点分析的意义有限，故未纳入重点分析范畴。本研究已通过单位内部审查，所有数据的使用与处理过程均严格遵守国家相关法律法规，符合相关隐私保护与科研伦理规范的要求。人工与 AI 预测技术路线图见图 1。

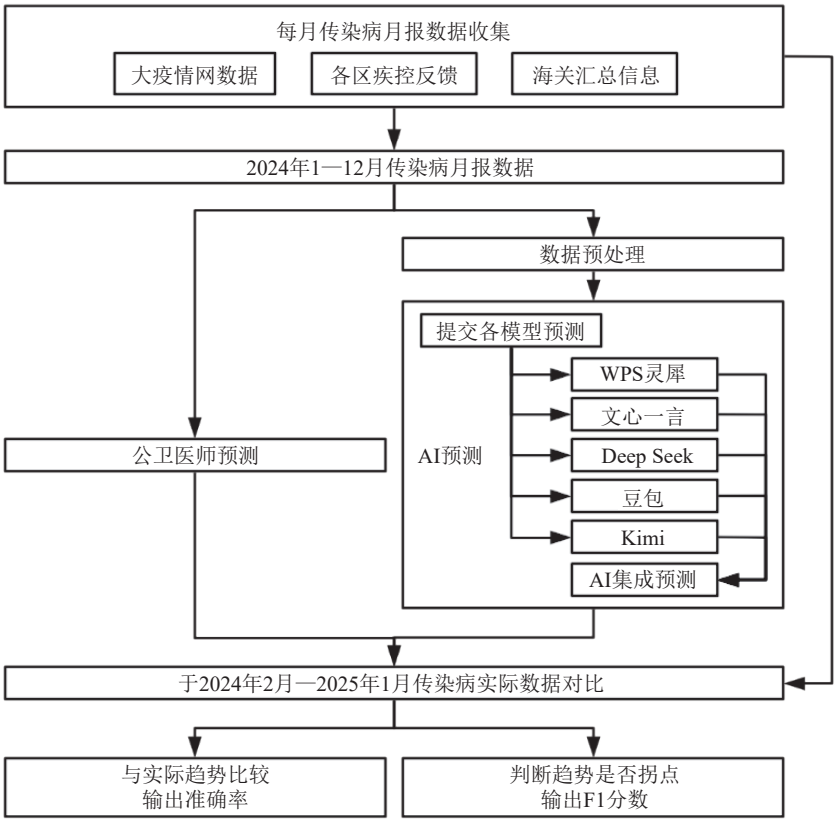



**图 1** 本研究涉及的人工预测及 AI 预测的技术路线图

**Fig. 1** Flowchart for manual prediction and AI prediction involved in this study

### 1.2 方法

1.2.1 人工预测　基于个人经验的层级式决策对次月传染病病例数趋势进行预测评估。首先由病种管理人依托其专业知识和既往经验对监测数据进行分析，形成初步预测结论；随后由具备管理职责的负责人(均为正高职称专家)结合技术合理性、资源可行性和政策适用性进行质控和审核。同时，使用专家会商法对传染病在次月的趋势进行研判调整。专家会商法基于广州市疾控中心半年度、季度、月度的突发公共卫生事件风险评估会议，参会人员为病种负责人和其他公卫领域的专家(会议人数根据专家工作安排略有变动，一般≥17 人，职称均为副主任医师及以上)。专家基于已有的监测数据、模型初步结果、现场调查信息等，通过正式会议对传染病次月趋势进行综合分析和讨论，判断指定传染病在未来数月的可能流行趋势。该预测方式按流程一般需要 1～2 d。

1.2.2 AI 模型预测　(1)AI 模型选用：鉴于不同大语言模型(LLM)在各自领域的独特优势，本研究选用了以下几款在各自领域中表现出色的模型：豆包(版本号：7.9.0)、Deep Seek(版本 Deep Seek-R1，使用其“深度思考”功能)、Kimi(版本号：v1.0)、文心一言(版本号 3.5)、灵犀(版本号：19770)。其中，豆包擅长内容创作、社区互动和多模态领域；Deep Seek 以开源共享、轻量化部署、低算力与低成本为主要特色；Kimi 支持超 200 万字的长文本处理(可达 200 万汉字)；文心一言具备强大的知识增强、检索增强与对话增强优势；WPS 灵犀聚焦智能办公场景。(2)AI 预测方法：基于预实验和参数调整，预测过程遵循以下指令：“假设您为华南地区公共卫生医师，专门负责传染病月度监测工作。请根据本市当前月的疾病状况，根据您的经验，预测次月本市的疫情趋势、主要防控场所及目标人群，并简要阐述病例数的预期变化(上升、平稳或下降)”，同时附上经过脱密化预处理的传染病数据，即为确保信息的保密性，所有能够明确指向具体地理位置的研究数据均被剔除，仅以“华南地区”作为地理背景的概括性描述，旨在保护个人隐私的同时，保留某些传染病(如登革热)与地理环境的关联性。此外，指令中“根据您的经验”旨在引导人工智能系统调用与特定传染病相关的背景知识库，包括气候季节性因素和历史疫情数据等。为简化传染病疫情趋势的量化分析，本研究要求预测模型明确指出预期病例数的变化趋势(上升、平稳或下降)，以便与实际状况进行对比分析。

1.3 评估指标　(1)根据各模型预测结果与实际趋势的一致性，将预测判定为“正确”或“错误”。若 AI 模型输出的结果为模糊性描述(如“难以确切判断”或“可能下降，但若未有效控制则仍可能上升”)，则归为“无效”。为减少单一 AI 模型偏差，本研究进一步计算了“AI 集成预测”指标，即当各模型对次月疫情趋势的预测结果存在分歧时，采用多数投票法将得票数超过 50% 的预测结果判定最终预测趋势[11]，并将该指标与“公卫医师”代表的人工预测结果进行比较。“预测准确率”定义为预测的疫情趋势与实际趋势完全一致的月份数占总预测月份的比例，即被认定为“正确”的预测数量占全部预测数量的比例。(2)“拐点”指当月疫情趋势与上月趋势相反的时点，即由上升转为下降或由下降转为上升[12]。“拐点处趋势准确率”是指在拐点处模型预测的趋势与实际趋势一致的比例。F1 分数被用于衡量预测模型判断拐点的能力，计算公式如下：

$$\frac{2\times\frac{TP}{TP+FP}\times\frac{TP}{TP+FN}}{\frac{TP}{TP+FP}+\frac{TP}{TP+FN}} \quad (1)$$

分数越高代表模型预测拐点的能力越强。其中，预测模型判定当月趋势与实际相符，且当月趋势与上月相反时，判定为“真阳性”(TP)；当月未出现拐点，模型预测为拐点，判定为“假阳性”(FP)；当月为拐点，模型未预测出相同趋势，判定为“假阴性”(FN)；当月未出现拐点，预测也非拐点，则为“真阴性”。根据 TP、FP、FN 预测月数算出 F1 分数[13]。考虑到部分传染病统计月份较少，或长期维持在低水平波动，本研究仅判断百日咳、登革热、流行性感冒、诺如病毒感染、手足口病等 5 种传染病的拐点。

## 2 结　果

2.1 不同传染病趋势预测比较(表 1)　共预测 99 条信息，公卫医师预测准确率为 84.8%，高于 AI 集成(73.7%)。在 17 种法定传染病中，AI 模型按准确率从高到低排序为：Kimi(70.7%)、豆包(68.7%)、灵犀(68.7%)、文心一言(65.7%)、Deep Seek(63.6%)。在 5 种重点关注的传染病中，总体看 LLM 模型对登革热趋势预测的准确率最高，4 种模型预测准确率均>80.0%，其中 AI 集成、灵犀准确率均为 81.8%，Deep Seek 与豆包准确率均为 90.0%；对手足口病的预测准确率次之，AI 集成、Deep Seek 和 Kimi 准确率均为 80.0%。值得注意的是，流行性感冒趋势 AI 集成预测的准确率仅为 40.0%，豆包模型准确率更低，为 20.0%。

**表 1** 广州市法定传染病模型预测结果比较(准确率/%)
**Table 1** Comparison of predictions for statutory infectious diseases in Guangzhou

| 病种 | 预测数据总数 | 公卫医师 | | AI集成 | | Kimi | | 灵犀 | | 豆包 | | 文心一言 | | Deep Seek | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 实发数 | % | 实发数 | % | 实发数 | % | 实发数 | % | 实发数 | % | 实发数 | % | 实发数 | % |
| 手足口病 | 5 | 4 | 80.0 | 4 | 80.0 | 4 | 80.0 | 3 | 60.0 | 3 | 60.0 | 3 | 60.0 | 4 | 80.0 |
| 流行性感冒 | 10 | 5 | 50.0 | 4 | 40.0 | 5 | 50.0 | 4 | 40.0 | 2 | 20.0 | 4 | 40.0 | 3 | 30.0 |
| 诺如病毒感染 | 7 | 7 | 100.0 | 5 | 71.4 | 6 | 85.7 | 5 | 71.4 | 6 | 85.7 | 4 | 57.2 | 5 | 71.4 |
| 登革热 | 11 | 11 | 100.0 | 9 | 81.8 | 9 | 81.8 | 9 | 81.8 | 10 | 90.9 | 7 | 63.6 | 10 | 90.9 |
| 百日咳 | 12 | 7 | 58.3 | 8 | 66.7 | 5 | 41.7 | 8 | 66.7 | 8 | 66.7 | 8 | 66.7 | 9 | 75.0 |
| 其他12种传染病 | 54 | 50 | 92.6 | 43 | 79.6 | 41 | 75.9 | 39 | 72.2 | 39 | 72.2 | 39 | 72.2 | 32 | 59.3 |
| 总计 | 99 | 84 | 84.8 | 73 | 73.7 | 70 | 70.7 | 68 | 68.7 | 68 | 68.7 | 65 | 65.7 | 63 | 63.6 |

2.2 传染病拐点预测结果(表 2) 依据 2024 年月报数据，流行性感冒、百日咳、诺如病毒感染、手足口病、登革热出现的拐点数分别为 3、3、2、1 和 1，总计 10 处。公卫医师的准确率最高，达 50.0%，其次是 Kimi，准确率为 30.0%；AI 集成与豆包的准确率均为 20%。本研究提取 5 种传染病拐点处的预测数据，将该位置的趋势定义为“拐点处趋势”，并统计各预测模型在拐点处预测趋势的准确数量。结果显示，公卫医师对拐点处趋势的预测准确率为 80.0%，其次是 Kimi，准确率为 40.0%，豆包的准确率为 30.0%，而 AI 集成、Deep Seek 和灵犀的准确率均仅为 10.0%。公卫医师预测的 F1 分数为 0.526，其次是 Kimi 为 0.300，AI 集成与豆包的 F1 分数均为 0.267，而 Deep Seek 为 0.125。说明 Kimi 在对拐点的预测优于其他 LLM 及其集成水平，但与公卫医师预测结果的准确性相比仍有差距。

**表 2** 广州市法定传染病拐点预测结果比较(准确率/%)
**Table 2** Comparison of predictions for the inflection points of statutory infectious diseases in Guangzhou

| 指标 | 总数 | 公卫医师 | | AI集成 | | Kimi | | 灵犀 | | 豆包 | | 文心一言 | | Deep Seek | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 实发数 | % | 实发数 | % | 实发数 | % | 实发数 | % | 实发数 | % | 实发数 | % | 实发数 | % |
| 拐点 | 10 | 5 | 50.0 | 2 | 20.0 | 3 | 30.0 | 2 | 20.0 | 1 | 10.0 | 0 | 0.0 | 0 | 0.0 |
| 拐点处趋势 | 10 | 8 | 80.0 | 1 | 10.0 | 4 | 40.0 | 3 | 30.0 | 1 | 10.0 | 2 | 20.0 | 1 | 10.0 |

## 3 讨　论

本研究结果表明，尽管 AI 集成预测(准确率 73.7%)在传染病流行趋势的预测中展现出一定的应用潜力，但公卫医师的准确率更高(84.8%)，尤其是在疫情拐点识别(F1 分数人工为 0.56，AI 集成为 0.27)上仍具有优势。在面对复杂、多变的传染病系统时，人类专家的经验与情境化判断能力是当前纯数据驱动的 AI 模型难以完全替代的。但一些基于专门设计与训练(而非通用大模型)的预测模型在特定传染病(如埃博拉)的预测中曾报告过优于或接近人类专家的表现[14]。这种差异可能源于本研究采用的是未经特定流行病学数据微调的通用 LLM，其核心能力在于语言处理与知识关联，而非专门的时序预测，这提示了模型专用性与任务匹配度的重要性。

本研究结果凸显了通用 LLM 在理解现实世界复杂因果机制方面的局限性，AI 在部分传染病拐点预测中表现不佳。疫情拐点的出现往往并非单纯由疾病自然史决定，而是防控措施(如紧急灭蚊、停课)、人群行为改变(如因恐慌减少聚集)、节假日效应(如开学季)等多种突发性社会因素共同作用的结果。公卫医师能够凭借其丰富的现场经验和对本地社情政情的深刻理解，将这些“隐性知识”纳入研判。而本研究所用的 AI 模型，尽管部分具备互联网搜索功能，但其信息整合与逻辑推理仍停留在较为表面的层次，难以像人类专家那样进行深度的、情境化的因果推断。这解释了本文中即使提供了丰富的流感预测的数据维度，但 AI 的表现仍未显著提升，因为病毒株变异、人群免疫水平动态变化等关键“信号”可能被淹没在数据“噪声”中，需要更为精细的特征工程与领域知识嵌入才能被有效捕捉。

对于“数据更全，预测反而不佳”的现象(如流感)，这挑战了“数据越多，预测越准”的简单假设。本研究观察到，即使流感的监测指标更为全面，其预测准确率却相对较低。这一方面可能与流感病毒本身的高变异性有关，导致历史数据的预测价值降低。另一方面，也反映了当前 AI 模型乃至传统预测方法在处理高维、异构数据时面临的“维度灾难”问题。若不能有效识别并聚焦于核心驱动因子，过多的冗余信息反而可能干扰模型

的判断。相比之下，公卫医师能够凭借经验进行信息筛选和加权，忽略次要干扰，抓住主要矛盾。提示未来的 AI 辅助预测系统，不应仅仅是数据的简单堆砌，而需要构建一个能够模仿专家思维、进行关键信息提取和优先级排序的智能框架。

尽管 AI 模型整体表现逊于人工，但其在特定方面展现的潜力不容忽视。本研究中，Kimi 和豆包等具备实时信息检索能力的模型表现相对较好，说明外部信息的及时补充对提升预测时效性具有积极作用。这与国内外关于融合多源数据（如社交媒体舆情、气象数据）以改进传染病预测的研究方向是一致的[5]。AI 在处理海量结构化数据、进行快速模式识别和自动化报告生成方面具有天然优势。这意味着，在未来构建“人机协同”模式时，AI 可以定位为一个强大的“信息助理”和“计算引擎”，负责完成数据清洗、初步分析、生成多情景模拟等基础性、重复性工作，从而将公卫医师从繁重的信息处理中解放出来，使其能更专注于需要深度思考和创造性解决问题的拐点研判、防控策略制定等高层级任务。

从方法学上看，本研究采用提示工程方式调动通用 LLM 进行预测，是一种低成本、易推广的探索性尝试。这种方法的好处在于无需复杂的模型训练，可直接利用现成工具。但其机制也存在明显局限性：LLM 本质上是基于已有语言模式进行概率生成，而非基于流行病学原理进行严谨的数学建模，其内部决策过程无法直接观察，使得结果难以解释和追溯，这在要求高可信度的公共卫生决策中是一个重大顾虑。因此，该方法更适合用于趋势的初步筛查和辅助分析，若要将其发展为可靠的预测工具，未来的研究必须转向开发融合了流行病学机制、具备更好可解释性的专用 AI 模型，或对现有 LLM 进行基于高质量公共卫生数据的领域微调。

本研究也存在局限性：（1）数据输入格式虽经初步标准化，但不同传染病的信息维度与颗粒度仍存在差异，这可能对 AI 模型的理解和表现产生潜在影响。（2）研究采用的 AI 模型版本为特定时间点的快照，而此类模型更新迭代迅速，其性能可能随时间变化。（3）本研究主要聚焦于短期（月度）趋势预测，未能探讨 AI 在长期预测或新发传染病早期预警中的潜力，这些领域或许更能发挥 AI 在处理非结构化信息方面的优势。

综上所述，本研究证实了在当前技术条件下，公卫医师在传染病流行趋势预测，特别是关键拐点识别中不可替代的核心价值。这一发现对政策与实践的启示在于，在积极推进“人工智能+”应用的同时，应避免技术至上主义，而应着力构建以专家智慧为主导、AI 技术为辅助的“人机协同”决策新模式。未来的工作应重点投入于数据标准化、领域专用模型开发以及人机交互流程设计，最终目标是形成一条将人类专家隐性知识显性化、将 AI 计算能力精准化的融合路径，从而全面提升传染病防控体系的智能化水平和响应效能。

**数据可用性声明**　根据国家相关法律法规和政策要求，本研究数据必须严格保密，不得擅自公开。在研究过程中，研究团队采取了一系列严格的安全措施来保护数据的安全。因此，无法公开原始数据。

## 参考文献


[1] Cassini A, Colzani E, Pini A, et al. Impact of infectious diseases on population health using incidence-based disability-adjusted life years (DALYs): results from the Burden of Communicable Diseases in Europe study, European Union and European Economic Area countries, 2009 to 2013[J]. Eurosurveillance, 2018, 23(16): 17-00454.

[2] Cannon J W, Zhung J, Bennett J, et al. The economic and health burdens of diseases caused by group A *Streptococcus* in New Zealand[J]. International Journal of Infectious Diseases, 2021, 103: 176-181.

[3] 周宇辉. 我国传染病流行现状与防控体系建设研究 [J]. 中国卫生政策研究, 2023, 16(4): 74-78.

[4] Yang L Y, Yang J, He Y, et al. Enhancing infectious diseases early warning: a deep learning approach for influenza surveillance in China[J]. Preventive Medicine Reports, 2024, 43: 102761.

[5] Deiner M S, Deiner N A, Hristidis V, et al. Use of large language models to assess the likelihood of epidemics from the content of tweets: infodemiology study[J]. Journal of Medical Internet Research, 2024, 26: e49139.

[6] Tariq M U, Ismail S B. AI-powered COVID-19 forecasting: a comprehensive comparison of advanced deep learning methods[J]. Osong Public Health and Research Perspectives, 2024, 15(2): 115-136.

[7] Banholzer N, Middelkoop K, Leukes J, et al. Estimating *Mycobacterium tuberculosis* transmission in a South African clinic: spatio-temporal model based on person movements[J]. PLoS Computational Biology, 2025, 21(2): e1012823.

[8] 谭晓文, 陈文芳, 王娜娜, 等. 国内不同大型语言模型对前列腺癌围术期护理与健康教育相关问题的查询响应与效果评价 [J]. 中华男科学杂志, 2024, 30(2): 151-156.

[9] 张志玲, 李强, 闫盈盈, 等. 大语言模型为药品不良反应报告者生成反馈信息的探究 [J]. 临床药物治疗杂志, 2024, 22(10): 38-42.

[10] Wu H, Li M Y, Zhang L. Comparing physician and large language model responses to influenza patient questions in the online health community[J]. International Journal of Medical Informatics, 2025, 197: 105836.

[11] Tian T, Zhu J, Qiaoben Y. Max-margin majority voting for learning from crowds[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(10): 2480-2494.

[12] 须成杰, 覃开舟. 基于 SEIR 模型的新型冠状病毒肺炎疫情分析 [J]. 计算机应用与软件, 2021, 38(12): 87-90.

[13] 周志华. 机器学习 [M]. 北京: 清华大学出版社, 2016: 32.

[14] Munday J D, Rosello A, Edmunds W J, et al. Forecasting the spatial spread of an Ebola epidemic in real time: comparing predictions of mathematical models and experts[J]. eLife, 2025, 13: RP98005.