· 调查报告与分析 ·

基于新型冠状病毒感染疫情数据的呼吸道传染病跨地区传播风险评估模型构建

张新杨[1,2]，王恒学[1]，马涛[1]，丁松宁[1]，顾润晖[1]，严子康[1]，丁洁[1,2]

1. 南京医科大学附属南京疾病预防控制中心，南京 210003；
2. 南京医科大学公共卫生学院，南京 211116
通信作者：丁洁，E-mail：yu2an2002@163.com

【摘　要】目的 构建呼吸道传染病跨地区传播风险评估模型，为评估呼吸道传染病在跨地区传播中的风险提供理论基础、实践经验和技术支持。**方法** 收集 2020 年 1 月—2022 年 12 月国家卫生健康委员会、百度迁徙平台及各省统计局等官方公开发布的中国新型冠状病毒感染（COVID-19）疫情相关数据，选取 2020 年湖北省武汉市疫情、2021 年河北省石家庄市疫情、2021 年陕西省西安市疫情、2022 年上海市疫情、2022 年广西壮族自治区北海市疫情、2022 年海南省三亚市疫情、2022 年内蒙古自治区呼和浩特市疫情、2022 年河南省郑州市疫情和 2022 年河北省石家庄市疫情共 9 个疫情事件，通过主成分分析法构建城市疫情综合输入风险指数计算公式，对跨地区疫情综合输入风险进行量化评估；将贡献率较高的主成分作为支持向量机模型的特征向量、城市间的实际疫情传播情况作为目标变量，使用支持向量机进行模型构建与预测，并采用沙普利加性解释（SHAP）方法对模型进行可解释性分析。**结果** 主成分分析提取出当地社会经济活动因子（F_1）、疫情来源地社会经济活动因子（F_2）、城市政策执行因子（F_3）、城市人口流入与地理距离相关因子（F_4）和城市疫情传播指数因子（F_5）共 5 个因子，累计方差贡献率为 82.32%；基于主成分得分系数和主成分方差贡献率得到综合输入风险指数得分的公式为：$F=0.217\ 8\times F_1+0.184\ 1\times F_2+0.155\ 6\times F_3+0.141\ 9\times F_4+0.123\ 8\times F_5$；支持向量机模型的预测准确率为 84.88%，精确率为 75.00%，召回率为 57.14%，F1 分数为 64.86%；SHAP 分析结果显示，F_1、F_2、F_3、F_4 和 F_5 的 SHAP 值分别为 0.15、0.09、0.08、0.08 和 0.12，对评估传播风险的贡献度分别为 29.23%、16.72%、15.26%、16.09% 和 22.70%。**结论** 本研究构建的基于新型冠状病毒感染疫情数据的呼吸道传染病跨地区传播风险评估模型在呼吸道传染病跨地区输入风险的量化评估和传播预测方面具有较好的适用性和准确性，可为呼吸道传染病防控策略的科学制定提供理论支持和实践参考。

【关键词】 呼吸道传染病；跨地区传播风险；新型冠状病毒感染疫情数据；评估模型构建

A model for assessing cross-regional transmission risk of respiratory infectious diseases based on COVID-19 epidemic data: a modelling analysis

ZHANG Xinyang[1,2], WANG Hengxue[1], MA Tao[1], DING Songning[1], GU Runhui[1], YAN Zikang[1], DING Jie[1,2] (*1. Nanjing Center for Disease Control and Prevention*, *Nanjing Medical University*, *Nanjing* 210003, *China*; *2. School of Public Health*, *Nanjing Medical University*, *Nanjing* 211116, *China*)
Corresponding author: DING Jie, E-mail: yu2an2002@163.com

【Abstract】 Objective To construct a cross-regional transmission risk assessment model for respiratory infectious diseases, providing a theoretical basis, practical experience, and technical support for evaluating the risk of cross-regional transmission. **Methods** Data related to the COVID-19 epidemic in China from January 2020 to December 2022, officially released by the National Health Commission, Baidu Migration, and provincial statistical bureaus, were collected. Nine epidemic events were selected: the 2020 Wuhan, Hubei province epidemic; the 2021 Shijiazhuang, Hebei province epidemic; the 2021 Xi'an, Shaanxi province epidemic; the 2022 Shanghai epidemic; the 2022 Beihai, Guangxi Zhuang Autonomous Region epidemic; the 2022 Sanya, Hainan province epidemic; the 2022 Hohhot, Inner Mongolia Autonomous Region epidemic; the 2022 Zhengzhou, Henan province epidemic; and the 2022 Shijiazhuang, Hebei province epidemic. Principal component analysis (PCA) was used to construct a comprehensive input risk index formula for urban epidemics to quantitatively assess the comprehensive cross-regional epidemic input risk. The principal components with higher contribution rates were used as feature vectors for a support vector machine (SVM) model, with the actual intercity epidemic transmission status as the target variable. The SVM was used for model construction and prediction, and the Shapley Additive exPlanations (SHAP) method was used for model interpretability analysis. **Results** Five factors were extracted by PCA: local socioeconomic activity factor (F_1), epidemic source socioeconomic activity factor (F_2), urban policy implementation factor (F_3), urban population inflow and geographical distance-related

开放获取：CC BY-NC-ND 4.0 DOI：10.11847/zgggws1146429
基金项目：南京市医学重点专科（传染病预防控制科）；南京市重大科技专项（2021-11005）；南京医科大学南京公共卫生研究院科研创新项目（NCX2304）
第一作者：张新杨（1998—），硕士在读，研究方向：急性传染病预防与控制。
收稿日期：2024-11-15 修回日期：2025-03-15 录用日期：2025-05-14
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

factor (F_4), and urban epidemic transmission index factor (F_5), with a cumulative variance contribution rate of 82.32%. Based on the principal component score coefficients and variance contribution rates, the formula for the comprehensive input risk index score was: $F=0.217\ 8\times F_1+0.184\ 1\times F_2+0.155\ 6\times F_3+0.141\ 9\times F_4+0.123\ 8\times F_5$. The SVM model achieved a prediction accuracy of 84.88%, precision of 75.00%, recall of 57.14%, and an F1-score of 64.86%. SHAP analysis showed that the SHAP values of F_1, F_2, F_3, F_4, and F_5 were 0.15, 0.09, 0.08, 0.08, and 0.12, respectively, contributing 29.23%, 16.72%, 15.26%, 16.09%, and 22.70% to the assessment of transmission risk. **Conclusions** The cross-regional transmission risk assessment model for respiratory infectious diseases constructed in this study, based on COVID-19 epidemic data, demonstrates good applicability and accuracy in quantifying the risk of cross-regional importation and predicting transmission, providing theoretical support and practical reference for the scientific formulation of respiratory infectious disease prevention and control strategies.

【Keywords】 respiratory infectious diseases; cross-regional transmission risk; COVID-19 epidemic data; evaluation model construction

近年来，新发及重点呼吸道传染病频发，严重威胁全球公共健康、社会稳定和经济发展[1]，科学评估呼吸道传染病的跨地区传播风险并及时识别高风险地区，对疾病防控至关重要。新型冠状病毒感染（coronavirus disease 2019, COVID-19）的全球流行，为各类评估模型提供了丰富的数据基础和验证条件[2–3]，有力推动了输入风险评估模型的发展。早期研究多聚焦疫情规模和人口学数据[4–6]，随着研究的深入，学者们逐步纳入防控措施效果、疫情空间分布特征、经济及医疗水平等因素[7–10]，更全面地刻画了跨地区输入风险，提升了评估模型的科学性和实用性。然而，在呼吸道传染病防控的新形势下，影响疾病传播的新因素和新变量不断出现，现有模型在数据降维和预测性能方面均存在不足，同时缺乏有效数据进行验证，导致了评估结果的准确性低、适用性差。因此，本研究收集 2020 年 1 月—2022 年 12 月国家卫生健康委员会、百度迁徙平台及各省统计局等官方公开发布的中国新型冠状病毒感染疫情相关数据，选取不同时间段 9 个共涵盖 57 个城市的疫情事件，通过主成分分析法构建城市疫情综合输入风险指数计算公式，对跨地区疫情综合输入风险进行量化评估，将贡献率较高的主成分作为支持向量机模型的特征向量、城市间的实际疫情传播情况作为目标变量，使用支持向量机进行模型构建与预测，并采用沙普利加性解释（shapley additive explanations, SHAP）方法对模型进行可解释性分析，旨在建立一套科学、实用的跨地区传播风险评估方案，为呼吸道传染病疫情风险评估与疫情研判提供理论基础和技术支持。结果报告如下。

1 资料与方法

1.1 资料来源　通过国家、省、市卫生健康部门的官方网站收集部分地区 2020 年 1 月—2022 年 12 月每日新增病例数和累计确诊病例数，通过各省统计局发布的统计年鉴获取人口、地方国内生产总值（gross domestic product，GDP）、医疗水平、实际利用外商资金等指标，从牛津大学 COVID-19 政府响应追踪系统（Oxford COVID-19 Government Response Tracker，OxCGRT）获取牛津政策严格指数（https://ourworldindata.org/covid-stringency-index），通过百度迁徙平台（https://qianxi.baidu.com/）获取百度迁徙指数和城内出行强度，通过百度百科获取城市地理面积，通过高德地图获取城市间地理距离，并根据全国 COVID-19 规模疫情发生情况选择具有代表性的城市结成城市对（输出–输入城市对）。考虑到疫情传播的时效性，选取病例输出城市发生规模疫情的首次报告病例后次日起第 1～5 d 的数据作为主要分析对象，最终纳入的疫情传播城市对包括：2020 年湖北省武汉市疫情（湖北省孝感市、湖北省黄冈市、湖北省荆州市、湖北省咸宁市、湖北省鄂州市、湖北省襄阳市、湖北省黄石市、湖北省宜昌市、湖北省荆门市、湖北省随州市、湖北省仙桃市、河南省信阳市、重庆市、河南省南阳市、河南省驻马店市、北京市、河南省周口市、上海市、天津市、浙江省宁波市、浙江省温州市、广东省广州市和广东省深圳市），2021 年河北省石家庄市疫情（河北省保定市、河北省邢台市、河北省衡水市和河北省邯郸市），2021 年陕西省西安市疫情（陕西省咸阳市、陕西省渭南市、陕西省宝鸡市和陕西省商洛市），2022 年上海市疫情（江苏省苏州市、浙江省嘉兴市、江苏省南通市、浙江省杭州市、江苏省无锡市和江苏省南京市），2022 年广西壮族自治区北海市疫情（广西壮族自治区南宁市、广西壮族自治区钦州市、广西壮族自治区玉林市和广西壮族自治区桂林市），2022 年海南省三亚市疫情（海南省海口市、海南省万宁市、海南省陵水县和海南省乐东县），2022 年内蒙古自治区呼和浩特市疫情（内蒙古自治区鄂尔多斯市、内蒙古自治区乌兰察布市、内蒙古自治区包头市和内蒙古自治区锡林郭勒盟市），2022 年河南省郑州市疫情（河南省开封市、河南省周口市、河南省新乡市和河南省洛阳市），2022 年河北省石家庄市疫情（河北省保定市、河北省邢台市、河北省衡水市和河北省邯郸市）。

1.2 指标体系构建(表 1) 本研究依据科学性、目的性、全面性、可比性、定量与定性相结合的筛选原则进行指标的选取，在综合考虑疫情输入风险的关键影响因素基础上，参考相关文献[5−7,9,11−17]筛选指标：当计算 A 地(病例输出城市)对 B 地(病例输入城市)的综合输入风险指数时：(1)考虑到疫情输入风险受迁入规模、来源地发病人数的影响，纳入 A 地对 B 地的每日迁入指数和 A 地每日新增病例数。其中，A 地对 B 地的迁入指数=(当日 A 地迁至 B 地的人口数×当日 A 地迁出规模指数)/当日 A 地迁出的总人口数。(2)考虑到不同地区医疗水平和卫生条件对疫情存在的影响，纳入 B 地卫生机构人员数。(3)考虑到人口流动对疫情传播的影响，纳入 A 地和 B 地人口密度、城内出行强度和两地间地理距离。其中，人口密度=地区常住人口数/地理面积；城内出行强度为该城市出行的人数与该城市居住人口数比值的指数化结果，可直接通过百度迁徙平台获得。(4)考虑到各地政府部门对疫情的管控能力，纳入 A 地和 B 地的牛津政策严格指数、生产总值和 B 地实际利用外商资金。其中，牛津政策严格指数通过追踪政府干预行为的 9 项指标计算获得(https://github.com/OxCGRT/covid-policy-tracker/blob/master/documentation/index_methodology.md)，可以对全国多个地区所实施的防疫政策进行量化比较。各项指标包括不同的措施分级和措施范围，以学校停课为例：在措施分级方面，未采取措施时取值为 0，建议关闭时取值为 1，部分关闭时取值为 2，全部关闭时取值为 3；在措施适用范围方面，取值为 0 表示该政策仅在市域内有效，取值为 1 表示该政策由该市所在省颁布，在全省范围内有效。对每项指标按最大值 100 进行标准化，第 t 日第 j 个指标的得分公式为：$I_{jt}=100\times[(V_{jt}-0.5\times(1-f_{jt})]/N_j$。其中，$V_{jt}$ 表示第 t 日第 j 个指标的措施等级；当 $V_{jt}=0$ 时，$I_{jt}=0$；f_{jt} 表示第 t 日第 j 个指标的措施范围值；N_j 表示第 j 个指标的最高等级值。第 t 日的政策严格指数为当日 9 个指标的算术平均值，值域为(0，100)。(5)以上共计 13 项指标纳入评估体系，其中，属性为正表示该指标与疫情输入风险呈正相关，属性为负表示该指标与疫情输入风险呈负相关；分别将各指标记为不同符号，如 *X1*、*X2*。具体见表 1。

表 1 呼吸道传染病跨地区传播风险评估指标体系

Table 1 Risk assessment indicator system for cross-regional transmission of respiratory infectious diseases

指标	单位	属性	符号	更新周期	来源
A地人口密度	万人/平方千米	正	*X1*	年	统计年鉴
A地GDP	亿元	正	*X2*	年	统计年鉴
A地对B地的每日迁入指数		正	*X3*	日	百度迁徙平台
A地每日新增病例数	例	正	*X4*	日	卫生健康委数据
A地城内出行强度		正	*X5*	日	百度迁徙平台
A地牛津政策严格指数		负	*X6*	日	OxCGRT
B地卫生技术人员数	万人	负	*X7*	年	统计年鉴
B地人口密度	万人/平方千米	正	*X8*	年	百度百科
B地城内出行强度		正	*X9*	日	百度迁徙平台
B地牛津政策严格指数		负	*X10*	日	OxCGRT
B地GDP	亿元	正	*X11*	年	统计年鉴
两地间地理距离	千米	负	*X12*	年	高德地图
B地实际利用外商资金	亿美元	正	*X13*	年	统计年鉴

注：在疫情传播中，A地被界定为病例输出城市，B地被界定为病例输入城市。

1.3 统计分析 采用 Excel 2021 建立数据库，应用 R 4.3.0 和 SPSS 27.0 软件进行统计分析。通过主成分分析法提取评价指标构建疫情综合输入风险指数计算公式，计算和分析各地的风险指数，并在此基础上采用支持向量机进行模型构建和预测。本研究主要参照文哲等[18]研究方法，为克服数据多维度和多重共线性问题，首先对 13 项指标进行中心标准化处理，使其均值为 0、标准差为 1，然后通过主成分分析法转化指标，按照方差贡献率对其赋以权重，构建疫情综合输入风险指数的数学公式；然后根据主成分分析结果，借鉴王纯杰等[19]的分析方法，采用 R 语言 e1071 包将城市对数据划分为训练集和测试集，并以得到的 m 个主成分因子作为其特征向量，使用训练好的模型对测试集进行预测。以城市对在我国 COVID-19 期间每天的实际传播情况作为金标准，将确有病例传播的情况标记为阳性、确无病例传播的情况标记为阴性，使用 confusionMatrix 函数评估模型的性能，并通过预测准确率、精确率、召回率和 F1 分数等指标对模型效果进行综合评价：(1)准

确率=(TP+TN)/(TP+FP+FN+TN)；(2)精确率=TP/(TP+FP)；(3)召回率=TP/(TP+FN)；(4)F1 分数=2×精确率×召回率/(精确率+召回率)。其中，TP 指模型正确预测为阳性的样本数量，FP 指模型错误预测为阳性的样本数量，TN 指模型正确预测为阴性的样本数量，FN 指模型错误预测为阴性的样本数量。使用 SHAP 方法对模型进行解释，评估各主成分因子对模型预测的贡献，揭示各因子对预测结果的影响程度。

2 结　果

2.1　主成分分析(表 2)　首先对 A 地人口密度、A 地 GDP、A 地对 B 地的每日迁入指数、A 地每日新增病例数、A 地城内出行强度、A 地牛津政策严格指数、B 地卫生技术人员数、B 地人口密度、B 地城内出行强度、B 地牛津政策严格指数、B 地 GDP、两地间地理距离和 B 地实际利用外商资金 13 项指标的原始数据进行标准化处理，正向指标保留原始值、负向指标采用倒数法变换以统一各指标的方向性，提高可比性。主成分分析中以特征值>1 为标准，提取前 5 个因子作为主成分，分别为当地社会经济活动因子(F_1)、疫情来源地社会经济活动因子(F_2)、城市政策执行因子(F_3)、城市人口流入与地理距离相关因子(F_4)和城市疫情传播指数因子(F_5)，累计方差贡献率为 82.32%。其中，F_1 在指标 B 地人口密度、B 地城内出行强度、B 地 GDP 和 B 地实际利用外商资金上的系数值较大，综合反映病例输入城市的社会经济活动水平、人口流动性和对外开放程度，方差贡献率为 21.78%；F_2 在指标 A 地人口密度和 A 地 GDP 上的系数值较大，综合反映病例输出城市的人口集中度和经济水平，方差贡献率为 18.41%；F_3 在指标 A、B 两地牛津政策严格指数上的系数值较大，政策严格指数通常反映政府对社会管理、公共卫生或者其他领域政策的执行力度，方差贡献率为 15.56%；F_4 在指标 A 地对 B 地的每日迁入指数和两地间地理距离上的系数值较大，每日迁入指数反映 B 地对外来人口的吸引力以及人口流动的规模和频率，两地间地理距离在一定程度上决定疾病传播的难易程度以及公共卫生响应的复杂性，方差贡献率为 14.19%；F_5 在指标 A 地每日新增病例数上的系数值较大，该指标综合反映疫情的传播速率以及该地区当前的健康风险及传染病的活跃度，方差贡献率为 12.38%。分别将上述各主成分的方差贡献率作为权重，得到综合输入风险指数的得分公式为：$F=0.217\ 8\times F_1+0.184\ 1\times F_2+0.155\ 6\times F_3+0.141\ 9\times F_4+0.123\ 8\times F_5$。

表 2　旋转后的成分矩阵图
Table 2　Rotated component matrix diagram

指标名称	成分				
	F_1	F_2	F_3	F_4	F_5
A地人口密度		0.910			
A地GDP		0.903			
A地对B地的每日迁入指数				0.919	
A地每日新增病例数					0.865
A地城内出行强度					
A地牛津政策严格指数			0.966		
B地卫生技术人员数					
B地人口密度	0.793				
B地城内出行强度	−0.725				
B地牛津政策严格指数			0.961		
B地GDP	0.823				
两地间地理距离				0.921	
B地实际利用外商资金	0.798				
特征值	2.831	2.393	2.023	1.845	1.610
方差贡献率/%	21.78	18.41	15.56	14.19	12.38
累计方差贡献率/%	21.78	40.19	55.75	69.94	82.32

注：在疫情传播中，A地被界定为病例输出城市，B地被界定为病例输入城市。

2.2　支持向量机的建立及预测效能评价　基于主成分分析结果，将 70% 的数据作为训练集、余下的 30% 数据作为测试集，将提取的 5 个主成分因子作为支持向量机的特征向量、城市对间的实际传播情况作为目标变量。结果显示，模型预测准确率为 84.88%(73/86)，精确率为 75.00%(12/16)，召回率为 57.14%(12/21)，F1 分数为 64.86%。

2.3　基于 SHAP 算法的可解释性分析(图 1)　结果显示，针对 B 地的当地社会经济活动因子(F_1)在所有评估传播风险的过程中具有最强的预测价值，随着 F_1 值的增加，传播发生的风险也随之上升；其次是城市疫情传播指数因子(F_5)、疫情来源地社会经济活动因子(F_2)等特征也对传播预测有重要贡献(图 1A)。此外，F_1、F_2、F_3、F_4 和 F_5 对评估传播风险的贡献度分别为 29.23%、16.72%、15.26%、16.09% 和 22.70%。特征因子对模型输出的影响分布进一步展示了各主成分因子与传播结果之间的正负关系(图 1B)，其中，当地社会经济活动因子(F_1)、疫情来源地社会经济活动因子(F_2)和城市疫情传播指数因子(F_5)的增加对预测结果有正向影响，表明其与传播风险上升正相关；而城市政策执行因子(F_3)和城市人口流入与地理距离相关因子(F_4)的增加对预测结果有负向影响，表明其与传播风险下降相关。

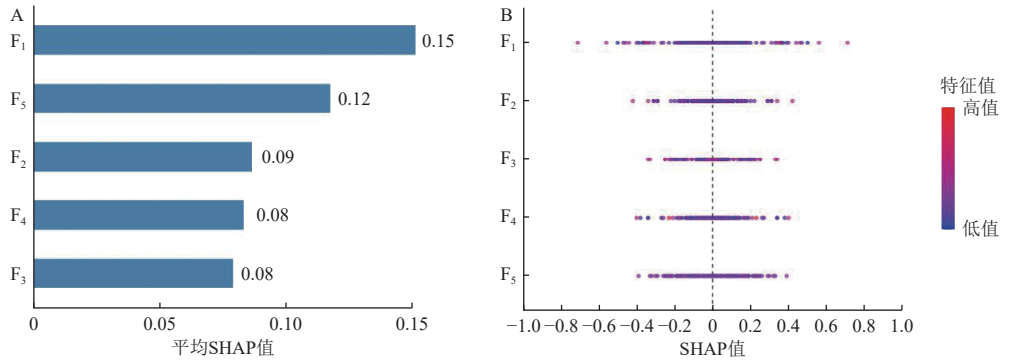


图 1　运用 SHAP 值解释预测模型

Fig. 1　Explaining the prediction model using SHAP values

3 讨　论

本研究基于现有传染病风险评估模型[4,9,17,20–21]，结合新型冠状病毒感染疫情的传播特点，构建了适用于呼吸道传染病的跨地区传播风险评估模型，研究中采用主成分分析法和支持向量机对城市间的输入风险进行定量评估并对城市间疫情传播风险进行预测，同时进行影响因素分析。结果显示，本研究构建的呼吸道传染病的跨地区传播风险评估模型具有较强的操作性和较高的准确性，为呼吸道传染病的早期预警和精准防控提供了实用的模型框架和可靠的分析手段。

从综合输入风险指数变化上看，不同城市对的综合输入风险指数在一定程度上反映了新型冠状病毒感染疫情的实际变化情况，表明本研究具有实际应用价值。目前，多种传染病风险评估模型在指标的选择上容易出现过度设计的问题，导致模型运算负担大，不利于实际应用。有研究表明，传染病监测信息利用率低的主要原因为数据整合不足、专业技能不足以及缺乏操作性强、容易分级响应的传染病预警指标等[22]，同时对一线工作人员的调查亦发现有效的预警措施需要具备操作方便和预警信息简洁易理解等特点[23–24]。本研究采用主成分分析法从复杂的指标评价体系中提取重点因素并形成风险指数计算公式，具备操作方便和预警信息直观等优点，有利于呼吸道传染病疫情的风险研判，相较于其他研究[25]，本研究在降低数据复杂性的同时，还具备高适应性和可操作性。

在预测模型的选择方面，Kalayou M H 等[26]研究中支持向量机模型展现出优异的预测性能，其准确率为 81%、灵敏度（召回率）为 78.2%、特异度为 83.3%、F1 分数为 81%。本研究中，模型预测准确率为 84.88%，精确率为 75.00%，召回率为 57.14%，F1 分数为 64.86%，表明模型在阴性样本识别方面具有明显优势。F1 分数是精确率和召回率的调和平均值，可综合衡量模型在类别不均衡情况下的性能，尽管由于样本量限制，本研究的召回率和 F1 分数相对较低，但整体预测精度和稳定性仍维持在较高水平，可弥补在疫情防控早期因数据收集困难导致分析研判准确率低的不足。此外，训练数据的配比选择会影响模型判断效果，未发生传播的城市对比例过高可导致支持向量机对部分高风险区域捕捉效果差，这一现象与类别不平衡问题的研究结果一致[27]。因此在实际应用中，可根据工作需求选择合适的配比，或将支持向量机与其他方法相结合，根据呼吸道传染病疫情防控的不同阶段、不同场景对模型灵敏度和特异度的需求建立多阶段混合模型，提升传播风险特征的识别效果，为呼吸道传染病的风险评估和预警策略提供更全面、科学的依据。

在呼吸道传染病输入风险的影响因素分析方面，本研究利用 SHAP 分析揭示了各主成分因子在呼吸道传染病跨地区传播预测中的贡献。结果显示，F_1 因子在传播风险预测中占据主导地位，贡献度达到 29.23%，表明呼吸道传染病在人口密集、人员流动性高、公共交通出行频繁的城市更容易发生传播，与 Kaiser J C 等[28]研究结果一致，因此在早期防控新发或重点呼吸道传染病时应对这些区域实施重点关注，采取严格的监测和控制措施最大程度地降低传播风险。F_5 和 F_2 因子贡献度分别为 22.70% 和 16.72%，均为推动呼吸道传染病跨区域传播的潜在因素，与金亚亚等[8]研究结果一致，发现疫情源头地的病例数增多时，经济发达和人口集中的地区往往具有更高的病例外溢风险；与本研究结果相似，刘云广等[29]进一步强调了人口密度在跨区域传播中的重要性，特别是流感等常见呼吸道传染病。此外，F_3 和 F_4 因子对呼吸道传染病的传播均具有抑制作用，结果显示，政策执行严格区域中传染病的传播受到有效抑制，同时较远地理位置距离和较少人员流动对传播风险亦有抑制作用，提示在新发或重点呼吸道传染病的防控中应提升政策执行

力，及时实施流动限制和隔离措施，以减少传播风险[30–31]。因此，本研究建议在新发或重点呼吸道传染病防控中应重点关注社会经济活动活跃和病例增长迅速的区域，强化政策执行，完善区域边界管理，建立针对高流动性场所的健康监测体系，并动态调整跨区域流动控制策略；同时在常规传染病研究中，还应深入探索社会经济活动和人口流动的长期影响，优化监测网络布局，为公共卫生策略的制定提供科学依据。

尽管本研究构建的基于新型冠状病毒感染疫情数据呼吸道传染病跨地区传播风险评估模型在呼吸道传染病跨地区输入风险的量化评估和传播预测方面具有较好的适用性和准确性，可为未来呼吸道传染病防控策略的科学制定提供理论支持和实践参考，但仍存在一定的局限性。首先，模型的性能在实际应用中可能受数据更新延迟、来源异质性和地区间数据完整性差异等因素的限制，因此未来研究应通过引入动态数据更新机制和整合多源数据的方法优化模型。其次，本研究是基于新型冠状病毒感染疫情数据构建的风险评估框架，部分指标（如牛津大学政策严格指数等）最初是应用于新型冠状病毒感染疫情，已有研究表明其可用于登革热和猴痘的防控政策，未来研究可纳入其他呼吸道传染病数据进一步验证模型的泛化能力和适用性[32–33]；此外，在特定场景下，仍可参考其计算方式并将其应用于不同传染病的风险评估。最后，由于所纳入的城市对数量有限，样本量可能不足以全面反映实际情况，在后续工作应考虑扩大样本量进行外部验证，增强结果的可靠性。

数据可用性声明　本分析中使用的所有输入数据的引用和元数据均可从公开数据库下载，具体来源包括：（1）疫情数据：来源于国家、省、市各级卫生健康部门官方网站或官方媒体；（2）人口流动数据：来源于百度迁徙平台（https://qianxi.baidu.com）；（3）社会经济数据和医疗资源数据：来源于各省统计年鉴、地方统计局官网；（4）政策响应数据：来自牛津大学新冠政府响应追踪系统（https://ourworldindata.org/covid-stringency-index）；（5）地理距离数据：基于高德地图获取（https://ditu.amap.com）。本研究使用的相关数据若因平台更新、调整、关闭，数据无法获取，作者不再单独提供相应数据，敬请理解。

参考文献

[1] 王岩, 陈俊利, 冯文佳, 等. 全球、高 SDI 国家和中国<5 岁儿童 1990—2019 年下呼吸道感染疾病负担变化趋势 [J]. 中国公共卫生, 2024, 40(7): 840-848.

[2] Badr H S, Zaitchik B F, Kerr G H, et al. Unified real-time environmental-epidemiological data for multiscale modeling of the COVID-19 pandemic[J]. Scientific Data, 2023, 10(1): 367.

[3] Soriano-Arandes A, Brett A, Buonsenso D, et al. Policies on children and schools during the SARS-CoV-2 pandemic in western Europe[J]. Frontiers in Public Health, 2023, 11: 1175444.

[4] 胡建雄, 刘涛, 肖建鹏, 等. 广东省新型冠状病毒肺炎输入风险评估与预警 [J]. 中华流行病学杂志, 2020, 41(5): 657-661.

[5] 沈思鹏, 魏永越, 赵杨, 等. 全球新型冠状病毒肺炎疫情对我国的输入风险评估 [J]. 中华流行病学杂志, 2020, 41(10): 1582-1587.

[6] 郑阳, 陈艳伟, 李爽, 等. 运用风险指数法对新冠肺炎疫情开展快速风险评估——以新发地市场疫情为例 [J]. 国际病毒学杂志, 2021, 28(5): 407-411.

[7] 金山, 王若希. 基于 PSR 模型的城市疫情风险评估指标体系构建——以新型冠状病毒肺炎为例 [C]//面向高质量发展的空间治理——2020 中国城市规划年会论文集 (01 城市安全与防灾规划). 成都: 中国城市规划学会, 成都市人民政府, 2021: 136-143.

[8] 金亚亚, 毛晓蒙, 王霞, 等. 新冠疫情的空间分布及影响因素——基于中国病例数据的动态分析 [J]. 陇东学院学报, 2021, 32(5): 16-23.

[9] 杨昕娉, 庞明樊, 梁作如, 等. 我国陆地接壤国家新冠病毒病疫情及输入风险评估 [J]. 疾病监测, 2020, 35(4): 283-287.

[10] 张奕, 孙晓东, 田睿, 等. 构建境外输入新型冠状病毒肺炎病例的快速风险评估体系 [J]. 中华流行病学杂志, 2022, 43(5): 663-668.

[11] 胡建雄, 何冠豪, 刘涛, 等. 新型冠状病毒肺炎疫情初期湖北省输出风险评估 [J]. 中华预防医学杂志, 2020, 54(4): 362-366.

[12] 王峥, 樊红. 湖北省 COVID-19 疫情时空聚类分析及社会经济影响因素研究 [J]. 武汉大学学报: 工学版, 2021, 54(9): 866-873.

[13] 李寻昌, 刘怡帆, 王雅倩, 等. 基于主成分分析的新型冠状病毒感染防控影响因素研究 [J]. 甘肃科学学报, 2023, 35(1): 49-56,63.

[14] 李辉智, 易大莉, 李高明. 重庆市新型冠状病毒肺炎疫情风险分区分级评估方法比较研究 [J]. 重庆医科大学学报, 2020, 45(7): 870-875.

[15] 韦原原, 江南, 陈云海, 等. 顾及时空对象空间相互作用的疫情风险评估建模与应用 [J]. 地球信息科学学报, 2021, 23(2): 274-283.

[16] 曹永旺, 刘樱, 周春山. 城市群视角下新冠肺炎疫情的时空扩散特征与影响因素 [J]. 地域研究与开发, 2021, 40(3): 1-7.

[17] 袁宏永, 陈涛, 黄丽达, 等. 全国及湖北省新冠肺炎综合风险指数分析及比较 [J]. 科技导报, 2020, 38(4): 21-28.

[18] 文哲, 向媛媛, 周李承, 等. “一带一路”合作及沿线国家口岸传染病输入风险评估 [J]. 中国国境卫生检疫杂志, 2022, 45(3): 210-215.

[19] 王纯杰, 张乐, 陈嘉, 等. 基于随机森林插补的缺失数据的印度肝脏病人数据支持向量机分类分析 [J]. 吉林师范大学学报: 自然科学版, 2020, 41(4): 36-40.

[20] 梁跃红, 何舰, 周晓昕, 等. 传染病暴发/流行风险评估指标体系的构建 [J]. 实用预防医学, 2023, 30(5): 621-625.

[21] 张婷瑜, 刘艳慧, 鲁影, 等. 传染病多点触发监测预警指标体系构建 [J]. 中国公共卫生, 2024, 40(4): 489-495.

[22] 柳东如. 湖北省统筹区域传染病监测预警体系建设实践与思考 [J]. 公共卫生与预防医学, 2023, 34(6): 1-4.

[23] 袁方, 任海玲, 赵梦, 等. 我国传染病预警监测模型研究进展综述 [J]. 价值工程, 2023, 42(33): 162-165.

[24] 程潇, 黄利群, 陈丹, 等. 我国口岸城市呼吸道传染病多源数据预警综合指标体系建立 [J]. 中国公共卫生, 2022, 38(9): 1160-1163.

[25] 李登峰, 龚政, 曾佩琪, 等. 新型冠状病毒感染疫情涉疫场所传播风险评估模型的建立及验证 [J]. 华南预防医学, 2024, 50(1): 87-90.

[26] Kalayou M H, Kassaw A A K, Shiferaw K B. Empowering child health: harnessing machine learning to predict acute respiratory infections in Ethiopian under-fives using demographic and health survey insights[J]. BMC Infectious Diseases, 2024, 24(1): 338.

[27] 张乐, 王如意, 杨慧, 等. 重采样技术在中老年居民糖尿病不平衡数据分类中的应用 [J]. 现代预防医学, 2023, 50(7): 1339-1344.

[28] Kaiser J C, Stathopoulos G T. Socioeconomic correlates of SARS-CoV-2 and influenza H1N1 outbreaks[J]. The European Respiratory Journal, 2020, 56(3): 2001400.

[29] 刘云广, 黄茜, 张乐, 等. 2004—2018 年中国流行性感冒发病率的时空分析及其预测模型研究 [J]. 中华疾病控制杂志, 2023, 27(2): 176-183,190.

[30] 崔明洁, 姚霞, 方昊然, 等. SARS 与 COVID-19 时空传播差异性及影响因素分析 [J]. 地球信息科学学报, 2021, 23(11): 1910-1923.

[31] 朱也凡, 张书志, 段春梅, 等. 新型冠状病毒肺炎疫情防控措施对上海市嘉定区法定传染病发病率的影响 [J]. 疾病监测, 2023, 38(2): 181-184.

[32] Islam J, Hu W B. Rapid human movement and dengue transmission in Bangladesh: a spatial and temporal analysis based on different policy measures of COVID-19 pandemic and Eid festival[J]. Infectious Diseases of Poverty, 2024, 13(1): 99.

[33] Huang Q R, Sun Y X, Jia M M, et al. Risk assessment for cross-border transmission of multi-country Mpox outbreaks in 2022[J]. Journal of Infection and Public Health, 2023, 16(4): 618-625.