· 流行病与统计方法 ·

基于症状等多源数据的急性呼吸道传染病发病趋势预测模型的构建与评价

邬安琪[1],文泽轩[2,3],吴强松[1],汪晨夕[1],施建华[1]
1. 徐汇区疾病预防控制中心,上海 200237;2. 复旦大学上海市重大传染病和生物安全研究院;
3. 复旦大学公共卫生学院流行病学教研室/公共卫生安全教育部重点实验室

摘要:目的 构建基于多源数据的遗传算法优化的支持向量机(Genetic Algorithm optimized Support Vector Machine, GA-SVM)模型预测急性呼吸道传染病并评价其预测效果,为建立呼吸道传染病早期预警体系提供参考。**方法** 根据2020—2022 年上海市徐汇区的症状监测、气象及大气污染和严格指数数据,在潜在预测变量中挑选最佳延迟周数的预测变量后,筛选出预测重要性最强的变量作为自变量。按 1∶4比例将全时间序列划分为验证集和训练集,利用遗传算法优化参数,以呼吸道传染病每周新增病例数为因变量构建 GA-SVM 模型。采用均方根误差、平均绝对百分比误差、预测相关系数和决定系数对模型预测结果进行评价。**结果** 预测重要性最强的变量为:延迟 2 周的严格性指数、延迟 1 周的症状监测病例数、延迟 1 周的最高气温、延迟 2 周的学校活动和延迟 1 周的臭氧(O_3)指数。建立的 GA-SVM 模型最优参数 C=18.04,γ=0.175 4,模型的平均均方根误差为 6.362,平均绝对百分比误差为 24.59%,平均预测相关系数和决定系数分别为 0.896 和 0.804。**结论** 该模型对于徐汇区急性呼吸道传染病报告病例数有着良好的预测效果,证实了 GA-SVM 应用于基于症状等多源数据实现呼吸道传染病预测的可行性,为多源数据应用于传染病早期预警提供方法参考。

关键词:多源数据;支持向量机;急性呼吸道传染病;症状监测;预测
中图分类号:R181.3 **文献标志码:**A **文章编号:**1003-8507(2025)02-220-07
DOI:10.20043/j.cnki.MPM.202407206

Construction and evaluation of a prediction model for the trend of acute respiratory infectious diseases based on multi-source data including symptom surveillance

WU An-qi*, WEN Ze-xuan, WU Qiang-song, WANG Chen-xi, SHI Jian-hua
*Xuhui Center for Disease Control and Prevention, Shanghai 200237, China

Abstract: Objective Toconstruct a Genetic Algorithm optimized Support Vector Machine (GA-SVM) model based on multi-source data predicting acute respiratory infectious diseases and toevaluate its predictive effectiveness, providing a reference for establishing an early warning system for respiratory infectious diseases. **Methods** Symptom surveillance cases, meteorological and atmospheric pollution, data and stringency index obtained from 2020 to 2022 were used as modeling and forecasting samples, respectively. By picking up the optimum lagging week number of the potential predictive variables and filter out the most important variables successively, the independent variables were obtained. Then the full time series data were divided into validation set and training set in a 1:4 ratio. The parameters were optimized by genetic algorithm. We used the weekly number of new cases of respiratory infectious diseases as the dependent variable to structure the GA-SVM model. The performance was evaluated based on the following metrics: root mean square error (RMSE), meansabsolute percentage error (MAPE), predictive correlation coefficient (PCC) and R-squared (R^2). **Results** The most important variables were stringency index with 2-weeks-lag, symptom surveillance cases with 1-week-lag, maximum temperature with 1-week-lag, school activities with 2-weeks-lag and O_3 index with 1-week-lag. The GA-SVM model performed best when C=18.04, γ=0.175 4 while average RMSE=6.362, average MAPE=24.59%, average PCC=0.896 and average R^2=0.804. **Conclusions** The model showsgood predictive performance for the reported cases of acute respiratory infectious diseases in Xuhui District, which confirms the feasibility of applying GA-SVM to multi-source data based on symptom monitoring for

基金项目:2021 年度徐汇区医学科研项目(SHXH202147)
作者简介:邬安琪(1989—),女,本科,公共卫生主管医师,研究方向:传染病预防与控制
通信作者:施建华, E-mail: 13381644949@189.cn

predicting respiratory infectious diseases, providing methodological references for the application of multi – source data in the early warning of infectious diseases.

Keywords: Multi – source data; SVM; Acute respiratory infectious diseases; Symptom surveillance; Prediction

急性呼吸道传染病包括流行性感冒、水痘、新型冠状病毒感染等,因其病原组成复杂、传播速度快、范围广的特点,一旦达到一定规模,干预措施控制效果相对有限,危害人群健康,造成严重疾病负担。我国虽已将强化监测预警纳入“十四五”规划,但目前我国的预警监测系统预警关口相对滞后、来源相对单一等问题仍旧存在[1]。基于症状的监测可以有效前移关口,更及时、敏感地触发预警信号,控制突发事件规模[2]。我国的症状监测虽然在 2003 年严重急性呼吸综合征流行后发展迅速,但目前仍存在症状监测方法应用不足、监测技术落后等问题[3]。随着多源数据应用的发展,机器学习作为人工智能的重要组成部分,在临床诊疗、病媒生物风险预测、慢性病影响因素研究等领域应用日益广泛[4-7],而在传染病监测预警方面应用尚在探索阶段。本研究将症状监测与机器学习方法结合,以急性呼吸道传染病症状指标为基础,同时纳入环境、气象、政策等多源数据,构建遗传算法优化的支持向量机模型(Genetic Algorithm optimized Support Vector Machine, GA – SVM)并评估其预测效果,为传染病的早期预警提供方法基础。

1　材料与方法

1.1　*数据来源*　症状监测病例数据来源于“上海市急性呼吸道感染综合监测平台”(以下简称“呼综平台”)中的徐汇区哨点医疗机构日报告病例数据。呼吸道传染病发病数据来源于“中国疾病预防控制信息系统”中的徐汇区医疗机构日报告病例数据。气象数据来源于“2345 天气王”网站。环境数据来源于上海市环境生态环境局网站,主要包括上海市 2020 年 3 月至 2022 年 1 月的最高气温、最低气温、最大温差及空气污染指数 AQI、$PM_{2.5}$、PM_{10}、SO_2、NO_2、O_3、CO 指数等指标。学校活动情况来源于官方新闻,包括寒假、暑假、启动线上教学及恢复线下教学的时间段。新冠疫情期间的呼吸道传染病疫情与政府响应措施密切相关,反映上海市政府疫情防控政策以及公共卫生干预措施的严格性指数通过牛津新冠政府响应追踪器[8-9]获得。由于研究中能够获得的部分呼吸道传染病发病病例稀少,受不确定性影响较大,故将症状监测数据、气象数据和严格指数数据进行预处理,将症状监测数据从日监测数据转换为周合计数据,将气象数据和严格指数从日数据转换为周均数据。

呼综平台自 2020 年 3 月启用,2022 年 2 月起受新冠流行影响因临床救治压力增大、医疗机构人手紧缺等原因,监测数据的报告数量及质量显著下降、明显失真,故不再利用该系统后期数据。自 2023 年 4 月起,徐汇区急性呼吸道综合监测数据切换至新平台“传染病综合监测预警平台”(原呼综平台停用)报送。由于新平台使用初期必填项设置缺陷、字段不足、医疗机构填报依从性较差等原因,新平台数据质量无法客观体现实际情况,截至研究时,尚无法纳入新数据进行训练或测试。故本次建模使用的数据时间段限于 2020 年 3 月 15 日—2022 年 1 月 30 日。

1.2　*病例纳入标准*

1.2.1　急性呼吸道传染病　法定报告传染病[10]中的急性呼吸道传染病,包括新型冠状病毒肺炎/新型冠状病毒感染、百日咳、流行性乙型脑炎、人感染 H_7N_9 禽流感、猩红热、流行性脑脊髓膜炎、麻疹、人感染高致病性禽流感、传染性非典型肺炎、白喉、风疹、流行性腮腺炎、流行性感冒和手足口病;以及上海市按照丙类传染病管理的水痘[11]。

1.2.2　症状监测病例　就诊距离发病 3 天内的发热伴咳嗽和/或咽痛症状的病例。

1.3　*研究方法*　利用 R 4.2.2 统计软件,基于 caret 软件包[12]进行预测变量选择,基于遗传算法和 e1071 软件包进行遗传算法求解最佳参数,构建支持向量机模型(Support Vector Machine, SVM),基于 ggplot2 软件包进行结果可视化。

1.3.1　变量筛选　将症状监测病例数、最高气温、最低气温、最大温差、AQI、$PM_{2.5}$、PM_{10}、SO_2、NO_2、O_3、CO、学生活动情况、严格性指数被列为潜在预测变量。各潜在预测变量被认为可能在 1 ~ 2 周的时间内对呼吸道传染病疫情产生影响,因此利用不同延迟周数下各潜在预测变量构建类泊松分布模型,并保留与报告传染病病例数之间的均方根误差(Root Mean Square Error, RMSE)最小的预测变量延迟周数,并根据最佳延迟周数处理原始数据集。使用向后选择的递归特征消除算法[13]计算保留不同变量数时进行建模预测的均方根误差,以及各预测变量的均方误差增加百分比(Percentage Increase in Mean Squared Error, %IncMSE),该值越大,认为对应模型中该变量的预测重要性越大。通过递归地训练模型、评估并移除预测重要性最弱的变量,从而确定各变量对模型预测精度的贡献程度。在每次迭代中,算法根据模型的表现评估剩余变量的重要性。方差膨胀系数(Variance inflation factor, VIF)被用于衡量处理后的数据集中各

预测变量之间的多重共线性严重程度，排除共线性[14-15]严重的潜在预测变量。最终保留的变量即为那些对预测模型贡献最大的预测变量。随后，将预测重要性较强的预测变量纳入 SVM 模型构建。

1.3.2 遗传算法参数求解 基本思路为首先任选一组参数作为 SVM 的初始参数，再根据实际需要对参数进行编码，从而构造第一代遗传群体。针对第一代遗传群体计算误差，误差越小，则适应度越大，适应度大的个体会被遗传给下一代，通过当前一代的群体算子进行交叉计算、变异等遗传处理即可产生下一代群体；再通过迭代步骤不断优化 SVM 参数，直至参数满足条件或达到最大迭代次数[16]。在本研究中，遗传算法的终止条件包括两个方面：一是当适应度值达到预设阈值或适应度值在连续多代的变化低于 0.001 时，算法将认为已收敛并停止迭代；二是当迭代次数达到 100 次时，即使未达到适应度收敛条件，也会终止算法。通过这些设置，我们保证了在优化过程中既能获得较优的 SVM 参数组合，同时避免了过多的计算资源消耗。

1.3.3 SVM 回归预测 通过纳入所选预测变量构建 SVM 模型。按照 1∶4 的比例将全时间序列划分为验证集和训练集，利用遗传算法优化 SVM 参数，以呼吸道传染病每周新增病例数为因变量，以 1.3.1 小节所挑选的最佳延迟周数的预测变量用于 SVM 回归计算。利用重抽样方法估计预测模型的不确定性。

1.3.4 预测准确度评价 采取 5 - 折交叉验证[17]测试预测模型对测试集的预测效果，使用 RMSE、平均绝对百分比误差（Mean Absolute Percentage Error，MAPE）、预测相关系数（Predictive Correlation Coefficient，PCC）和决定系数（R^2）并绘制预测值与实际值的时间序列图来对模型预测结果进行评价。

2 结 果

2.1 *数据结构* 2020 年 3 月 15 日至 2022 年 1 月 30 日共报告 2 372 例急性呼吸道传染病病例，包括：水痘 1 364 例，流行性感冒 533 例，手足口病 266 例，新型冠状病毒肺炎/新型冠状病毒感染 114 例，流行性腮腺炎 79 例，猩红热 13 例，流行性脑脊髓膜炎 2 例，风疹 1 例，其间无百日咳、流行性乙型脑炎、人感染 H_7N_9 禽流感、麻疹、人感染高致病性禽流感、传染性非典型肺炎、白喉病例报告。同期共报告症状监测病例 54 019 例。

2.2 *徐汇区呼吸道传染病疫情描述性分析* 徐汇区 2020 年 3 月至 2022 年 1 月的传染病报告病例数与症状监测病例数在 2020 年 9 月份后呈现相似的变化趋势。总体上，症状监测病例数越高，呼吸道传染病报告病例数也会随之升高。需要注意的是，由于 2020 年 4 月 27 日起实行“学校师生出现发热等症状凭《上海市学校方便就诊卡》至医疗机构免费检测核酸”和中小学校逐步复课的影响，症状监测病例数出现了小高峰，而类似高峰并未在传染病报告病例数的变化中观察到。除此之外，研究阶段内包含了 2 次冬春季（2020 年 10 月至 2021 年 2 月和 2021 年 10 月至 2022 年 2 月），对应阶段为呼吸道传染病的高发季节，这种季节性的波动在传染病报告和症状监测病例数中均有所体现。详见图 1。

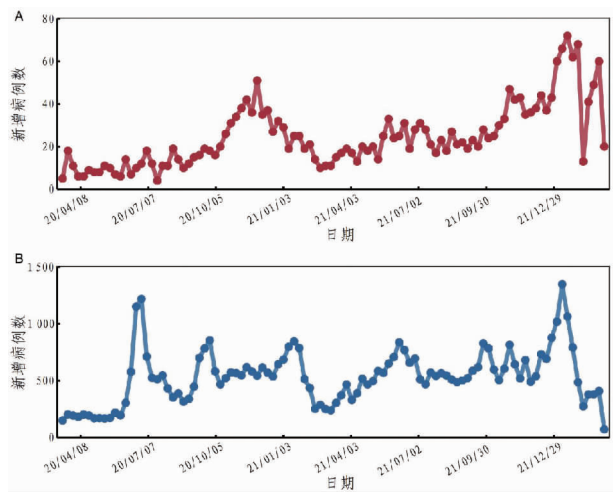
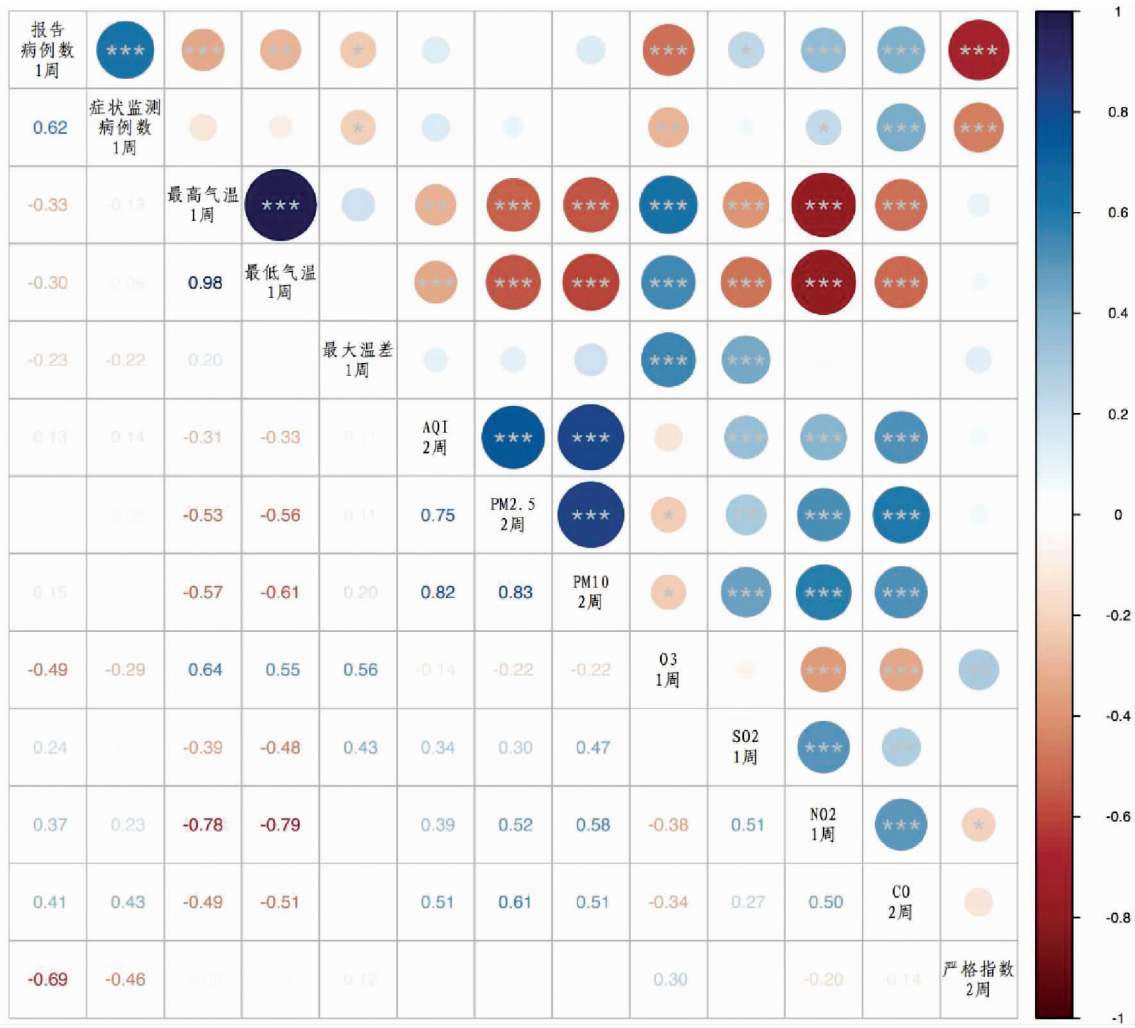


图 1 2020 年 3 月 15 日—2022 年 1 月 30 日急性呼吸道传染病报告病例数（A）与呼吸道症状监测病例数（B）变化趋势

Fig. 1 Trends in the number of reported acute respiratory infectious diseases cases (A) and respiratory symptom surveillance cases (B) from March 15, 2020, to January 30, 2022

2.3 *潜在预测变量最佳延迟周数与相关性* 表 1 显示了类泊松分布模型计算各潜在预测变量在不同延迟周数下建模时的均方根误差，其中症状监测病例数、气温、O_3、SO_2 和 NO_2 在 1 周延迟情况下均方根误差最小，其他潜在预测变量在 2 周延迟后均方根误差最小。在考虑了各自可能的滞后效应后，症状监测病例数、最高气温、O_3、NO_2、CO、严格性指数、学校活动这些变量有着较小的 RMSE，可能成为更具预测效果的潜在预测变量。通过图 2 显示症状监测病例数、气温、O_3、CO 和严格指数均与报告病例数有较强相关性，其中症状监测病例数和 CO 表现出正相关，O_3 和严格指数表现出较强的负相关。除此之外，多种环境因素之间表现出较强的相关性（特别是气温）。

2.4 *变量筛选* 2.3 小节所示的预测变量在调整滞后效应后被纳入随机森林模型，利用递归特征消除算



注：* $0.01 \leqslant P < 0.05$；* * $0.001 \leqslant P < 0.01$；* * * $P < 0.001$。

图 2　各潜在预测变量与急性呼吸道传染病报告病例数的相关系数

Fig. 2　Correlation coefficients between potential predictive variables and reported cases of acute respiratory infectious diseases

表 1　预测变量各延迟周数下的均方根误差

Table 1　Root mean square error for predictive variables at each lag week

潜在预测变量	各延迟周数下的均方根误差	
	1 周	2 周
症状监测病例数	11.78[a]	12.13
最高气温	12.88[a]	13.29
最低气温	13.14[a]	13.47
最大温差	13.89[a]	14.09
AQI	14.14	14.01[a]
$PM_{2.5}$	13.95	13.91[a]
PM_{10}	14.08	13.99[a]
O_3	12.27[a]	12.82
SO_2	14.10[a]	14.22
NO_2	12.78[a]	12.84
CO	12.30	12.06[a]
严格性指数	12.00	11.81[a]
学校活动	12.34	11.63[a]

注：a 最小均方根误差的延迟周数为最佳延迟周数。

法进行变量筛选，筛选结果由表 2 所示。预测重要性最强的 3 个变量为严格性指数、症状监测病例数和最高气温，而最低气温、O_3 和学校活动 3 个变量由于一定的预测重要性，也被纳入共线性检验。在排除掉最低气温（VIF = 58.93）后，其余潜在预测变量 VIF 均在 1.88 ~ 2.98 之间，将被用于构建 SVM 模型。

表 2　潜在预测变量的预测误差和预测重要性

Table 2　The prediction errors and predictive importance of the potential variables

潜在预测变量	RMSE	R^2	MAE	% Inc MSE(%)
症状监测病例数	7.676	0.735	5.792	49.78
最高气温	7.197	0.741	5.304	33.86
最低气温	8.113	0.712	5.972	30.86
最大温差	8.164	0.716	5.947	
AQI	8.064	0.725	6.066	
$PM_{2.5}$	7.139	0.777	5.275	
PM_{10}	7.184	0.778	5.331	
O_3	7.669	0.744	5.718	23.44
SO_2	7.391	0.757	5.493	
NO_2	7.635	0.743	5.727	

（续表）

潜在预测变量	RMSE	R^2	MAE	% Inc MSE(%)
CO	7.825	0.735	5.799	
严格指数	7.794	0.731	5.706	84.75
学校活动	7.764	0.736	5.716	26.64

注：MAE 为平均绝对误差，% IncMSE 为均方误差增加百分比。

2.5　SVM 模型预测及预测效果评估　最终，严格性指数、症状监测病例数、最高气温、学校活动和 O_3 指数 5 个变量被用于 GA－SVM 模型的建立，设置遗传种群数量为 100，遗传代数为 100。经反复训练获得遗传算法最优适应度曲线，如图 3 所示，最终搜索出最优参数 C = 18.04，γ = 0.175 4。

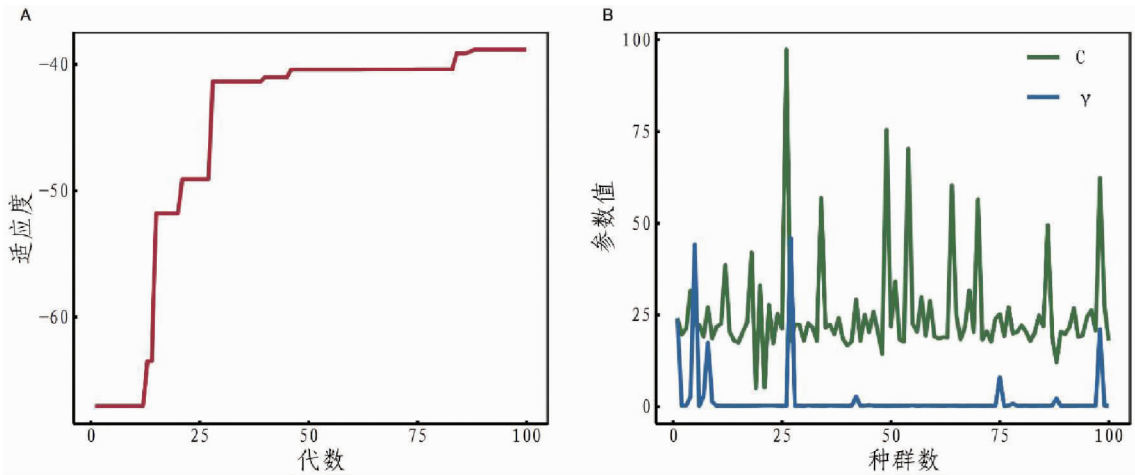


图 3　遗传算法适应度曲线（A）和各种群参数值（B）

Fig. 3　Genetic algorithm fitness curve (A) and population parameters C and γ values (B)

将搜索到的最优参数代入 GA－SVM 模型构建，采取 5－折交叉验证测试预测模型对测试集的预测效果，得到相应的预测结果（图 4），除了 2020 年 12 月的部分高值预测效果较差外，其余周内的累计新增报告病例数均能得到较好预测。表 3 进一步展示了 5－折交叉验证下，每一轮次中对测试集的预测效果。GA－SVM 模型的平均 RMSE 为 6.362，平均 MAPE 为 24.59%，平均 PCC 和 R^2 分别达到了 0.896 和 0.804，说明模型对于急性呼吸道传染病报告病例数有着良好的预测效果。

表 3　5－折交叉验证下的模型预测效果和误差均值

Tabel 3　Model prediction performance and mean error under 5－fold cross－validation

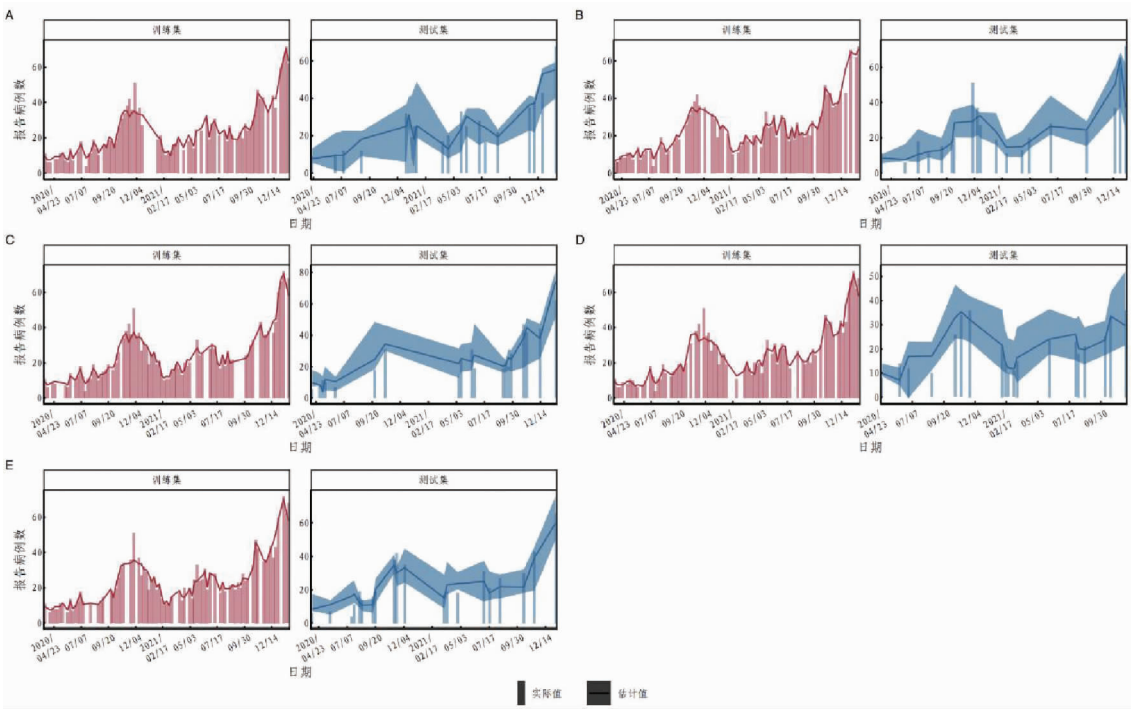
交叉验证轮次	RMSE	MAPE(%)	PCC	R^2
1	5.796	19.17	0.904	0.817
2	10.620	22.42	0.799	0.639
3	5.790	25.92	0.933	0.870
4	3.704	17.27	0.916	0.838
5	5.902	38.17	0.931	0.866
均值	6.362	24.59	0.896	0.804

3　讨论

本研究构建 GA－SVM 回归模型预测上海市徐汇区 2020 年 3 月 15 日—2022 年 1 月 30 日期间的急性呼吸道传染病的流行趋势，预测效果良好，可作为急性呼吸道传染病传统监测预测预警的有效补充。

SVM 回归是一种常见的机器学习算法，由 Vapnik 提出的支持向量机[18]演变而来。SVM 是一种基于统计学习理论的机器学习方法，具有强大的非线性映射能力和处理复杂性的能力，比传统神经网络具有更好的泛化能力，对于时间序列预测中可能存在的水平变化具有较好的容忍性。本研究受客观因素影响训练样本量偏少、数据分布稀疏且波动较大，很多常用的时间序列模型均不适用，而 SVM 恰能在有效解决数据分布不均匀的难点的同时获得不错的预测效果。国内有研究也证实了 SVM 模型在呼吸道疫情（流行性感冒）方面的出色的跟踪预测能力[19]。在一项比较多种时间序列模型在我国传染病监测数据中的应用研究表明 SVM 模型在大多数情况下优于 SVM 模型[20]。ARIMA－SVM 组合模型在肺结核发病趋势预测中也获得了优于单纯 ARIMA 模型的预测效果[21]。本研究中由于监测系统调整和某些重大公共卫生事件的发生，单纯时间序列模型的预测效果很差，无法使用。

症状作为疾病诊断前的发生过程，基于其开展的监测能有效利用从就诊、实验室检测到医生诊断等过程中产生的时间差，更早探察出疫情苗子，前移防控关口，从而实现传染病疫情的早期预警，其敏感性和



注：A ~ E 为 5 - 折交叉验证下模型 1 ~ 模型 5 的训练效果和预测效果。

图 4　5 - 折交叉验证下的模型训练（红色）和预测效果（蓝色）

Fig. 4　Model training (red) and prediction performance (blue) under 5 - fold cross - validation

及时性在传染病监测预警和公共卫生保障中得到应用证实[22-23]。本研究模型在医疗机构症状监测数据的基础上，挖掘利用气象和社会政策及管控措施等多源数据与疾病发生之间的潜在关联，实现了多种数据 SVM 模型的联动，顺应大数据发展的应用前景，为今后的新发传染病分析预警、探索新发传染病传播机制、构建全面的传染病监测管控方面打下基础[24]。

受客观因素影响，建模使用的是 2020 年 3 月至 2022 年 1 月上海市徐汇区的监测数据。一方面，因区级传染病报告数据体量十分有限，难以实现不同急性呼吸道传染病的单独预测。本研究组曾尝试使用 GA - SVM回归模型构建“流行性感冒”单病种的预测模型，获得的预测效果不佳，原因在于本地流感病例报告在 2021 年底前处于低水平，2021 年 11 月起受到流行性感冒的诊断标准修改影响，报告病例数陡增，然而陡增后的训练数据十分有限（仅持续到 2022 年 1 月），也很大程度上直接导致了模型预测的平均绝对百分比误差升至 24.59%。提示使用 GA - SVM 回归模型方法进行预测时，如遇到数据特性发生骤变的情况，需要有足够的后期数据用于模型训练，以保证模型的预测效能。另一方面，此时间段的呼吸道传染病发病情况受到新冠疫情的影响极大，加之人群流动、民众生活习惯改变、人群构成变化等本研究尚未纳入的因素也可能影响着疾病的发生，故在有条件引入更多的多源数据后，应根据不同阶段的实际情况按需对模型进行调整，例如引入稳定后的新监测平台中的新增数据或增加新的参数指标等来进行进一步的训练和测试、调整模型核函数进一步提升预测效果等。下一阶段，通过真实疫情处置验证来寻找合理的预警阈值对于提高模型的实用效能至关重要。

利益冲突声明　本研究不存在任何利益冲突

参考文献

[1]　杨维中，兰亚佳，吕炜，等. 建立我国传染病智慧化预警多点触发机制和多渠道监测预警机制[J]. 中华流行病学杂志，2020，41(11)：1753 - 1757.
Yang WZ, Lan YJ, Lv W, et al. Establishment of multi - point trigger and multi - channel surveillance mechanism for intelligent early warning of infectious diseases in China[J]. Chinese Journal of Epidemiology, 2020, 41(11): 1753 - 1757. (In Chinese)

[2]　杨维中. 传染病预警理论与实践[M]. 北京：人民卫生出版社，2012.
Yang WZ. Early warning of infectious disease theory and practice [M]. Beijing: The People's Health Publishing House, 2012. (In Chinese)

[3]　杨津，冯录召，赖圣杰，等. 急性呼吸道传染病症状监测及预警技术的现状与展望[J]. 中华流行病学杂志，2023，44(1)：60 - 66.
Yang J, Feng LZ, Lai SJ, et al. Syndrome surveillance and early warning technology for acute respiratory infectious diseases: current status and future development [J]. Chinese Journal of Epidemiology, 2023, 44(1): 60 - 66. (In Chinese)

[4]　侯玉梅，张晨阳，苏艳林. 基于支持向量机的缺血性脑卒中患病风险预测[J]. 现代预防医学，2019，46(15)：2692 - 2695，

2700.
Hou YM, Zhang CY, Su YL. Risk prediction of ischemic stroke based on support vector machine[J]. Modern Preventive Medicine, 2019, 46(15):2692 - 2695, 2700. (In Chinese)

[5] 朱碧云,王妮,陈卉,等. 基于实验室指标的新型冠状病毒肺炎鉴别诊断模型[J]. 北京生物医学工程,2022,41(5):483 - 487.
Zhu BY, Wang N, Chen H, et al. Diagnosis model of COVID - 19 based on laboratory indicators[J]. Beijing Biomedical Engineering, 2022, 41(5): 483 - 487. (In Chinese)

[6] 黄嘉,梅鹏程,黄超,等. 基于弹性网络模型大队列老年人认知功能障碍风险因素研究[J]. 中国预防医学杂志,2022,23(12): 936 - 941.
Huang J, Mei PC, Huang C, et al. Risk factors of cognitive impairment in the elderly based on the elastic network model[J]. China Preventive Medicine, 2022, 23(12): 936 - 941. (In Chinese)

[7] 公衍峰,罗卓韦,冯家鑫,等. 基于监督式机器学习模型的上海市小尺度湖北钉螺扩散趋势预测研究[J]. 中国血吸虫病防治杂志,2022,34(3):241 - 251.
Gong YF, Luo ZW, Feng JX, et al. Prediction of trends for fine - scale spread of Oncomelania hupensis in Shang - hai Municipality based on supervised machine learning models[J]. Chinese Journal of Schistosomiasis Control, 2022, 34(3): 241 - 251. (In Chinese)

[8] Oxford University. Oxford COVID - 19 Government Response Tracker (OxCGRT)[EB/OL]. [2024 - 11 - 17]. https://github.com/OxCGRT/covid - policy - dataset.

[9] Hale T, Angrist N, Goldszmidt R, et al. A global panel database of pandemic policies (Oxford COVID - 19 Government Response Tracker)[J]. Nature Human Behaviour, 2021, 5(4): 529 - 538.

[10] 中华人民共和国国家疾病预防控制局. 传染病目录[EB/OL]. [2024 - 11 - 18]. https://www.ndcpa.gov.cn/jbkzzx/c100041/common/list.html.
The State Administration of Disease Control and Prevention of the People's Republic of China. Catalogue of infectious diseases[EB/OL]. [2024 - 11 - 18]. https://www.ndcpa.gov.cn/jbkzzx/c100041/common/list.html. (In Chinese)

[11] 上海市人民政府.《上海市传染病防治管理办法》(沪府令 60 号)[EB/OL]. [2024 - 11 - 17]. https://www.shanghai.gov.cn/nw41492/20200823/0001 - 41492_54567.html.
Shanghai Municipal People's Government. Procedures of Shanghai Municipality for the Prevention and Control of Infectious Diseases (Decree of Shanghai Municipal People's Government No. 60)[EB/OL]. [2024 - 11 - 17]. https://www.shanghai.gov.cn/nw41492/20200823/0001 - 41492_54567.html. (In Chinese)

[12] Kuhn M. 20 recursive feature elimination[EB/OL]. [2024 - 11 - 17]. https://topepo.github.io/caret/recursive - feature - elimination.html.

[13] Kuhn M, Johnson K. Applied predictive modeling[EB/OL]. [2024 - 11 - 17]. http://link.springer.com/10.1007/978 - 1 - 4614 - 6849 - 3.

[14] Yaya SN, Ahinkorah BO, Ameyaw EK, et al. Proximate and socio - economic determinants of under - five mortality in Benin, 2017/2018[J]. BMJ Global Health, 2020, 5(8): e002761.

[15] Krishna CVM, Rao GA, Anuradha S. Analysing the impact of contextual segments on the overall rating in multi - criteria recommender systems[J]. Journal of Big Data, 2023, 10(1): 16.

[16] 王晨晖,刘立申,任佳,等. 主成分分析法和遗传算法优化的支持向量机模型在地震伤亡人数预测中的应用[J]. 地震,2020,40(3):142 - 152.
Wang CH, Liu LS, Ren J, et al. Application of support vector machine model optimized by PCA and genetic algorithm to predicting the number of earthquake casualties[J]. Earthquake, 2020, 40(3): 142 - 152. (In Chinese)

[17] 肖文. 基于机器学习的我国粮食产量预测研究[D]. 兰州:兰州财经大学,2024.
Xiao W. Research on grain yield prediction in China based on machine learning[D]. Lanzhou: Lanzhou University of Finance and Economics, 2024. (In Chinese)

[18] Cai F, Cherkassky V. Generalized SMO algorithm for SVM - based multitask learning[J]. IEEE Transactions on Neural Networks and Learning Systems, 2012, 23(6): 997 - 1003.

[19] 梁凤. 北京、辽宁省 2011 - 2016 年整合百度搜索及传统监测数据的流感支持向量机回归预测模型构建[D]. 沈阳:中国医科大学,2020.
Liang F. The construction of Support Vector Machine Regression Model to forecast influenza epidemic by integrating Baidu search queries and traditional surveillance data in Beijing and Liaoning province from 2011 to 2016[D]. Shenyang: China Medical University, 2020. (In Chinese)

[20] Nsoesie EO, Oladeji O, Abah ASA, et al. Forecasting influenza - like illness trends in Cameroon using Google Search Data[J]. Scientific Reports, 2021, 11(1): 6713.

[21] 杨美涛,王彦丁,李志强,等. ARIMA - SVM 组合模型在肺结核发病趋势预测中的应用[J]. 现代预防医学,2023,50(11):1921 - 1926.
Yang MT, Wang YD, Li ZQ, et al. Application of ARIMA - SVM combination model in predicting the inci - dence trend of pulmonary tuberculosis[J]. Modern Preventive Medicine, 2023, 50(11): 1921 - 1926. (In Chinese)

[22] 陈勇辉,王云辉,王伟,等. 学校因病缺课监测对传染病防控的预警效果[J]. 中国学校卫生,2020,41(3):465 - 467.
Chen YH, Wang YH, Wang W, et al. The early warning effect of school absence monitoring on infectious disease prevention and control due to illness[J]. Chinese Journal of School Health, 2020, 41(3): 465 - 467. (In Chinese)

[23] 黄春萍,宋姝娟,刘牧文,等. G20 杭州峰会期间症状监测机制的建立与应用[J]. 中国公共卫生管理,2020,36(2):204 - 206.
Huang CP, Song SJ, Liu MW, et al. Establishment and application of symptom monitoring mechanism of G20 summit in Hangzhou[J]. Chinese Journal of Public Health Management, 2020, 36(2): 204 - 206. (In Chinese)

[24] 霍添琪,孙晓宇,刘昊,等. 大数据技术在新发传染病管理中的研究进展[J]. 中国数字医学,2021,16(6):91 - 98.
Huo TQ, Sun XY, Liu H, et al. Research progress of big data in the management of emerging infectious diseases[J]. China Digital Medicine, 2021, 16(6): 91 - 98. (In Chinese)

收稿日期:2024-07-11