·论著·一次研究·

基于安徽省医疗卫生机构人员数量构建卫生人力资源最优组合预测模型的策略探讨

王逸凡，韦昕宇，汪国英，沈铮先，尹青青，王　斌

安徽医科大学公共卫生学院（合肥 230032）

【摘要】目的 基于安徽省医疗卫生机构人员数量数据，构建多种卫生人力资源组合预测模型并对比分析，探讨构建最优组合预测模型的策略。**方法** 于《安徽统计年鉴》获取1988—2022年安徽省医疗卫生机构人员数量数据，使用Holt模型、ARIMA模型、NNAR模型、GM（1，1）和随机森林模型构建所有可能的组合模型，分别预测2018—2022年安徽省医疗卫生机构人员数量，对比各模型的均方误差（MSE）、平均绝对误差（MAE）和平均绝对百分比误差（MAPE）3项误差指标，筛选最优组合模型。**结果** 在所有的组合预测模型中，GM（1，1）与随机森林模型构建的组合模型预测误差最小，其MSE、MAE和MAPE分别为146 101 146、10 587.465、2.267，该组合模型为最优组合预测模型。**结论** 选择一定数量的子模型，再通过排列组合的方式生成所有可能的组合模型，并根据误差指标进行筛选，可准确、有效地确定预测误差最小的组合模型。该策略能科学、合理地筛选多种模型，最终得到最优组合预测模型。

【关键词】卫生人力资源；灰色模型；随机森林模型；组合预测模型

【中图分类号】R197.322　　**【文献标识码】**A

Exploring strategies for constructing optimal combined forecasting models for health human resources based on personnel numbers in Anhui Province's healthcare institutions

WANG Yifan, WEI Xinyu, WANG Guoying, SHEN Zhengxian, YIN Qingqing, WANG Bin

School of Public Health, Anhui Medical University, Hefei 230032, China
Corresponding author: WANG Bin, Email: phili_wc@163.com

【Abstract】Objective To construct and compare various combined forecasting models for health human resources based on personnel number in Anhui Province's healthcare institutions, to explore strategies for developing the optimal combined forecasting model. **Methods** Data on the personnel number in healthcare institutions in Anhui Province from 1988 to 2022 was obtained from the *Anhui Statistical Yearbook*. All possible combined models were constructed using the Holt model, ARIMA model, NNAR model, GM (1,1) , and random forest model. The personnel number in Anhui Province's healthcare institutions from 2018 to 2022 was predicted. The optimal combined model was selected by comparing 3 error metrics of mean squared error (MSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). **Results** The combined model of the GM(1,1) and random forest achieved the lowest prediction error among all combined

DOI: 10.12173/j.issn.1004-5511.202602035
基金项目：国家自然科学基金面上项目（81573217）
通信作者：王斌，博士，教授，博士研究生导师，Email：phili_wc@163.com

models, with MSE, MAE and MAPE values were 146,101,146, 10,587.465 and 2.267, which was the optimal combined forecasting model. Conclusion A certain number of individual models are initially selected, and all possible combinations are generated. These combined models are then evaluated using error metrics to effectively identify the one with the smallest prediction error. This strategy can scientifically and rationally screen various models, ultimately yielding the optimal combined forecasting model.

【Keywords】Health human resources; Grey model; Random forest model; Combined forecasting model

卫生人力资源作为医疗卫生体系的核心构成要素，是保障公共卫生服务可及性、提升医疗服务质量的关键支撑。卫生人力资源规划的科学性，核心在于对未来人力需求与供给的精准预判，而预测模型与方法的选择直接决定了规划结果的合理性与实践可操作性。目前，已有多种相对有效的卫生人力资源预测方法，例如卫生人力/人口比值法、灰色模型（grey model，GM）、时间序列模型等[1]。然而，仅运用单一模型进行预测研究，往往存在限制性和不足[2-5]。因此，选择合适的子模型建立组合预测模型，可规避单一模型缺陷、减小误差、提高预测精度[6]。目前，已有多项研究使用组合模型对卫生人力资源进行预测[7-10]，但关于最优子模型组合形式、权重分配方案的选取标准尚未形成统一结论，何种组合方式能构建出精度最高、误差最小的预测模型，仍缺乏系统、深入的研究与论证，一定程度上限制了组合预测模型在卫生人力资源规划工作中的推广应用。

基于此，本研究结合历年安徽省医疗卫生机构人员数量的变化情况，探讨卫生人力资源最优组合预测模型的科学构建方法，明确最优子模型组合方式与权重分配策略，弥补当前卫生人力资源组合预测模型研究的不足，为我国卫生人力资源预测及科学规划提供参考策略。

1 资料与方法

1.1 资料来源

本研究所用的数据资料来源于1989—2023年的《安徽统计年鉴》，提取1988—2022年安徽省医疗卫生机构人员数量数据。

1.2 预测模型构建

选用5种在相关文献中已被验证适用于时间序列数据分析的单一预测模型[1-5]，分别为Holt模型、差分自回归移动平均（auto regressive integrated moving average，ARIMA）模型、神经网络自回归（neural network auto regressive，NNAR）模型、GM（1，1）和随机森林模型。采用RStudio 4.4.1软件建模并构建出所有可能的组合模型。基于1988—2017年安徽省医疗卫生机构人员数据，预测2018—2022年人员数量。

1.2.1 Holt模型

Holt模型由美国统计学家Holt在简单指数平滑法的基础上提出。基于RStudio 4.4.1软件的forecast包holt函数构建模型，初始水平（L_0）值通过序列前2期观测值的均值确定，初始趋势（b_0）值由前2期观测值的线性回归斜率估算得出。模型的水平平滑系数（alpha）、趋势平滑系数（beta）采用函数内置自动寻优机制，以极大似然估计（maximum likelihood estimate，MLE）为优化准则，通过L-BFGS-B数值优化算法求解最优参数，结合观测值逐步更新水平值和趋势值，最后完成目标值预测[11]。

1.2.2 ARIMA模型

ARIMA由美国统计学家Box和英国统计学家Jenkins提出。通过RStudio 4.4.1软件的forecast包auto.arima函数自动确定模型最优阶数，函数依据校正赤池信息准则（Akaike information criterion with correction，AICc）进行最优模型筛选，最终确定自回归阶数（p）、差分次数（d）、移动平均阶数（q），并据此建立模型开展预测[12]。

1.2.3 NNAR模型

NNAR模型结合了时间序列分析中的自回归方法及机器学习中的神经网络模型。构建该模型时，先对原始数据进行差分和归一化处理，使时间序列平稳且适配神经网络训练。该模型属于无分布假设的机器学习模型，对数据正态性无严格要求，因此无需进行正态性检验。采用RStudio 4.4.1软件的forecast包实现自动参数寻优确定关键参数，输入层节点数（p）、季节性滞后项个数

（P）、隐藏层节点数（k）、模型重复训练次数（repeats）、Box-Cox 变换参数（lambda），最终使用训练完成的模型实现指标预测[13]。

1.2.4 GM（1,1）

GM（1,1）是灰色系统理论中最常用的预测模型，由邓聚龙教授于1982年提出。该模型在少量数据或不完全数据条件下表现出色，适用于小样本预测。其核心构成包括原始序列、累加生成序列、灰色微分方程、背景值和模型参数，建模过程简洁高效，主要涵盖数据预处理、累加生成、参数估计、模型检验和预测还原5个关键步骤[14-15]。

该模型构建完成后，还需使用后验差比值（C）和小误差概率（P）对模型的拟合效果进行检验，模型拟合效果判定标准：若$C<0.35$且$P\geqslant0.95$，模型精度等级评为1级，拟合效果好；若$0.35\leqslant C<0.50$或$0.80\leqslant P<0.95$，模型精度等级评为2级，拟合效果合格；若$0.50\leqslant C<0.65$或$0.70\leqslant P<0.80$，模型精度等级评为3级，拟合效果勉强；若$C\geqslant0.65$或$P<0.70$，模型精度等级评为4级，拟合效果不合格[3]。

1.2.5 随机森林模型

随机森林模型是由统计学家Breiman和Cutler在2001年提出的一种基于分类树的算法，其通过对大量分类树的汇总，可提高模型的预测精度，是取代传统机器学习方法的新模型[16]。采用RStudio 4.4.1软件的randomForest包构建随机森林模型前，先对数据进行差分和归一化处理，使时间序列平稳的同时便于后续训练机器学习模型。再通过自相关函数（autocorrelation function，ACF）与模型试算对比确定最优滞后阶数（lag），在此基础上构建滞后数据集并划分为训练集和测试集。然后设置随机森林参数，使用训练集训练模型，并通过测试集对决策树数量（ntree）和节点分裂变量数（mtry）进行参数调优。最后使用训练完成的模型对目标值进行预测，并采用绝对误差和相对误差对模型预测效果进行综合评价，其中相对误差数值越小越好，< 20% 说明预测精度达到要求，< 10% 说明预测精度达到较高要求[17]。

1.2.6 组合模型

使用变权重组合法构建组合模型，即先计算各子模型的预测值与实际值的误差平方和，再用该项指标的倒数计算子模型在组合模型中的权重，使得误差较小的模型获得较大的权重，提高组合模型的预测精度。具体步骤如下：①根据公式1计算误差平方和（sum of squared errors，SSE）；②根据公式2计算权重（w_i）；③根据公式3计算组合模型预测值（$\widehat{y_k}$），其中，y_k是第k个实际值，$\widehat{y_{k,\,i}}$是第i个子模型对于第k个数据点的预测值，n是总的数据量。采用排列组合的方式获得5种子模型构建组合模型的所有可能情况，再运用上述的变权重组合法构建组合模型。

$$SSE_i=\sum_{k=1}^{n}\left(y_k-\widehat{y_{k,\,i}}\right)^2 \qquad \text{公式1}$$

$$w_i=\frac{\dfrac{1}{SSE_i}}{\sum\limits_{j=1}^{m}\dfrac{1}{SSE_j}} \qquad \text{公式2}$$

$$\widehat{y_k}=\sum_{i=1}^{m}w_i\widehat{y_{k,\,i}} \qquad \text{公式3}$$

1.3 组合模型寻优

计算并比较各组合模型的均方误差（mean square error，MSE）、平均绝对误差（mean absolute error，MAE）、平均绝对百分比误差（mean absolute percentage error，MAPE），数值越低，预测精度越高，以筛选出最优组合模型[18]。

2 结果

2.1 子模型预测结果分析

2.1.1 Holt模型

根据holt函数内部寻优结果可知，当alpha=0.8619、beta=0.306时模型的预测效果最好，据此构建Holt模型对2018—2022年安徽省医疗卫生机构人员数量进行预测。结果显示，各年度预测值分别为423 115、438 835、454 555、470 276、485 996人，见表1。各年份预测值的相对误差分别为0.875%、3.470%、9.670%、9.510%、12.200%，5年内仅单一年份相对误差高于10%，预测精度良好，见表2。

2.1.2 ARIMA模型

根据auto.arima函数基于AICc准则的寻优结果可知，当p=0、d=2、q=1时模型的预测效果最好，据此构建ARIMA模型对2018—2022年安徽省医疗卫生机构人员数量进行预测。结果显示，

各年度预测值分别为423 894、440 258、456 622、472 985、489 349人，见表1。各年份预测值的相对误差分别为0.693%、3.160%、9.260%、8.980%、11.600%，5年内仅单一年份相对误差高于10%，预测精度良好，见表2。

2.1.3 NNAR模型

根据数据特征与自动寻优结果，确定NNAR模型最优参数：输入p=2、P=0、k=2、repeats=100、lambda=NULL。其中，k=2由模型自动寻优获得，为适配本研究数据特征的最优取值；repeats=100表示神经网络重复训练集成的次数，用于提升模型稳定性。使用上述参数构建模型对2018—2022年安徽省医疗卫生机构人员数量进行预测，结果显示，各年度预测值分别为421 262、439 485、457 808、472 385、487 667人，见表1。各年份预测值的相对误差分别为1.310%、3.330%、9.030%、9.100%、11.900%，5年内仅单一年份相对误差高于10%，预测精度良好，见表2。

2.1.4 GM（1,1）

以1988—2017年安徽省医疗卫生机构人员数量为原始序列构建GM（1,1），对2018—2022年该项指标进行预测。结果显示，各年度预测值分别为455 795、481 499、501 521、532 155、566 985人，见表1。该模型拟合效果的检验结果显示，P=1，C=0.43，表示模型拟合效果合格。各年份预测值的相对误差分别为6.780%、5.910%、0.339%、2.400%、2.460%，均小于10%，进一步验证了模型具有良好的预测精度，见表2。

2.1.5 随机森林模型

经自相关分析与多组滞后阶数对比试算，确定lag=2，据此创建滞后数据集并划分为训练集和测试集。根据测试集参数调优结果（ntree=100，mtry=2）构建随机森林模型，对2018—2022年安徽省医疗卫生机构人员数量进行预测。结果显示，各年度预测值分别为426 628、448 224、469 913、494 034、519 053人，见表1。各年份预测值的相对误差分别为0.052%、1.410%、6.620%、4.930%、6.200%，均小于10%，预测精度良好，见表2。

2.2 组合模型寻优及预测结果分析

基于最优子集思路，对5种子模型进行排列组合，得到共26种可能的组合模型。使用变权重组合法计算组合模型的权重系数，并预测2018—2022年安徽省医疗卫生机构人员数量，最后分析模型精度，见表3。结果显示，GM（1,1）与随机森林模型构建的组合模型（权重系数为0.61和0.39）的各年度预测值分别为444 443、468 538、489 209、517 306、548 315人。各年份预测值的相对误差分别为4.1%、3.1%、2.8%、0.5%、0.9%，全部小于10%，预测精度良好。在所有的组合预测模型中，GM（1,1）与随机森林模型构建的组合模型预测误差最小，其MSE、MAE和MAPE分别为146 101 146、10 587.465、2.267，该组合模型为最优组合预测模型，见表4。

表1 各子模型的预测结果

Table 1. Prediction results of each sub-model

年份	实际值（人）	预测值（人）				
		Holt模型	ARIMA模型	NNAR模型	GM（1，1）	随机森林模型
2018年	426 851	423 115	423 894	421 262	455 795	426 628
2019年	454 616	438 835	440 258	439 485	481 499	448 224
2020年	503 226	454 555	456 622	457 808	501 521	469 913
2021年	519 677	470 276	472 985	472 385	532 155	494 034
2022年	553 360	485 996	489 349	487 667	566 985	519 053

表2 各子模型预测结果的绝对误差及相对误差

Table 2. Absolute and relative errors of prediction results from each sub-model

年份	Holt模型		ARIMA模型		NNAR模型		GM（1，1）		随机森林模型	
	绝对误差	相对误差	绝对误差	相对误差	绝对误差	相对误差	绝对误差	相对误差	绝对误差	相对误差
2018年	3 736.495	0.008 75	2 957.126	0.006 93	5 588.610	0.013 10	28 943.562	0.067 80	223.239	0.000 52
2019年	15 781.156	0.034 70	14 358.253	0.031 60	15 131.453	0.033 30	26 883.397	0.059 10	6 392.253	0.014 10
2020年	48 670.817	0.096 70	46 604.379	0.092 60	45 418.083	0.090 30	1 704.605	0.003 39	33 312.640	0.066 20
2021年	49 401.478	0.095 10	46 691.505	0.089 80	47 292.178	0.091 00	12 477.815	0.024 00	25 643.426	0.049 30
2022年	67 364.139	0.122 00	64 010.632	0.116 00	65 693.295	0.119 00	13 625.010	0.024 60	34 306.645	0.062 00

表3 组合模型的预测结果、绝对误差及相对误差

Table 3. The prediction results， absolute error， and relative error of the combined model

年份	预测值（人）	绝对误差	相对误差	年份	预测值（人）	绝对误差	相对误差	年份	预测值（人）	绝对误差	相对误差
A_B				A_C				A_D			
2018年	423 524	3 326.639	0.008	2018年	422 152	4 699.380	0.011	2018年	450 389	23 537.962	0.055
2019年	439 583	15 032.877	0.033	2019年	439 173	15 443.386	0.034	2019年	474 442	19 826.264	0.044
2020年	455 642	47 584.115	0.095	2020年	456 246	46 979.772	0.093	2020年	493 753	9 473.274	0.019
2021年	471 701	47 976.353	0.092	2021年	471 372	48 304.886	0.093	2021年	521 919	2 242.377	0.004
2022年	487 759	65 600.591	0.119	2022年	486 865	66 495.494	0.120	2022年	553 589	228.620	< 0.001
A_E				B_C				B_D			
2018年	425 795	1 055.894	0.002	2018年	422 594	4 257.053	0.010	2018年	450 046	23 194.599	0.054
2019年	445 999	8 617.460	0.019	2019年	439 876	14 740.206	0.032	2019年	474 067	19 451.058	0.043
2020年	466 273	36 952.588	0.073	2020年	457 208	46 018.361	0.091	2020年	493 430	9 796.191	0.020
2021年	488 403	31 274.178	0.060	2021年	472 689	46 988.231	0.090	2021年	521 492	1 814.652	0.003
2022年	511 219	42 141.401	0.076	2022年	488 518	64 841.851	0.117	2022年	552 994	366.046	< 0.001
B_E				C_D				C_E			
2018年	425 927	923.795	0.002	2018年	449 693	22 842.077	0.054	2018年	425 277	1 573.671	0.004
2019年	446 182	8 433.533	0.019	2019年	474 076	19 459.798	0.043	2019年	446 024	8 591.857	0.019
2020年	466 507	36 718.635	0.073	2020年	493 798	9 428.334	0.019	2020年	466 866	36 359.508	0.072
2021年	488 640	31 036.976	0.060	2021年	521 594	1 917.062	0.004	2021年	488 585	31 092.288	0.060
2022年	511 442	41 918.263	0.076	2022年	552 970	389.728	0.001	2022年	511 154	42 206.477	0.076
D_E				A_B_C				A_B_D			
2018年	444 434	17 582.570	0.041	2018年	422 757	4 093.971	0.010	2018年	446 282	19 430.524	0.046
2019年	468 538	13 921.935	0.031	2019年	439 550	15 066.317	0.033	2019年	469 143	14 526.762	0.032
2020年	489 209	14 016.501	0.028	2020年	456 377	46 849.328	0.093	2020年	487 996	15 229.575	0.030
2021年	517 306	2 371.092	0.005	2021年	471 933	47 744.259	0.092	2021年	514 333	5 343.666	0.010
2022年	548 315	5 045.226	0.009	2022年	487 728	65 632.039	0.119	2022年	543 630	9 730.159	0.018
A_B_E				A_C_D				A_C_E			
2018年	425 399	1 451.653	0.003	2018年	445 965	19 113.542	0.045	2018年	424 869	1 981.584	0.005
2019年	444 804	9 812.457	0.022	2019年	469 132	14 516.075	0.032	2019年	444 668	9 947.774	0.022
2020年	464 264	38 961.694	0.077	2020年	488 293	14 933.406	0.030	2020年	464 545	38 681.446	0.077
2021年	485 194	34 483.432	0.066	2021年	514 395	5 282.081	0.010	2021年	485 132	34 545.438	0.067
2022年	506 666	46 693.676	0.084	2022年	543 575	9 785.130	0.018	2022年	506 409	46 951.263	0.085
A_D_E				B_C_D				B_C_E			
2018年	442 133	15 281.521	0.036	2018年	445 739	18 888.292	0.044	2018年	424 994	1 857.241	0.004
2019年	465 332	10 715.965	0.024	2019年	468 893	14 277.107	0.031	2019年	444 842	9 773.811	0.022
2020年	485 469	17 756.875	0.035	2020年	488 100	15 125.645	0.030	2020年	464 767	38 459.426	0.076
2021年	512 230	7 447.263	0.014	2021年	514 145	5 532.308	0.011	2021年	485 387	34 289.701	0.066
2022年	541 588	11 771.547	0.021	2022年	543 220	10 139.771	0.018	2022年	506 684	46 675.733	0.084
B_D_E				C_D_E				A_B_C_D			
2018年	442 003	15 152.240	0.036	2018年	441 749	14 898.456	0.035	2018年	442 993	16 141.766	0.038
2019年	465 192	10 575.722	0.024	2019年	465 172	10 556.435	0.023	2019年	465 244	10 628.198	0.023
2020年	485 354	17 872.416	0.036	2020年	485 572	17 654.012	0.035	2020年	484 028	19 197.846	0.038
2021年	512 062	7 615.242	0.015	2021年	512 102	7 574.681	0.015	2021年	508 819	10 857.786	0.021
2022年	541 338	12 022.224	0.022	2022年	541 289	12 070.604	0.022	2022年	536 274	17 086.498	0.031
A_B_C_E				A_B_D_E				A_C_D_E			
2018年	424 701	2 150.355	0.005	2018年	440 182	13 331.483	0.031	2018年	439 949	13 097.589	0.031
2019年	443 905	10 710.797	0.024	2019年	462 651	8 035.082	0.018	2019年	462 627	8 011.190	0.018
2020年	463 174	40 052.132	0.080	2020年	482 385	20 841.191	0.041	2020年	482 575	20 651.453	0.041
2021年	483 030	36 646.736	0.071	2021年	508 034	11 643.176	0.022	2021年	508 060	11 616.791	0.022

续表1

年份	预测值（人）	绝对误差	相对误差	年份	预测值（人）	绝对误差	相对误差	年份	预测值（人）	绝对误差	相对误差
2022年	503 457	49 902.575	0.090	2022年	536 003	17 356.842	0.031	2022年	535 946	17 414.130	0.032
B_C_D_E				A_B_C_D_E							
2018年	439 855	13 004.497	0.031	2018年	438 394	11 543.287	0.027				
2019年	462 530	7 913.709	0.017	2019年	460 462	5 845.542	0.013				
2020年	482 501	20 724.807	0.041	2020年	480 062	23 164.028	0.046				
2021年	507 953	11 723.842	0.023	2021年	504 665	15 012.472	0.029				
2022年	535 780	17 579.936	0.032	2022年	531 435	21 925.267	0.040				

注：A、B、C、D、E分别代表Holt模型、ARIIMA模型、NNAR模型、灰色GM（1，1）模型和随机森林模型，其中，A_B代表Holt模型和ARIIMA模型构建的组合模型，以此类推。

表4 所有预测模型的3项误差指标

Table 4. Three error indicators for all predictive models

模型	MSE	MAE	MAPE
A	1 922 057 592	36 990.817	7.140
B	1 732 865 963	34 924.379	6.733
C	1 775 030 961	35 824.724	6.927
D	380 937 879	16 726.879	3.579
E	597 034 781	19 975.641	3.842
A_B	1 821 293 972	35 904.115	6.926
A_C	1 844 538 817	36 384.584	7.029
A_D	208 387 974	11 061.700	2.446
A_E	838 968 231	24 008.304	4.624
B_C	1 753 089 037	35 369.140	6.829
B_D	203 145 075	10 924.509	2.415
B_E	828 134 125	23 806.240	4.583
C_D	198 632 942	10 807.400	2.389
C_E	829 285 467	23 964.760	4.619
D_E	146 101 146	10 587.465	2.267
A_B_C	1 805 138 580	35 877.183	6.926
A_B_D	188 748 563	12 852.137	2.712
A_B_E	997 162 338	26 280.582	5.063
A_C_D	184 539 940	12 726.047	2.685
A_C_E	999 389 509	26 421.501	5.094
A_D_E	171 538 894	12 594.634	2.605
B_C_D	184 561 980	12 792.625	2.694
B_C_E	986 502 361	26 211.182	5.052
B_D_E	172 677 056	12 647.569	2.613
C_D_E	169 628 350	12 550.838	2.592
A_B_C_D	230 382 483	14 782.419	3.022
A_B_C_E	1 111 353 743	27 892.519	5.378
A_B_D_E	222 693 941	14 241.555	2.882
A_C_D_E	220 082 049	14 158.231	2.863
B_C_D_E	221 552 795	14 189.358	2.868
A_B_C_D_E	282 016 334	15 498.119	3.089

注：A、B、C、D、E分别代表Holt模型、ARIIMA模型、NNAR模型、灰色GM（1，1）模型和随机森林模型，其中，A_B代表Holt模型和ARIIMA模型构建的组合模型，以此类推；MSE. 均方误差；MAE. 平均绝对误差；MAPE. 平均绝对百分比误差。

3 讨论

本研究基于安徽省医疗卫生机构人员数量数据，构建Holt模型、ARIMA模型、NNAR模型、GM（1,1）及随机森林模型5种单一预测模型，并通过排列组合方式构建26种组合预测模型，经多维度误差指标筛选，最终确定GM（1,1）与随机森林模型构建的组合模型为最优预测模型。该组合模型的MSE、MAE、MAPE均低于其他单一模型及组合模型，绝对误差与相对误差控制在较低水平，表明该模型的预测误差较低、预测精度良好。

各单一模型的预测结果显示，不同模型在不同时期的预测性能存在明显差异。采用GM（1,1）预测，2018—2019年预测值的相对误差分别为6.780%、5.910%，明显高于同期随机森林模型（0.052%、1.410%），而2020—2022年预测值相对误差较小，提示GM（1,1）更适于捕捉序列后期平稳增长趋势，对前期波动的拟合效果相对有限；随机森林模型则在预测初期表现出更高的精度，对数据波动的适应性更强。不同模型在时序方面的性能差异，也进一步印证了组合预测的必要性。

相较于单一模型，组合预测能充分整合不同模型的信息优势：GM（1,1）适于捕捉序列整体增长趋势，对小样本、渐进变化数据具有良好适应性[17]；随机森林模型能挖掘历史数据中的非线性关系与潜在规律，对波动特征具有较强捕捉能力[19]。二者组合有效弥补了单一模型在结构假设、数据适应性上的不足，使得短期及中期预测结果更稳健可靠[20]。

本研究仍存在一定局限。首先，子模型类型仍可进一步拓展，后续可纳入支持向量机、长短期记忆网络等模型以丰富模型库；其次，组合方式以简单组合为主，未引入自适应权重、集成学习等优化策略；再次，评价指标以传统误差指标为主，对预测稳定性与鲁棒性的探讨仍可深化；最后，组合模型可能存在过拟合风险，先排列组合所有可能模型，再根据测试集误差选最优，可能导致最终选出的模型在测试集上表现优异，但其泛化能力可能被高估。未来研究可进一步优化模型组合机制，构建更全面的评价体系，提升预测方法的适用性与泛化性能。

综上，在卫生人力资源时间序列预测中，“先选取一定数量具有不同预测特性的子模型，再通过排列组合生成所有潜在组合模型，最后依据科学的误差评价指标进行多维度筛选”的策略具有明显科学性与优越性。该方法不依赖主观赋权与经验判断，通过量化指标全面比较模型性能，可系统降低模型选择偏倚，高效定位精度最优的组合预测方案，为同类卫生资源预测提供了可复制的建模思路。在预测结果的实际意义方面，该组合模型能够较为精准地反映安徽省医疗卫生人员数量的变化规律，可为卫生人力资源规模预判提供较为可靠的数据支撑，为卫生行政部门制定卫生人才发展规划、全面深化卫生人才体制改革、持续推进卫生人才队伍建设提供可靠的决策依据。

伦理声明：不适用

作者贡献：研究设计：王逸凡、王斌；数据采集：王逸凡、韦昕宇、汪国英、沈铮先、尹青青；数据分析、论文撰写：王逸凡；论文审定：王斌、韦昕宇、汪国英、沈铮先、尹青青

数据获取：本研究中使用和（或）分析的数据可在《安徽统计年鉴》获取

利益冲突声明：无

致谢：不适用

参考文献

1 张瑜洁, 王健, 王辛, 等. 基于文献计量法的卫生人力预测模型研究现状分析[J]. 中国医院, 2022, 26(2): 43-46. [Zhang YJ, Wang J, Wang X, et al. Analysis of research status of health workforce prediction model based on bibliometric method[J]. Chinese Hospitals, 2022, 26(2): 43-46.]

2 李慧超, 谢学勤, 邓小虹. 北京市医疗机构卫生专业技术人员需求预测——基于人力人口比值法[J]. 中国卫生政策研究, 2013, 6(4): 56-59. [Li HC, Xie XQ, Deng XH. Forecasting of health professionals in Beijing medical institutions: using a human population ratio method[J]. Chinese Journal of Health Policy, 2013, 6(4): 56-59.]

3 吕邦亮, 汤质如, 彭婧, 等. 基于灰色GM(1,1)模型的安徽省“十四五”期间医疗卫生资源配置预测研究[J]. 卫生软科学, 2024, 38(1): 52-57. [Lyu BL, Tang ZR, Peng J, et al. Prediction of medical and health resources allocation in Anhui Province during the 14th Five-Year Plan period based on grey GM(1, 1) model[J]. Soft Science of Health, 2024, 38(1): 52-57.]

4 林志添, 张健明, 丁海峰. 基于ARIMA模型的我国长三角地区卫生人力资源需求预测分析[J]. 中国医疗管理科学, 2021, 11(3): 5-11. [Lin ZT, Zhang JM, Ding HF. Predictive analysis of healthcare human resources demand in the Yangtze River Delta Region based on the ARIMA model[J]. Chinese Journal of Medical Management Sciences, 2021,11(3): 5-11.]

5 石丛, 王健. 基于BP神经网络和时间序列的我国卫生人力资源研究[J]. 中国初级卫生保健, 2013, 27(11): 22-24. [Shi C, Wang J. Study of health human resource in China based on the model of BP-

ANN and Arima[J]. Chinese Primary Health Care, 2013, 27(11): 22-24.]
6 谢俏丽. 基于组合预测模型的湖北省卫生人力资源需求预测研究[D]. 武汉:华中科技大学, 2016. [Xie QL. Research on the demand forecast of health human resources in Hubei province based on combination forecasting model[D]. Wuhan: Huazhong University of Science and Technology, 2016.]
7 朱泉同, 高山. 基于组合预测模型的江苏省卫生人力资源需求预测探讨[J]. 中国卫生统计, 2020, 37(6): 862-865. [Zhu QT, Gao S. Discussion on the forecast of health human resource demand in Jiangsu province based on combination forecasting model[J]. Chinese Journal of Health Statistics, 2020, 37(6): 862-865.]
8 侯雅楠, 王丹, 陈芸, 等. 山东省卫生人力资源组合预测模型构建及应用[J]. 卫生软科学, 2021, 35(4): 72-75, 79. [Hou YN, Wang D, Chen Y, et al. Construction and application of combination forecasting model of health human resources in Shandong province[J]. Soft Science of Health, 2021, 35(4): 72-75, 79.]
9 徐瑞璞, 钱国宏, 路杰, 等. 基于组合模型预测短期甘肃省医护人才需求[J]. 中国卫生统计, 2024, 41(2): 287-290. [Xu RP, Qian GH, Lu J, et al. Predicting short-term demand for medical and nursing talents in Gansu province based on a combined model[J]. Chinese Journal of Health Statistics, 2024, 41(2): 287-290.]
10 黄锐. 基于最优加权组合模型的重庆市卫生人力资源需求预测研究[D]. 重庆: 重庆医科大学, 2022. [Huang R. Research on the forecast of health human resource demand in Chongqing based on the optimal weighted combination model[D]. Chongqing: Chongqing Medical University, 2022.]
11 苗开超. 基于指数平滑模型的农产品价格预测研究[D]. 合肥: 合肥工业大学, 2009. [Miao KC. Research on agricultural product price forecasting based on exponential smoothing model[D]. Hefei: Hefei University of Technology, 2009.]
12 李志超, 刘升. 基于ARIMA模型、灰色模型和回归模型的预测比较[J]. 统计与决策, 2019, 35(23): 38-41. [Li ZC, Liu S. Prediction comparison based on ARIMA model, grey model, and regression model[J]. Statistics & Decision, 2019, 35(23): 38-41.]
13 张欣, 刘振球, 袁黄波, 等. 神经网络自回归模型在丙肝发病趋势和预测研究中的应用[J]. 中国卫生统计, 2020, 37(4): 524-526. [Zhang X, Liu ZQ, Yuan HB, et al. Application of neural network autoregressive model in the study of hepatitis C incidence trend and prediction[J]. Chinese Journal of Health Statistics, 2020, 37(4): 524-526.]
14 陈嘉琳. 基于灰色GM(1,1)模型的广东省卫生总费用预测分析[J]. 中国医疗管理科学, 2021, 11(5): 5-11. [Chen JL. Prediction of total expenditure on health in Guangdong province based on grey GM (1,1) model[J]. Chinese Journal of Medical Management Sciences, 2021, 11(5): 5-11.]
15 毕慧, 马丹华, 许桂丽, 等. 基于年龄-时期-队列模型的中国物质使用障碍疾病负担及预测研究[J]. 药物流行病学杂志, 2024, 33(7):760-769.[Bi H, Ma DH, Xu GL, et al. Study on the disease burden and prediction of substance use disorder in China based on age-period-cohort model[J]. Chinese Journal of Pharmacoepidemiology, 2024, 33(7): 760-769.]
16 李欣海. 随机森林模型在分类与回归分析中的应用[J]. 应用昆虫学报, 2013, 50(4): 1190-1197. [Li XH. Using "random forest" for classification and regression[J]. Chinese Journal of Applied Entomology, 2013, 50(4): 1190-1197.]
17 林小龙, 张杰, 林伟. 2022—2026年我国鼻咽癌发病率与死亡率的预测: 基于GM(1,1)和ARIMA模型[J]. 医学新知, 2025, 35(9): 1017-1023. [Lin XL, Zhang J, Lin W. Prediction of incidence and mortality rates of nasopharyngeal carcinoma in China from 2022 to 2026: based on GM(1,1) and ARIMA models[J]. Yixue Xinzhi Zazhi, 2025, 35(9): 1017-1023.]
18 吴国平, 袁有树, 王志伟. 河南省中医类医院卫生人力资源需求预测组合模型的构建[J]. 郑州大学学报(医学版), 2025, 60(6): 808-812. [Wu GP, Yuan YS, Wang ZW. Construction of a combination model for predicting human resource demand in traditional Chinese medicine hospitals in Henan Province[J]. Journal of Zhengzhou University (Medical Sciences), 2025, 60(6): 808-812.]
19 刘吉莉, 王凤美, 刘阳, 等. 基于机器学习方法构建幽门螺杆菌感染的风险预测模型[J]. 药学前沿, 2025, 29(2): 265-276. [Liu JL, Wang FM, Liu Y, et al. Construction of the risk prediction model of Helicobacter pylori infection based on machine learning method[J]. Frontiers in Pharmaceutical Sciences, 2025, 29(2): 265-276.]
20 刘鸿宇, 孙玉凤, 王健. 卫生人力资源需求预测研究进展、问题探讨及展望[J]. 中国卫生事业管理, 2016, 33(11): 828-830, 860. [Liu HY, Sun YF, Wang J. Progress, problems and trend of projection of HRH demand research[J]. Chinese Health Service Management, 2016, 33(11): 828-830, 860.]

收稿日期：2026 年 02 月 08 日 修回日期：2026 年 04 月 02 日
本文编辑：杨 燕 曹 越

引用本文：王逸凡, 韦昕宇, 汪国英, 等. 基于安徽省医疗卫生机构人员数量构建卫生人力资源最优组合预测模型的策略探讨[J]. 医学新知, 2026, 36(6): 657-664. DOI: 10.12173/j.issn.1004-5511.202602035.
Wang YF, Wei XY, Wang GY, et al. Exploring strategies for constructing optimal combined forecasting models for health human resources based on personnel numbers in Anhui Province's healthcare institutions[J]. Yixue Xinzhi Zazhi, 2026, 36(6): 657-664. DOI: 10.12173/j.issn.1004-5511.202602035.