

DOI:10.13288/j.11-2166/r.2025.11.007

基于机器学习 Voting 集成算法的慢性咳嗽 中医证候诊断模型构建

白逸晨¹, 秦苏杨¹, 周崇云¹, 史利卿², 季坤², 张楚楚¹, 李盼飞³, 崔唐明¹, 李海燕¹✉

1. 中国中医科学院中医药信息研究所, 北京市东城区东直门内南小街 16 号, 100700; 2. 北京中医药大学东方医院; 3. 浙江中医药大学医学技术与信息工程学院

[摘要] **目的** 探索慢性咳嗽中医证候诊断机器学习模型的构建及采用 Voting 集成算法进行优化的方法。**方法** 回顾性收集北京中医药大学东方医院呼吸科 921 例慢性咳嗽患者的病例资料, 通过标准化处理提取 84 项临床特征, 进行中医证候类型判定。筛选例数 > 50 的证候类型所属病例数据形成慢性咳嗽中医证候诊断专病数据集。采用合成少数类过采样技术 (SMOTE) 平衡数据后, 构建 Logistic 回归 (LR)、决策树 (DT)、多层感知机 (MLP) 和引导聚集 (Bagging) 4 种基础模型, 通过硬投票方式融合为 Voting 集成算法模型, 并运用准确率、召回率、精确率、F1 分数、受试者工作特征 (ROC) 曲线、ROC 曲线下面积 (AUC) 及混淆矩阵评价模型性能。**结果** 921 例慢性咳嗽患者例数 > 50 的证型为湿热郁肺证 (294 例)、风邪伏肺证 (103 例)、寒饮伏肺证 (102 例)、痰热郁肺证 (64 例)、肺阳亏虚证 (54 例)、痰湿阻肺证 (53 例) 6 种证候类型, 共计 670 例, 故为专病数据集。6 种证候类型的患者高频症状可见咳嗽、咳痰、异味诱咳、咽痒、咽痒则咳、冷风诱咳等。构建的 4 种基础模型中, MLP 模型的中医证候诊断效能最佳 (测试集中准确率 0.9104, AUC 0.9828); 与 4 种基础模型相比, Voting 集成算法模型性能表现最优, 在训练集和测试中准确率分别为 0.9289 和 0.9253, 过拟合差异为 0.0036, 测试集中 AUC 值为 0.9836, 较所有基础模型的准确率和 AUC 均有所改善, 且对湿热郁肺证 (AUC 0.9984) 和风邪伏肺证 (AUC 0.9970) 诊断效果更优。**结论** Voting 集成算法有效整合多种机器学习优势, 集成后的慢性咳嗽中医证候诊断模型效能得到了进一步优化, 具有较高的准确性和更强的泛化能力。

[关键词] 慢性咳嗽; 机器学习; 证候; 诊断模型; Voting 集成算法

慢性咳嗽作为呼吸科最常见的病症之一, 归属于中医学“久咳”“顽咳”等范畴。中医证候学研究^[1-2]表明, 慢性咳嗽常见风邪伏肺证、湿热郁肺证、痰湿阻肺证、寒饮伏肺证、肺阳亏虚证等证候, 临床常见多种证候相兼为患。辨证论治作为中医学基本特点之一, 辨证的准确性直接影响着临床疗效, 但传统辨证方法存在主观性较强的问题。近年来, 机器学习技术为中医智能辨证提供了新思路^[3-4], 证候诊断属于机器学习分类任务。Voting

集成算法可实现将多个分类器的诊断结果进行汇总融合, 从而提高模型的诊断性能^[5-7]。本研究选取 Logistic 回归 (logistic regression, LR)、决策树 (decision tree, DT)、多层感知机 (multilayer perceptron, MLP) 和引导聚集 (bootstrap aggregating, Bagging) 4 种机器学习算法构建证候诊断基础模型, 运用 Voting 集成算法进行模型融合, 初步建立慢性咳嗽中医证候诊断的集成模型, 以期慢性咳嗽中医证候诊断提供一种客观化、可量化的辨证工具, 并为中医智能辨证研究提供方法学参考。本研究获得北京中医药大学东方医院伦理委员会批准 (批号: 2024-04)。

基金项目: 国家中医药管理局中医药数字化基础性标准研究项目 (GHC-2023-ZFGM-018-03); 中国中医科学院科技创新工程 (CI2021B002); 北京中医药大学东方医院医疗机构制剂与新药研发项目 (2023-DYXY-005)

✉ 通讯作者: lihy@mail.cintcm.ac.cn

1 资料

1.1 数据来源

本研究回顾性收集 2009 年 1 月 1 日至 2022 年 12 月 31 日就诊于北京中医药大学东方医院呼吸科诊断为慢性咳嗽的患者资料,共纳入患者 921 例,男性 298 例,女性 623 例,男女比例为 1:2.09,平均年龄 (49.99±15.22) 岁,中位病程 24 (6,72) 个月,存在药物过敏者 88 例 (9.55%),存在既往病史者 220 例 (23.89%)。

1.2 诊断及辨证标准

西医诊断标准参照中华医学会呼吸病学分会哮喘学组《咳嗽的诊断与治疗指南 (2021)》^[8] 中慢性咳嗽的诊断标准制定:以咳嗽为唯一或者主要症状,咳嗽病程>8 周,且胸部 X 线检查无明显异常。

中医证候判断标准根据《咳嗽中医诊疗专家共识意见 (2021)》^[9]、《咳嗽的诊断与治疗指南 (2021)》^[8] 的中医部分、《周仲瑛实用中医内科学》^[10] 的咳嗽病部分、《中医临床诊疗术语 第 2 部分:证候》^[11] 及课题组慢性咳嗽中医证候学研究结果^[2,12-15],拟定慢性咳嗽的风邪伏肺证、湿热郁肺证、寒饮伏肺证、痰热郁肺证、痰湿阻肺证、肺热津伤证、肺胃热盛证、肺阳亏虚证、肺脾阳虚证、肺阴亏虚证、肝气犯肺证等 20 个主要证候类型诊断标准,并按照文献^[11] 中的证候诊断信息拟定病例中涉及的次要证候类型诊断标准。主要证候类型与次要证候类型同时存在的病例则判定为复合证候类型,主要证候和次要证候以外的证候类型以及病例数≤2 例的证候类型均归入其他证候类型进行合并统计。具体见附表 (扫描本文二维码获取)。

1.3 纳入标准

1) 符合西医慢性咳嗽诊断标准及中医证候类型辨证标准;2) 病例记录完整。

1.4 排除标准

1) 合并有严重的原发性心、肝、肺、肾、血液疾病或影响患者生存的严重疾病;2) 妊娠或哺乳期妇女。

2 方法

2.1 中医四诊信息采集及规范

根据慢性咳嗽西医诊断标准、中医证候判断标准及临床流行病学调查要求,设计慢性咳嗽中医四诊信息采集表,主要内容包括患者的一般资料、既往史、既往诊疗情况、症状、体征、舌脉信息等。将纳入研究的病例资料的证候名称和临床特征中的

症状、舌象、脉象名称参照《中医临床常见症状术语规范》^[16]、《中医症状鉴别诊断学》^[17]、《中医诊断学》^[18] 中具有相同语义的名称进行表述,实现证候名称和临床特征中症状、舌象、脉象名称的标准化、规范化处理。本研究共提取临床特征 84 个 (人口学特征 2 个,临床信息 4 个,症状 55 个,舌象 16 个,脉象 7 个),全部纳入诊断模型的构建中。入选病例资料的审核及每个病例的中医证候判别均由 2 名工作年限≥3 年且具有中级及以上职称的中医师联合完成。

2.2 数据录入及数据集准备

将纳入的慢性咳嗽患者的中医症状、体征及舌脉信息分为“0”(无)、“1”(有),采用双人双机独立录入的方法,运用 Epidata 3.1 软件依据慢性咳嗽中医四诊信息采集表构建的电子数据库,录入数据后进行数据审核和校对。在数据采集、录入前,对参与人员进行统一培训,保证病例资料数据采集的一致性和可靠性。对于证候类型信息的处理,则将纳入病例按证候类型进行整体排序,筛选例数>50 的证候类型所属病例数据,形成慢性咳嗽中医证候诊断专病数据集。

2.3 中医证候诊断机器学习模型构建与评价

使用 Python 3.12 软件进行专病数据集数据平衡处理、基础模型训练与测试、集成模型训练与测试、模型评价,从而完成慢性咳嗽中医证候诊断模型的构建与评价。

2.3.1 样本拆分与数据平衡处理 使用 Python.sklearn.model_selection 的 train_test_split 功能将专病数据集按照 7:3 的比例分为训练集 (469 例) 和测试集 (201 例),同时保证每种中医证候类型的在训练集和测试集中的比例均为 7:3。由于专病数据集中的各证候类型样本量差异较大,存在数据不平衡现象,容易出现模型性能下降、泛化能力差、收敛速度慢等情况,因此采用合成少数类过采样技术 (SMOTE) 对训练集进行过采样处理以提升模型对少数类样本的识别能力。为保持训练集总样本量基本不变,对训练集中样本量高于均值的证候类型数据采用欠采样,随机删除多余数据,保留总样本量的均值数量,对病例数少于均值的证候类型则采用 SMOTE 技术进行过采样处理,补足病例数至总样本量的均值数量,从而提升模型拟合度^[19]。各基础模型和 Voting 集成算法模型均在同一数据平衡处理后的训练集中进行训练,在相同测试集中进行模型测试。

2.3.2 基础模型训练与测试 在相同的训练集中分别使用 LR、DT、MLP、Bagging 4 种基础模型机器学习算法进行独立的中医证候诊断任务的训练, 后在相同的测试集中分别对 4 种基础模型进行测试, 使用模型评价指标对各基础模型在训练集与测试集中的效能进行评价。

2.3.3 集成模型训练与测试 使用融合了 4 种基础模型的 Voting 集成算法, 在与 4 个基础模型训练使用的相同训练集中, 采用硬投票的方式对 4 种基础模型的预测结果进行多数投票, 生成最终预测结果, 完成 Voting 集成模型的训练。后使用与基础模型相同的测试集进行 Voting 集成模型的测试, 使用模型评价指标对 Voting 集成模型在训练集与测试集中的效能进行评价。

2.3.4 模型评价指标 使用准确率、召回率、精确率、F1 分数、受试者工作特征 (ROC) 曲线、ROC 曲线下面积 (AUC) 及混淆矩阵来评价 4 种基础模型及 Voting 集成模型的性能。在模型评价的混淆矩阵中, 真正例是模型将临床诊断为某一证候的 (正例) 数据预测为该证候 (正例) 的数量; 假正例是模型将临床诊断为非某一证候 (负例) 的数据预测为该证候 (正例) 的数量; 真负例是模型将临床诊断为非某一证候 (负例) 的数据预测为非此证候 (负例) 的数量; 假负例是模型将临床诊断为某一证候 (正例) 的数据预测为非此证候 (负例) 的数量。准确率是模型性能的基本指标, 召回率、精确率、F1 分数则是模型不同维度性能的评价。准确率用于衡量模型整体的预测性能; 召回率用于衡量模型对正例样本的覆盖能力, 即避免漏检的能力; 精确率用于衡量模型预测正例数据的可靠性, 即避免假阳性的能力; F1 分数可平衡召回率、精确率, 用于兼顾衡量模型避免漏检与避免假阳性的性能^[20]23-45。ROC 曲线以假阳性率为 x 轴, 以真阳性率为 y 轴, 可展示不同阈值下模型整体的分类性能; ROC 曲线与 AUC 值对类别分布不敏感, 用于评价数据分布不均条件下的模型性能, 同时可对模型在各类样本中的预测能力进行更均衡的量化评估; 混淆矩阵则能够反映模型对不同类别预测结果的具体信息^[3]。

2.4 统计学方法

使用 SPSS 25.0 软件进行描述统计, 正态分布的计量资料以均值±标准差 ($\bar{x}\pm s$) 表示, 非正态分布的计量资料以中位数 (下四分位数, 上四分位数) [$M(P_{25}, P_{75})$] 表示, 计数资料以例 (%) 表示。

3 结果

3.1 921 例患者中医证候及临床特点

表 1 示, 921 例慢性咳嗽患者例数>50 的证型为湿热郁肺证、风邪伏肺证、寒饮伏肺证、痰热郁肺证、肺阳亏虚证、痰湿阻肺证 6 种证候类型, 共计 670 例, 故为专病数据集。专病数据集中患者平均年龄 (49.29±15.10) 岁, 中位病程 24 (6, 72) 个月。表 2 示, 患者临床特征可见女性患者占比较高, 为 66.12%; 有药物过敏的患者 56 例, 有既往病史的患者 147 例, 冬季发病者最多, 高频症状可见咳嗽、咳痰、异味诱咳、咽痒、咽痒则咳、冷风诱咳、口干、咳白痰等, 常见舌象为黄苔、腻苔、白苔、红舌、淡红舌, 脉象表现则以滑脉、弦滑脉、细脉为主。

表 1 921 例慢性咳嗽患者中医证候类型分布情况
Table 1 Distribution of TCM syndrome types among 921 patients with chronic cough

证候类型	例数 (%)	证候类型	例数 (%)
湿热郁肺证	294 (31.92)	肺气亏虚证	8 (0.87)
风邪伏肺证	103 (11.18)	胃气上逆证	6 (0.65)
寒饮伏肺证	102 (11.07)	肝火犯肺证	5 (0.54)
痰热郁肺证	64 (6.95)	肺肾亏虚证	5 (0.54)
肺阳亏虚证	54 (5.86)	瘀阻肺络证	4 (0.43)
痰湿阻肺证	53 (5.75)	肺胃热盛证	4 (0.43)
肺热津伤证	40 (4.34)	肺脾气虚证	3 (0.33)
肺脾阳虚证	24 (2.61)	复合证型	83 (9.01)
肺阴亏虚证	19 (2.06)	其他证型	39 (4.23)
肝气犯肺证	11 (1.19)		

3.2 慢性咳嗽中医证候诊断模型构建

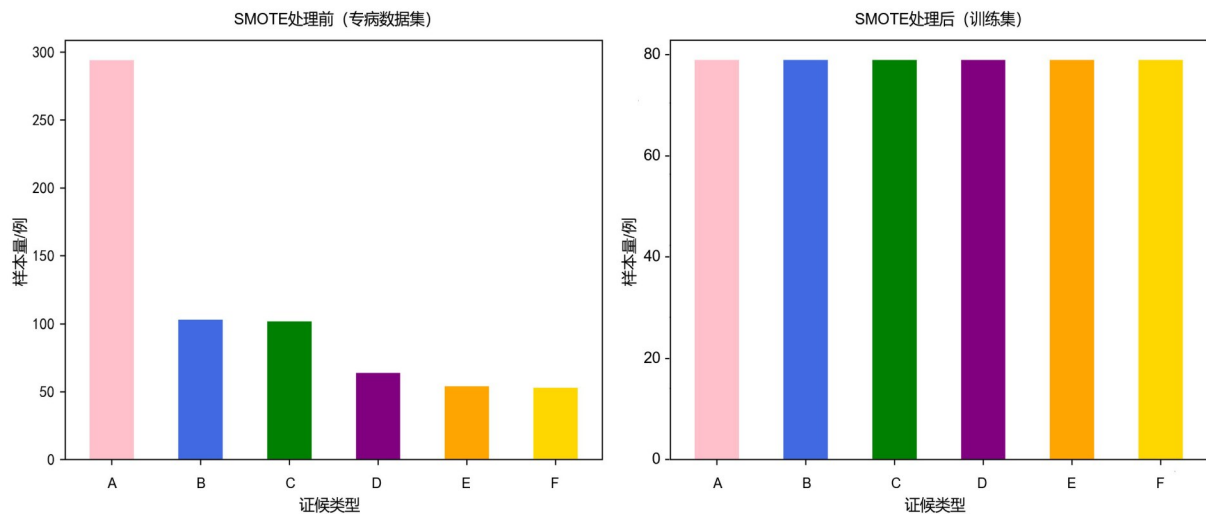
3.2.1 数据平衡处理结果 图 1 示, 专病数据集原始数据中湿热郁肺证、风邪伏肺证、寒饮伏肺证、痰热郁肺证、肺阳亏虚证、痰湿阻肺证 6 种证候的分布存在明显不平衡的问题。为提升各模型在训练中的泛化能力, 在保持训练集总样本量基本不变的情况下, 对湿热郁肺证病例数据欠采样处理, 保留训练集总样本量的均值 79 例, 对其余 5 种证候数据进行 SMOTE 处理。将各证候样本量补足 79 例, 处理后训练集的数据分布趋于平衡。

3.2.2 中医证候诊断模型的整体诊断效能 表 3、图 2 示, 在训练集和测试集中, 4 种基础模型的准确率、召回率、精准率及 F1 分数均在 0.69 以上, 4 种基础模型 ROC 曲线的 AUC 值均>0.9。其中, MLP 模型的准确率、精准率、召回率及 F1 分数最高, 对于证候的预测有着较好的诊断价值。在测试

表2 670例专病数据集中慢性咳嗽患者临床特征分布情况(≥20例)

Table 2 Distribution of clinical characteristics among patients with chronic cough in the specialized dataset of 670 cases

临床特征	例数 (%)	临床特征	例数 (%)	临床特征	例数 (%)
一般资料		症状		症状	
性别(女)	443(66.12)	口苦	137(20.45)	咽痛	36(5.37)
临床信息		口黏	122(18.21)	鼻塞流涕	33(4.93)
药物过敏	56(8.36)	口渴欲饮	120(17.91)	咳黄痰	28(4.18)
既往病史	147(21.94)	受凉则咳	112(16.72)	烧心	27(4.03)
发病季节(冬)	259(38.66)	饮不解渴	100(14.93)	纳差	22(3.28)
发病季节(春)	198(29.55)	大便黏	96(14.33)	气管痒	21(3.13)
发病季节(夏)	124(18.51)	咳黏痰	92(13.73)	舌象	
发病季节(秋)	89(13.28)	反酸	90(13.43)	黄苔	357(53.28)
症状		胃怕凉	86(12.84)	腻苔	291(43.43)
咳嗽	670(100.00)	咽干	83(12.39)	白苔	266(39.70)
咳痰	417(62.24)	全身怕冷	77(11.49)	红舌	263(39.25)
异味诱咳	393(58.66)	四肢怕凉	64(9.55)	淡红舌	262(39.10)
咽痒	347(51.79)	身怕热	64(9.55)	厚苔	92(13.73)
咽痒则咳嗽	335(50.00)	饮食不当诱咳	61(9.10)	舌尖红	51(7.61)
吸冷风咳嗽	325(48.51)	大便干	55(8.21)	舌边尖红	43(6.42)
口干	309(46.12)	便溏	55(8.21)	胖大舌	21(3.13)
咳白痰	293(43.73)	咳稀痰	53(7.91)	脉象	
季节因素诱咳	154(22.99)	咳则遗尿	49(7.31)	滑脉	247(36.87)
后背凉	151(22.54)	自汗	40(5.97)	弦滑脉	230(34.33)
感冒后易咳	148(22.09)	尿黄	40(5.97)	细脉	138(20.60)



注：A，湿热郁肺证；B，风邪伏肺证；C，寒饮伏肺证；D，痰热郁肺证；E，肺阳亏虚证；F，痰湿阻肺证；SMOTE，合成少数类过采样技术。

图1 慢性咳嗽患者SMOTE处理前专病数据集与SMOTE处理后训练集6种中医证候类型分布情况

Figure 1 Distribution of six TCM syndrome types in the specialized dataset and the SMOTE-processed training set of chronic cough patients

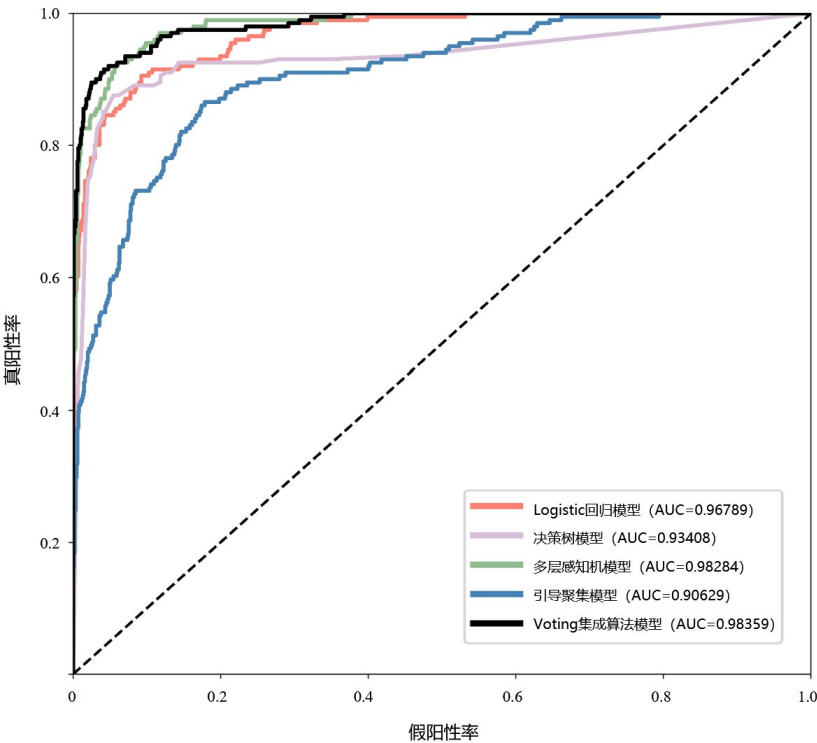
集中，Voting集成算法模型的准确率、精准率、召回率、F1分数及AUC值分别为0.9253、0.8722、0.8853、0.8772和0.9836，与4种基础模型在测试集中的各评价指标相比，均有不同程度的提升，使诊断模型的诊断效能得以优化。此外，通过对比各

模型在训练集和测试集中的各评价指标的表现，可见Voting集成算法模型在测试集和训练集中的准确率和AUC值各相差0.0036和0.0071，与4个基础模型相比，差值更小，显示了Voting集成算法模型具有更强的泛化能力。

表 3 慢性咳嗽患者临床特征预测中医证候的 4 种基础模型及 Voting 集成模型在训练集和测试集中的效能比较
Table 3 Performance comparison of four base models and the Voting ensemble model in predicting TCM syndromes from clinical features of chronic cough patients on training and test sets

模型类型	准确率		精准率		召回率		F1 分数		AUC 值	
	训练集	测试集	训练集	测试集	训练集	测试集	训练集	测试集	训练集	测试集
LR	0.9015	0.8955	0.8451	0.8337	0.8676	0.8375	0.8433	0.8269	0.9802	0.9679
DT	0.7886	0.7761	0.7343	0.7256	0.7069	0.6972	0.7124	0.6930	0.9421	0.9341
MLP	0.9141	0.9104	0.8829	0.8631	0.8854	0.8703	0.8725	0.8613	0.9903	0.9828
Bagging	0.8601	0.8507	0.8004	0.7838	0.8597	0.8487	0.7902	0.7739	0.9197	0.9063
Voting	0.9289	0.9253	0.8807	0.8722	0.8921	0.8853	0.8812	0.8772	0.9907	0.9836

注：LR，Logistic 回归模型；DT，决策树模型；MLP，多层感知机模型；Bagging，引导聚集模型；Voting，Voting 集成算法模型；AUC，受试者工作特征曲线下面积。



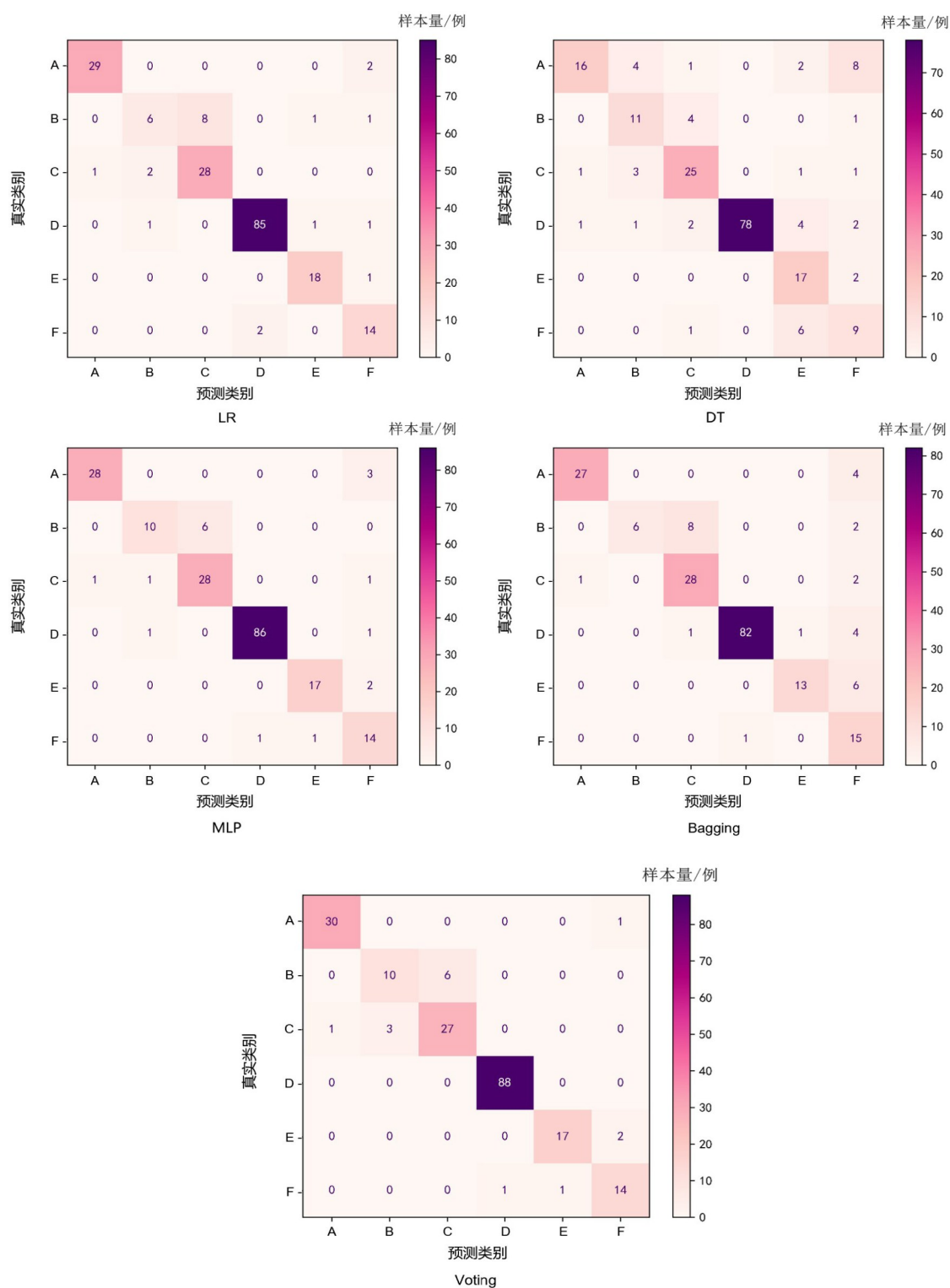
注：ROC，受试者工作特征；AUC，ROC 曲线下面积。

图 2 慢性咳嗽测试集患者临床特征预测中医证候的 4 种基础模型及 Voting 集成模型的 ROC 曲线

Figure 2 ROC curves of four base models and the Voting ensemble model for predicting TCM syndromes from clinical features of chronic cough patients in the test set

3.2.3 中医证候诊断模型对各证候的诊断能力
图 3 示，在 4 种基础模型对各证候的预测结果中，与其他基础模型相比，LR 模型对风邪伏肺证（29 例）、寒饮伏肺证（28 例）和痰热郁肺证（18 例）预测正确的例数分别为更高，而对肺阳亏虚证（6 例）预测正确的例数最低；DT 模型对肺阳亏虚证（11 例）的预测有着更好的效果，而对痰湿阻肺证（9 例）的预测效果较差；MLP 模型对湿热郁肺证（86 例）和寒饮伏肺证（28 例）的预测具有更好性能表现；Bagging 模型则对寒饮伏肺证（28 例）和

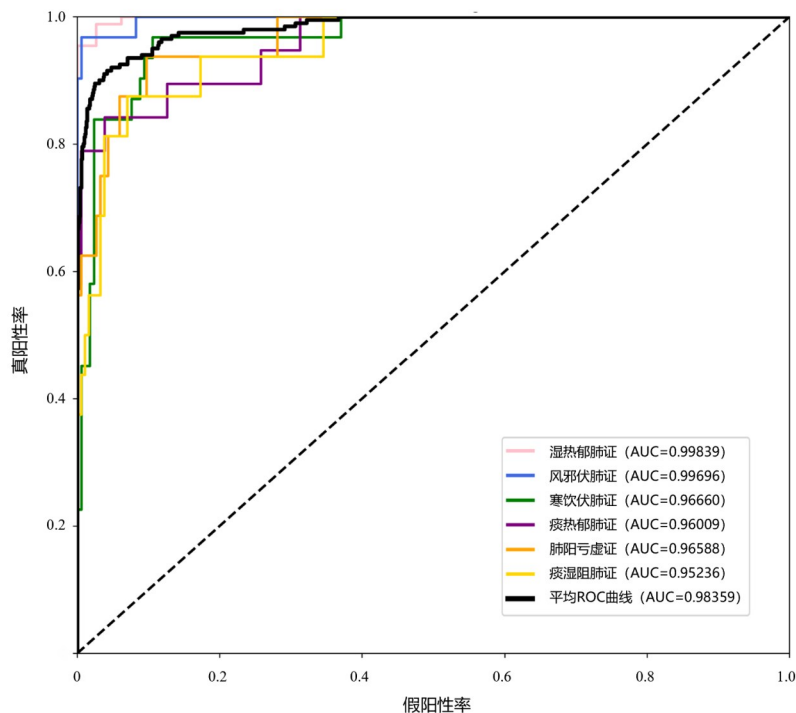
痰湿阻肺证（15 例）的预测效果更佳，而对肺阳亏虚证（6 例）和痰热郁肺证（13 例）的预测效果较差。与 4 个基础模型相比，Voting 集成算法模型对每个证候的预测正确的例数均处于较高水平，对风邪伏肺证、肺阳亏虚证、寒饮伏肺证、湿热郁肺证、痰热郁肺证和痰湿阻肺证 6 个证候预测正确的例数分别为 30 例、10 例、27 例、88 例、17 例、14 例；与基础模型中最高的预测正确例数相比，Voting 集成算法模型对风邪伏肺证、湿热郁肺证预测正确的例数分别提高了 1 例和 2 例。图 4 示，Voting 集成模



注：LR，Logistic 回归模型；DT，决策树模型；MLP，多层感知机模型；Bagging，引导聚集模型；Voting，Voting 集成算法模型；A，湿热郁肺证；B，风邪郁肺证；C，寒饮伏肺证；D，痰热郁肺证；E，肺阳亏虚证；F，痰湿阻肺证；图中单元格内的数字表示对应类别的样本例数，单元格的颜色越深表明样本例数越大。

图3 慢性咳嗽测试集患者临床特征预测中医证候的4种基础模型及Voting集成模型的混淆矩阵图

Figure 3 Confusion matrices of four base models and the Voting ensemble model for predicting TCM syndromes based on clinical features of chronic cough patients in the test set



注：ROC，受试者工作特征；AUC，ROC 曲线下面积。

图4 慢性咳嗽测试集患者临床特征预测6种中医证候类型的
Voting集成模型ROC曲线

Figure 4 ROC curves of the Voting ensemble model for predicting each TCM syndrome
based on clinical features of chronic cough patients in the test set

型对湿热郁肺证和风邪伏肺证预测的 AUC 值均达到了 0.99 以上，与相比于其他证候的预测效果，具有更好的预测性能。

4 讨论

中医辨证是一种非线性的、复杂又兼具逻辑性的思维过程，注重了整体观与群体智慧集成，而多种机器学习算法通过不同的计算原理可为中医证候诊断提供多种实现方案^[3]。机器学习算法可以不同原理计算临床特征与证候诊断间的关联性，将临床中多维度的临床特征信息整合，这与中医整体审察的诊断原则高度契合。集成学习算法则可通过整合多个基础模型的预测结果来提升整体预测的性能。本研究使用的 Voting 集成算法模型融合了 LR、DT、MLP 和 Bagging 4 种基础机器学习算法，其中 LR 为线性特征筛选，DT 通过树形结构和规则路径实现直观的可视化，MLP 擅长非线性关系的挖掘，Bagging 可通过引入数据扰动来提升模型稳健性^{[20] 171-189}。本研究结果显示，4 种基础模型均具有一定的诊断能力（准确率均>0.7），但受模型自身算法的限制，单一模型的整体诊断效能有待加强，对于不同证候的诊断效果也存在较大差异，存在对

不同证候预测效能较好与较差共存的现象，如 LR 模型对风邪伏肺证、寒饮伏肺证和痰热郁肺证的预测能力较好，但同时对于肺阳亏虚证预测性能最差；DT 模型对肺阳亏虚证的预测有着更好的效果，但同时对于痰湿阻肺证的预测效果较差，这也反映了单一思维的局限性。本研究将 Voting 集成算法引入慢性咳嗽辨证模型构建，结果显示融合了 4 种基础模型的 Voting 集成算法模型有着优于 4 种模型的中医证候诊断性能（测试集中准确率 0.9253，AUC 0.9836）；而在对每个中医证候的诊断性能评价中，Voting 集成算法模型则对各个证候均表现出较高的预测效能，且对湿热郁肺证（AUC 0.9984）和风邪伏肺证（AUC 0.9970）诊断效果更优。4 种机器学习基础模型模拟了不同的思维方式，体现了中医辨证思维的多样性，通过 Voting 集成算法多数投票原则对基础模型的预测结果进行集成，体现了利用经验集成、群体智慧以优化诊断的临床思维。

慢性咳嗽病程较长，病情反复发作，易受环境、气候、饮食、失治误治等多种因素的影响^[8]。“肺为娇脏，寒热皆所不宜”（《医学源流论》），多种因素致使慢性咳嗽出现复杂病机。观其症状及

证候类型的分布,可见风邪伏肺为慢性咳嗽的基本病机,湿热、寒饮、痰热、痰湿、阳虚等病机常见,多种病机共存构成了慢性咳嗽常见复合证候类型的临床特点。在临床实践中疾病常见单一证型、复合证型共存的情况,具体表现为存在重叠或相似症状较多,多种病机的混杂也使得单一证型与复合证型分类间的差异度下降,降低了诊断模型构建的成功率。因此,结合临床专科专病的证候分布实际与中医辨证思维,建立适应中医临床辨证实际的诊断模型为临床所亟需。对于慢性咳嗽中医证候的诊断,与 4 种基础模型相比,Voting 集成算法模型拥有了更强的诊断性能。然而基于机器学习和集成算法的中医预测模型仍是一种“黑箱”模型,相较于中医经典辨证方式,其可解释性较差,对于提高预测模型可解释方法尚待进一步研究。

本研究采用 Voting 集成算法融合了 4 种基础模型,构建的集成模型使得对慢性咳嗽中医证候诊断的效能得以优化,提高了慢性咳嗽中医辨证的准确性与客观性;同时采用的 SMOTE 技术实现对不平衡数据的预处理,提高了模型的泛化能力。然而本研究中病例数据来源单一、缺乏外部验证集对模型泛化能力的检验,尚需通过多中心前瞻性研究进一步验证。在今后的研究中,可立足于中医临床辨证的思维方式,指导构建整合多种机器学习的智能体模型,探索基于临床辨证知识与临床多模态数据双驱动的智能辨证方法,不断优化诊断模型的效能,实现诊断效能与可解释性的双重优化,为更好地解决临床数字化问题提供实证参考。

参考文献

- [1]李敏静,叶柏春,郭莉,等.基于因子分析法的慢性咳嗽中医证候规律研究[J].浙江中西医结合杂志,2019,29(9):773-776.
- [2]季坤,史利卿,宋欢,等.基于因子分析的 771 例慢性咳嗽患者中医证候分布特点研究[J].现代中医临床,2023,30(2):30-36.
- [3]郭小川,冯贞贞,刘文瑞,等.基于 Stacking 集成算法的中医证候诊断模型建立:以肺癌为例[J].中医杂志,2024,65(17):1775-1783.
- [4]姚帅君,闫敬来,杜彩凤,等.基于集成学习构建围绝经期综合征中医智能诊断模型[J].中医杂志,2023,64(6):572-580.
- [5]周志华.集成学习:基础与算法[M].北京:电子工业出版社,2020:65-68.
- [6]JAIN R, SINGH P, ABDELKADER M, et al. Efficient lung cancer detection using computational intelligence and ensemble learning [J]. PLoS One, 2024, 19 (9) : e0310882.
- [7]VICTOR CO, ADAUGO FO. Enhancing the weighted voting ensemble algorithm for tuberculosis predictive diagnosis[J]. Sci Rep, 2021, 11(1): 14806.
- [8]中华医学会呼吸病学分会哮喘学组.咳嗽的诊断与治疗指南(2021)[J].中华结核和呼吸杂志,2022,45(1):13-46.
- [9]世界中医药学会联合会肺系病专业委员会,中华中医药学会肺系病分会,天津中医药大学第二附属医院.咳嗽中医诊疗专家共识意见(2021)[J].中医杂志,2021,62(16):1465-1472.
- [10]周仲瑛,薛博瑜.周仲瑛实用中医内科学[M].北京:中国中医药出版社,2012:231-246.
- [11]国家市场监督管理总局,国家标准化管理委员会.中医临床诊疗术语 第 2 部分:证候[S].北京:中国标准出版社,2021:5-258.
- [12]刘智霖,史利卿,马建岭,等.难治性慢性咳嗽的“多维度病机”认识与证治浅探[J].北京中医药大学学报,2022,45(5):507-512.
- [13]白逸晨,史利卿,季坤,等.慢性咳嗽肺脾相关病机探讨[J].现代中医临床,2018,25(4):42-45.
- [14]史利卿,马建岭,杨凯,等.慢性咳嗽中医病因病机及证候学规律的认识与探讨[J].现代中医临床,2017,24(5):1-5.
- [15]白逸晨.慢性咳嗽肺脾相关病机探讨及临床特点观察[D].北京:北京中医药大学,2019.
- [16]黎敬波,马力.中医临床常见症状术语规范[M].北京:中国医药科技出版社,2005:1-86.
- [17]姚乃礼.中医症状鉴别诊断学[M].2版.北京:人民卫生出版社,1984:19-389.
- [18]季绍良,成肇智.中医诊断学[M].北京:人民卫生出版社,2011:62-174.
- [19]王晓霞,李雷孝,林浩.SMOTE 类算法研究综述[J].计算机科学与探索,2024,18(5):1135-1159.
- [20]周志华.机器学习[M].北京:清华大学出版社,2016.

Construction of a Diagnostic Model for Traditional Chinese Medicine Syndromes of Chronic Cough Based on the Voting Ensemble Machine Learning Algorithm

BAI Yichen¹, QIN Suyang¹, ZHOU Chongyun¹, SHI Liqing², JI Kun², ZHANG Chuchu¹, LI Panfei³, CUI Tangming¹, LI Haiyan¹

1. Institute of Information on Traditional Chinese Medicine, China Academy of Chinese Medical Sciences, Beijing, 100700; 2. Dongfang Hospital, Beijing University of Chinese Medicine; 3. School of Medical Technology and Information Engineering, Zhejiang Chinese Medical University

ABSTRACT **Objective** To explore the construction of a machine learning model for the diagnosis of traditional Chinese medicine (TCM) syndromes in chronic cough and the optimization of this model using the Voting ensemble algorithm. **Methods** A retrospective analysis was conducted using clinical data from 921 patients with chronic cough treated at the Respiratory Department of Dongfang Hospital, Beijing University of Chinese Medicine. After standardized processing, 84 clinical features were extracted to determine TCM syndrome types. A specialized dataset for TCM syndrome diagnosis in chronic cough was formed by selecting syndrome types with more than 50 cases. The synthetic minority over-sampling technique (SMOTE) was employed to balance the dataset. Four base models, logistic regression (LR), decision tree (dt), multilayer perceptron (MLP), and Bagging, were constructed and integrated using a hard voting strategy to form a Voting ensemble model. Model performance was evaluated using accuracy, recall, precision, F1-score, receiver operating characteristic (ROC) curve, area under the curve (AUC), and confusion matrix. **Results** Among the 921 cases, six syndrome types had over 50 cases each, phlegm-heat obstructing the lung (294 cases), wind pathogen latent in the lung (103 cases), cold-phlegm obstructing the lung (102 cases), damp-heat stagnating in the lung (64 cases), lung *yang* deficiency (54 cases), and phlegm-damp obstructing the lung (53 cases), yielding a total of 670 cases in the specialized dataset. High-frequency symptoms among these patients included cough, expectoration, odor-induced cough, throat itchiness, itch-induced cough, and cough triggered by cold wind. Among the four base models, the MLP model showed the best diagnostic performance (test accuracy: 0.9104; AUC: 0.9828). Compared with the base models, the Voting ensemble model achieved superior performance with an accuracy of 0.9289 on the training set and 0.9253 on the test set, showing a minimal overfitting gap of 0.0036. It also achieved the highest AUC (0.9836) in the test set, outperforming all base models. The model exhibited especially strong diagnostic performance for damp-heat stagnating in the lung (AUC: 0.9984) and wind pathogen latent in the lung (AUC: 0.9970). **Conclusion** The Voting ensemble algorithm effectively integrates the strengths of multiple machine learning models, resulting in an optimized diagnostic model for TCM syndromes in chronic cough with high accuracy and enhanced generalization ability.

Keywords chronic cough; machine learning; syndrome; diagnosis model; Voting ensemble learning

(收稿日期: 2025-02-25; 修回日期: 2025-05-06)

[编辑: 柴倩云]