DOI:10.13193/j.issn.1673-7717.2025.06.005

基于机器学习方法预测胃癌前状态的比较研究

沈翊康[1],刘海煜[1],王少丽[2],张小平[1],胡晓楠[1],楼一方[1],刘震[2],谢晓霞[1],赵玉凤[1]
(1.中国中医科学院中医药数据中心,北京 100700;2.中国中医科学院广安门医院,北京 100053)

摘要:胃癌是消化道肿瘤中发病率较高的一种,对人们的健康构成严重威胁。早期预警、早期诊断可以提高5年生存率60%以上。近年来,机器学习方法在胃癌前状态预测方面取得了丰富的成果。首先从内镜图像、生化标志物、基因表达、中医证候等方面论述胃癌前状态的显著性特征,并结合机器学习方法预测的胃癌前状态研究成果,比较分析基于传统机器学习方法、深度学习方法和多种融合方法的优缺点。提出开展多中心、大样本队列研究并建设共享数据平台,实现多模态数据融合,增强模型的可解释性3点建议,为胃癌的"早诊早治"提供方法学借鉴。

关键词:胃癌;早诊早治;疾病状态预测;机器学习
中图分类号:R2-03 **文献标志码**:A **文章编号**:1673-7717(2025)06-0031-06

Comparative Research on Predicting Precancerous Gastric Conditions Using Machine Learning

SHEN Yikang[1], LIU Haiyu[1], WANG Shaoli[2], ZHANG Xiaoping[1], HU Xiaonan[1],
LOU Yifang[1], LIU Zhen[2], XIE Xiaoxia[1], ZHAO Yufeng[1]
(1. Data Center of Traditional Chinese Medicine, China Academy of Chinese Medical Sciences, Beijing 100700, China;
2. Guang'anmen Hospital, China Academy of Chinese Medical Sciences, Beijing 100700, China)

Abstract: Gastric cancer is one of the most common malignant tumors in the digestive tract and poses a serious threat to human health. Early warning and diagnosis can increase the 5-year survival rate by more than 60%. In recent years, machine learning methods have achieved rich results in predicting the precancerous state of gastric cancer. This article first discussed the significant features of the precancerous state of gastric cancer from the aspects of endoscopic images, biochemical markers, gene expression, and traditional Chinese medicine syndromes, and analyzed the advantages and disadvantages of traditional machine learning methods, deep learning methods and various fusion methods by combining the research results of the precancerous state of gastric cancer predicted by machine learning methods. It proposed three suggestions for carrying out multi-center, large-sample cohort studies and building shared data platforms, realizing multi-modal data fusion and enhancing the interpretability of the model, providing methodological references for the "early diagnosis and treatment" of gastric cancer.

Keywords: gastric cancer; early diagnosis and treatment; disease state prediction; machine learning

胃癌(Gastric Cancer,GC)是消化道肿瘤中最常见的一种。根据世界卫生组织数据,2020年中国胃癌新发病例约47.9万,死亡约37.4万,分别占全球的44.0%和48.6%[1]。胃癌在中国恶性肿瘤发病率中排名第四,病死率第三。2003—2015年,胃癌5年生存率从27.4%增至35.1%,仍低于日本和韩国。胃癌早期症状不明显,多数患者确诊时已是中晚期,中晚期患者即使接受手术,5年生存率仍低于30%,而早期发现并及时治疗的患者,5年生存率可超过90%[2]。因此,通过综合运用内镜检查、肿瘤标志物检测等多种诊断技术,早期发现胃癌患者的前驱病变,实现早诊早治,对于提高我国胃癌患者的生存率和生存质量具有至关重要的意义。

慢性萎缩性胃炎是胃癌最常见的前驱病变。目前,公认的胃癌发展进程是"慢性萎缩性胃炎→肠上皮化生→异型增生→胃腺癌"(Correa模型)[3],这是一个多因素、多步骤的演变过程。其中肠化生通常被认为是"炎-癌转化"进程中不可逆的癌前状态[4],对早期预测胃癌有重要意义。"炎-癌转化"进程受感染、环境、饮食、遗传等多种因素共同影响,这一过程十分复杂[5]。如何提前发现患者的胃癌前状态,对于降低胃癌发病率至关重要。

近年来,研究者采取病理、生化、流行病学等多种方法,发现胃癌患者早期预警标志物,精准预测胃癌前状态,成为降低胃癌发病率的有效手段[6]。同时,随着大数据技术和信息技术的快速发展,以机器学习算法为核心的人工智能方法,被临床上更广泛地应用到胃癌早期诊断和早期预警中[7-8]。影响人工智能辅助诊断与预测准确性的因素主要包括临床表征、机器学习算法的选择和优化、数据集质量等方面。因此,为了提高基于机器学习方法的胃癌早期诊断和预测准确性,本文从内镜图像、生化标志物、基因、临床证候等多个方面描述胃癌"炎-癌转化"进程的显著特征,比较不同类型机器学习方法在胃癌前状态预测中的实用性,归纳总结胃癌前状态预测的发展趋势,为人工智能辅助医疗决策提供重要的参考

基金项目:国家自然科学基金面上项目(82374621);中国中医科学院创新工程项目(CI2021A05042)
作者简介:沈翊康(1995-),男,江苏无锡人,博士在读,研究方向:真实世界中医药大数据分析与应用。
通讯作者:赵玉凤(1978-),女,黑龙江五常人,研究员,博士研究生导师,博士,研究方向:真实世界中医药大数据分析与应用。E-mail:snowmanzhao@163.com。

中国知网 https://www.cnki.net

价值。

1 胃癌前状态的显著性特征分析

根据病理组织学检查结果，胃癌演变进程分为胃癌前状态、胃癌前病变、胃癌3个主要阶段[11]。胃癌前状态包括胃黏膜萎缩和肠化生，二者可统称为慢性萎缩性胃炎（Chronic Atrophic Gastritis，CAG）[10]。胃癌前病变状态为胃上皮内瘤变。胃癌是指原发于胃的上皮源性恶性肿瘤[9]。上述疾病状态主要借助内镜图像进行诊断，各阶段疾病状态的内镜特点如表1所示。

了解胃癌前状态的显著特征对于提前发现患者的胃癌前状态具有重要的意义。这些特征包括流行病学特征、生化标志物特征、基因表达特征、细胞学特征、中医证候特征等。通过综合识别分析这些特征，有助于早期筛查胃癌高风险人群并及时加以干预措施。

1.1 胃癌前状态的流行病学特征

由于胃癌的演变进程受到环境、饮食、遗传等多种因素的共同影响，因此近年来众多流行病学调查致力于寻找影响胃癌前状态发生及转归的相关因素，例如幽门螺杆菌感染、高盐饮食、吸烟、饮酒和家庭遗传史等。幽门螺杆菌感染是最为明确的危险因素之一，其与慢性胃炎和胃黏膜肠化生的发生密切相关[43]。高盐饮食被认为可以损伤胃黏膜，增加患胃癌的风险。吸烟和饮酒不仅可以直接损伤胃黏膜，还可能通过诱发胃酸分泌过多来间接促进胃癌的发生[44]。此外，家族中有胃癌病例的人群，其患病风险也显著增加，提示遗传因素在胃癌前状态的形成和发展中起到重要作用。然而其研究结果存在较多争议，在各地区、人群中分布也不尽相同[17]。

1.2 胃癌前状态的生物学特征

胃癌前状态是胃癌发生发展的关键环节。深入研究其生物学特征，对于早期诊断、预防和治疗胃癌具有重要意义。通过解析胃癌前状态的生理病理机制，我们可以揭示胃癌发生的内在规律，为精准医学提供理论基础。

1.2.1 生化标志物特征 在胃癌前状态的发展过程中，一些生物标志物的水平会发生变化，这些变化可以作为诊断和监测疾病进展的指标。其中，血清胃蛋白酶原（pepsinogen，PG）可分为PGⅠ和PGⅡ两种，在胃黏膜出现炎症时，血清PGⅠ和PGⅡ水平增加；而在胃黏膜萎缩时，PGⅠ和PGⅡ水平下降，且PGⅠ/PGⅡ的比值降低。胃泌素17可受到胃酸的反馈调节，通常胃体黏膜萎缩时胃泌素17升高。PGⅠ/PGⅡ比值和胃泌素17联合检测已被证实可用于筛查胃黏膜萎缩，包括胃窦或胃体黏膜萎缩[11]。一项包含12 112例患者的前瞻性研究表明[12]，PGⅠ和PGⅠ/PGⅡ比值降低与胃癌的高风险有关，而胃泌素17水平过高或过低同样与胃癌的高风险有关，表明采用联合筛查策略将有助于检测胃癌的高风险人群。此外，血清肿瘤抗原也是判断CAG癌变进展和监测复发的重要血清学指标，包括癌胚抗原（Carcinoembryonic antigen，CEA）、糖类抗原（Carbohydrate antigen，CA）199、糖类抗原724等，但血清肿瘤抗原用于胃癌前病变筛查的价值仍有待进一步验证[5]。

1.2.2 胃癌分子亚型分类及其基因组特征 癌症基因组图谱（The Cancer Genome Atlas，TCGA）项目通过综合基因组、转录组、甲基化组和蛋白组测序分析，对胃癌的基因表达特征进行了深入的研究。在对295例胃癌患者样本的分析中[13]，研究者发现胃癌可分为4个分子亚型：①表达Epstein－Barr病毒的肿瘤，特点是反复出现PIK3CA突变、极度DNA超甲基化和JAK2、CD274以及PDCD1LG2的扩增；②微卫星不稳定的肿瘤，表现出突变率升高，包括编码可靶向致癌信号蛋白基因的突变；③基因组稳定的肿瘤，富含弥漫性组织学亚型；④染色体不稳定的肿瘤，表现出明显的非整倍性和受体酪氨酸激酶的局灶性扩增。这一分类方式将为胃癌的详细分类和精准预测提供基因层面的支持。

1.2.3 细胞学特征 解痉多肽表达化生（Spasmolytic Polypeptide Expressing Metaplasia，SPEM）细胞是胃黏膜上皮细胞在受损后发生的一种适应性转化，这种转化既是一种修复机制，也是胃癌发展的潜在风险因素。在胃壁黏膜层受到损伤时，表层细胞会经历乳头状增生等病理变化以进行修复。SPEM细胞的出现是这一过程中的关键步骤，它既是胃黏膜对损伤的适应性反应，又可能诱发胃癌。当该细胞的正常分化轨迹发生改变，可能导致细胞生长控制能力的丧失，进而增加癌变的风险[14]。此外，研究还发现，SPEM细胞在慢性肠炎的背景下可能发展为肠化生[15]，或者不经过肠化生阶段而直接发展为异型增生[16]。

1.3 胃癌前状态的中医学特征

1.3.1 胃癌前状态的中医病名与病因病机 中医古籍中未见慢性萎缩性胃炎或胃癌等病名，按《慢性胃炎中医诊疗专家共识意见（2017）》[19]，可根据临床症状，如胃脘胀满、疼痛、胃灼热等，将胃癌前状态归于“胃脘痛”“痞满”“嘈杂”等范畴。病因则主要与脾胃虚弱、情志失调、饮食不节、药物、外邪（幽门螺杆菌感染）等多种因素有关[45]。因胃在生理上以和降为顺，在病理上因滞而病[20]，而上述病因能够损伤脾胃，致运化失司，升降失常，而发生气滞、湿阻、寒凝、火郁、血瘀等，演变为胃癌前状态，进而导致胃癌[18-19]。

1.3.2 胃癌前状态的中医证候特点 共识意见[21]主要将CAG分为肝胃气滞、肝胃郁热、脾胃湿热、脾胃气虚、脾胃虚寒、胃阴亏虚及胃络瘀血证，并给出了具体定义。黄远程等[22]通过文献统计发现肝胃不和证和脾胃气虚证为慢性非萎缩性胃炎和CAG的核心证候，其中慢性非萎缩性胃炎病机以胃气滞为显著，CAG病机以脾气虚更为显著。胃阴亏虚、胃络瘀阻是CAG区别于慢性非萎缩性胃炎的主要证候特点。李星等[23]通过文献统计发现胃癌中医证型以瘀毒内阻、脾胃气虚、脾胃虚寒、气血虚、肝胃不和、痰湿凝结、胃热伤阴为主，这符合胃癌本虚标实的病机。刘北石[24]通过文献统计发现胃癌早期以肝胃不和型居多，中晚期胃癌以脾胃虚弱型、气血两虚型、瘀毒内阻型、痰湿凝滞型居多；胃癌的辨证分型与年龄有关，老人

表1 胃癌进程中疾病状态的内镜观察特点比较

疾病状态	病理名称	医学定义	内镜下图像特点
胃癌前状态	胃黏膜萎缩	胃腔固有腺体减少，胃黏膜变薄，胃小凹变浅	黏膜红白相间，以白色为主，皱襞变平甚至消失，部分黏膜血管显露；可伴有黏膜颗粒或结节状等表现
胃癌前状态	肠化生	胃黏膜出现肠腺上皮	黏膜欠光滑或灰色斑
胃癌前病变	胃上皮内瘤变	以出现不同程度的细胞和结构异型性为特征的胃黏膜上皮增生，包括胃低级别上皮内瘤变和高级别上皮内瘤	低级别上皮内瘤变：黏膜结构改变轻微；高级别上皮内瘤变：黏膜腺体结构异型性明显

中国知网 https://www.cnki.net

多虚证，年轻人多实证。

比较上述各类胃癌前状态的显著特征如表2。由此可以，当前针对各类胃癌前状态的评估方法多样且复杂，不同特征的检测在成本、患者接受度和准确性等方面存在差异，因此未来的研究方向应着重于开发更有效、更经济、更可接受的综合评估策略，以便于早期发现和干预胃癌前状态。

2 基于机器学习方法预测胃癌前状态的比较分析

机器学习是一种人工智能领域的技术，旨在让计算机系统能够通过数据学习以完成特定任务并不断改进其性能[25]。根据模型的复杂性和学习方式可将机器学习方法分为两类：传统机器学习方法倾向于手动设计特征和模型，主要基于特征工程和模型选择，通常适用于结构化数据和较简单的任务；深度学习则更侧重于自动特征提取和端到端学习，通过构建多层神经网络来学习数据的复杂特征和模式，适用于大规模数据和高度非线性问题[26]。

2.1 基于传统机器学习方法的胃癌前状态预测

日益发展的机器学习技术能够有效地整合胃癌前状态的复杂多维特征，为预测胃癌的演变进程提供了强大的分析工具。支持向量机（SVM）、随机森林（RF）、逻辑回归等传统机器学习方法已经被应用于对这些特征进行分析和预测，以建立与胃癌发病之间的关联性。例如，TANINAGA J 等[27]展示了利用 XGBoost 模型对胃癌风险因素进行预测的基本特性。他们通过逐步添加更多的影响因素，观察到模型学习结果的 ROC 曲线（如插页Ⅷ图 1 中 ModelA ~ ModelE 所示）。随着所学习的因素增多，AUC 值也逐步增大，表明模型能够有效地学习数据并做出准确的预测。

SHANG Z 等[28]采用频繁模式生长（FP - growth）分析了舌象参数图像信息与内镜图像的相关性规律，发现紫绀舌、滑毛舌、黄毛舌和斑点舌依次与胃窦黏膜充血或水肿和胃窦黏膜红斑相关。BRINDHA S K 等[29]使用了 Logistic 回归、朴素贝叶斯、多层感知机（Multi - Layer Perceptron，MLP）、支持向量机（Support Vector Machine，SVM）、随机森林（Random Forest，RF）五种机器学习算法从人群的饮食和生活方式特征数据中预测早期 GC 影响因素。结果表明，朴素贝叶斯、Logistic 回归和多层感知器是最适合的分类算法。SAFDARI R 等[30]开发了一个采用模糊逻辑表示领域知识并用于筛选和评估胃癌风险水平的模糊专家系统，通过专家访谈选择了 13 个特征作为导致胃癌的影响因素，然后进一步分类并建模。结果表明，该系统与医学专家的声明的一致性为 87%，灵敏度算为 92.1%。因此传统机器学习方法可以为胃癌前状态预测构建良好的模型。

2.2 基于深度学习方法的胃癌前状态预测

传统机器学习算法在处理患者标签化数据方面具有一定预测能力，但对于被广泛认可的胃癌诊断标准——内镜检测、病理学和 CT 图像等数据，传统机器学习算法显得力不从心。此时，需要借助深度学习方法来提取特征。

HUANG C R 等[31]利用经过改进的特征选择与神经网络（Recursive Feature Selection Neural Network，RFSNN），学习分类 104 名患者的内窥镜图像与 Hp 相关的胃组织学特征，发现具有较好效果。且 RFSNN 对于预测胃萎缩、肠化生以 Hp 相关胃炎的严重程度非常准确（>80%）。LI C 等[32]为实现癌性胃黏膜和正常胃黏膜样本拉曼光谱的自动分类，建立了一个以 SENet 突出重要特征，后将其放入 LSTM 网络中，以融合频谱的重要特征之间的关系。结果显示，该模型准确率和 AUC 分别达到 96.20% 和 0.99。YUAN L 等[33]通过临床前瞻性多中心研究发现，三种基于舌像的深度学习诊断模型（APINet、TransFG、DeepLabV3 +）在内部验证中 GC 患者的 AUC 值达到了 0.88 ~ 0.92，在独立外部验证中达到了 0.83 ~ 0.88，显著优于 8 个血液标志物组合的传统诊断方法。

总之，深度学习模型构建可以采取两种策略，一种是直接利用神经网络算法进行全部的预测，另一种是利用神经网络算法提取出标签化特征，然后在已提取特征的基础上进行预测，两种方法均有一定优势。

2.3 基于融合机器学习方法的胃癌前状态预测

除了直接应用已有的单一模型进行胃癌预测，近来研究人员开始探索多种算法相结合的方式，整合它们在特征提取、模式识别和数据处理等方面的能力，以提高预测效率和准确性，进一步提升胃癌预测的综合性能。

BARADARAN REZAEI H 等[34]提出一种基于 Ridge 回归、Elastic Net、Logistic 回归、随机森林、梯度提升决策树和深度神经网络等机器学习算法的集成机器学习算法以预测 GC 的早期发病率和病死率。该方法先筛选出有诊断价值的特征，然后分别用单一模型进行训练，最后使用 Stacking 方法对单一模型进行集成。结果表明，集成方法预测 GC 的早期发病率和病死率达到 97.9% 和 76.3%。ZHOU 等[35]提出了一种神经网络，将图像的浅层特征（包括形态、统计、亚区域和基于模型的特征）与深层特征（全连接卷积神经网络提取）融合，用于癌前病变识别。对比传统机器学习（SVM、RF、反向传播神经网络）和深度学习方法（GoogLeNet、ResNet、ResNeXt），融合特征方法表现更好，因为它能克服浅层特征不足以区分疾病和深层特征过拟合等问题。YACOB Y M 等[36]依次使用了预训练的

表2 胃癌前状态显著性特征的比较

状态角度	特点	优点	缺点
内镜特征	可以发现胃黏膜的异常变化，如黏膜增厚、溃疡、结节等	能够直接观察到病变部位	有一定的侵入性和风险，不适用于大规模筛查
生化标志物特征	检测血清中的特定化学物质	非侵入性、简便快速	特异性和敏感性不高，易受其他因素干扰，存在误诊和漏诊的可能性
基因特征	通过分子生物学方法检测相关基因的突变或表达异常	能够从遗传层面揭示个体患病的风险	成本和技术要求较高，且仅能检测特定的遗传变异
细胞特征	通过细胞学检查，观察胃黏膜表面细胞的形态、结构等性质	能够发现细胞学上的异常变化	需要专业技术人员进行解读，成本较高
流行病学特征	通过人群流行病学调查，分析胃癌发病的相关因素和规律	揭示胃癌的发病机制和危险因素	受到调查对象的选择、样本量和调查方法等因素的影响
中医特征	从中医理论的角度分析胃癌前状态的体质特点、病因病机	能够整体、动态地评估患者的状况	缺乏客观化的检测手段，诊断标准较为模糊

中国知网 https://www.cnki.net

表 3 胃癌前状态预测方法的比较

方法类型	特点	优点	缺点
传统机器学习方法	需要人工设计特征	适用于标签化数据和简单特征，训练速度较快	对特征工程依赖性强，需要人工标注
深度机器学习方法	能够自动提取高级特征和复杂模式	能够处理非线性关系和复杂数据，具有较高的预测准确性和泛化能力	训练需要大量数据和计算资源
融合机器学习方法	结合了传统方法和深度学习方法的优势	提高了预测性能和鲁棒性，能够充分利用不同方法的优势	需要额外的模型集成和特征融合策略

ShuffleNet Version 1、CCA 特征融合、Relief－F 特征选择方法以实现对 Hp 相关慢性胃炎的内镜图像分类，并获得了较高的总体准确率。WANG S 等[37]提出了一种多实例深度学习方法，首先使用基于 ResNet 的传统网络生成每个补丁实例的特征和异常概率，然后采用局部－全局特征融合和注意力机制聚合实例特征。他们在数据集上对胃切片进行分类，结果表明该方法效果显著，可有效区分 GC、Dys 和正常状态。

比较上述三类基于机器学习方法预测胃癌前状态的特点如表 3。由此可见，三类基于机器学习方法在预测胃癌前状态时各有优势。传统方法依赖人工设计特征，适用于简单数据；深度学习方法自动提取特征，适用于复杂数据但需大量资源；融合方法结合前两者优势，提升性能但需复杂策略。因此，在选择预测胃癌前状态的方法时，应综合考虑数据量、计算资源以及模型的复杂度，以确定最合适的机器学习方法。

3 提高胃癌前状态预测准确性的方法

3.1 构建高质量大规模临床数据集

目前，胃癌检测的人工智能研究多为样本量有限的横截面研究，缺乏对胃癌图像的动态监测以及与生化指标、中医证候等多因素的整合研究[4]。此外，由于样本量有限和公共数据库中内镜图像等关键数据的缺失，不同研究方法难以进行有效比较，导致结果具有一定的局限性。为解决这些问题，建议开展多中心、大样本队列研究，并构建高质量共享数据库。这将有助于全面考虑影响胃癌检测的各种因素，提高研究的信度和可靠性，为胃癌的早期检测和精准治疗提供更有力的科学依据。

3.2 多模态数据融合

当前胃癌预测研究中，虽然一些学者尝试融合多种机器学习算法，但这些尝试相对简单。作为一种多因素多步骤的疾病，胃癌病因的多样性和复杂性决定了单一模型的局限性。因此，探索多模态数据融合方法对于提高胃癌预测的准确性和可靠性至关重要。通过融合不同类型的数据，如临床、基因表达和图像数据，可以从多个角度提供更全面准确的预测。研究人员应探索多模态数据联合处理和融合算法，如协同训练（Co－training）[38]、注意力机制（Attention Mechanisms）[39]模型等，利用不同数据类型之间的相互关联和互补信息，优化分类模型性能，实现更精确的胃癌预测。

3.3 构建面向真实世界临床数据的因果学习框架

机器学习算法在当前的医学领域的应用日益广泛，然而，大多数研究仅停留在关联分析层面，难以深入挖掘潜在的因果关系。在临床研究中，因果推断对于理解疾病发生机制、评估干预措施的效果具有重要意义。然而，真实世界的临床数据往往存在高维性、噪声、混杂等问题，传统的因果推断方法难以直接应用。因此，迫切需要开发能够处理复杂临床数据的因果学习算法。

3.4 增强算法可解释性

尽管胃癌检测领域中的人工智能技术发展迅速，但其不可解释性导致临床医生缺乏信任，难以推广[40]。因此，急需将可解释方法应用于胃癌前状态预测模型中。研究人员需要探索

构建可解释的 ML 模型，如提供局部解释的 LIME（Local Interpretable Model－agnostic Explanations）技术[41-44]，并给出决策支

持的解释和可视化工具，以帮助医生理解和信任这些预测模型，从而增强医生和患者对预测结果的信任和接受度。

4 小结

中西医临床数据驱动的胃癌前状态预测研究对于早期发现和干预胃癌具有重要意义。本文通过综合现代医学和中医学的研究进展，探讨了胃癌前状态的多样化特征，主要包括内镜图像、生化标志物、基因和中医证候等方面。由于特征的多样性和复杂性，预测胃癌前状态颇具挑战，因此需要结合机器学习算法来提高效率与准确性。研究比较了传统机器学习深度学习和融合机器学习方法的优缺点，发现融合机器学习方法在多源异构的胃癌前状态数据处理方面具有优势。并进一步提出了开展多中心大样本队列研究、建设共享数据平台和实现多模态数据融合等建议。未来研究需要整合多学科知识，探索新的预测模型和方法，以更好地服务于临床实践，提高早期诊断率和治疗效果，降低胃癌的发病率和病死率。

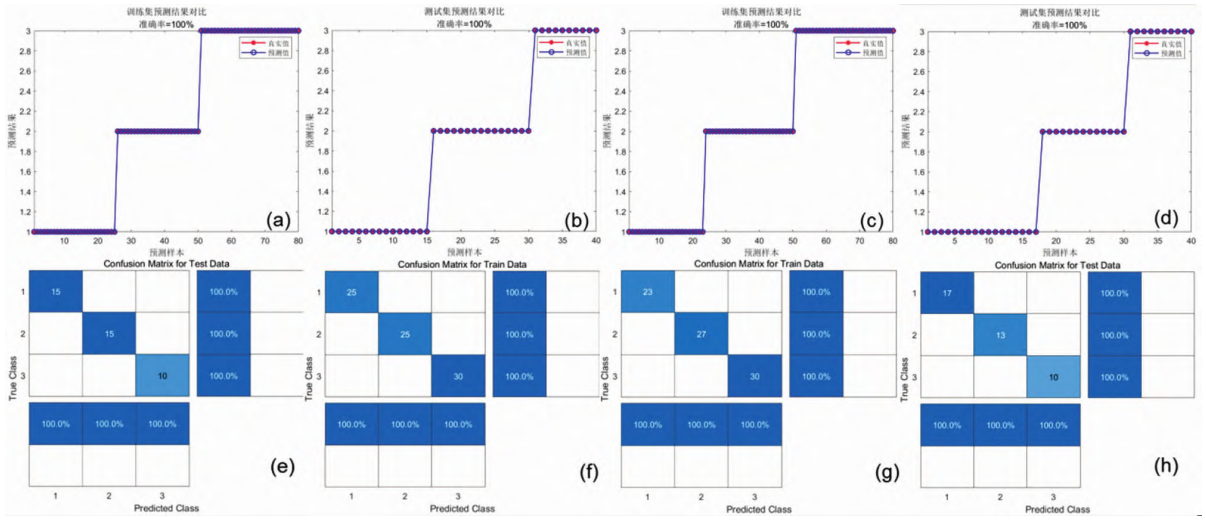
参考文献

[1] SUNG H, FERLAY J, SIEGEL R L, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA Cancer J Clin, 2021, 71(3): 209－249.
[2] 赫捷，陈万青，李兆申，等. 中国胃癌筛查与早诊早治指南（2022，北京）[J]. 中国肿瘤，2022，31(7)：488－527.
[3] CORREA P, PIAZUELO M B. Helicobacter pylori infection and gastric adenocarcinoma[J]. US Gastroenterol Hepatol Rev, 2011, 7(1): 59－64.
[4] 唐旭东，张泰，卞立群，等. 萎缩性胃炎与胃癌前病变诊疗研究的关键问题及其对策[J]. 中国中西医结合消化杂志，2023，31(5)：317－322.
[5] 邹天慧，廖晓宏，陈萦晅. 慢性萎缩性胃炎癌变的预警、早期诊断与预防[J]. 临床荟萃，2019，34(5)：407－411.
[6] 张开光，提高胃癌前病变及早期胃癌检出率的系列研究[N]. 安徽省，安徽省立医院，2019－06－27.
[7] NIU P H, ZHAO L L, WU H L, et al. Artificial intelligence in gastric cancer: Application and future perspectives[J]. World journal of gastroenterology, 2020, 26(36): 5408.
[8] BHARDWAJ P, BHANDARI G, KUMAR Y, et al. An investigational approach for the prediction of gastric cancer using artificial intelligence techniques: a systematic review[J]. Archives of Computational Methods in Engineering, 2022, 29(6): 4379－4400.
[9] 国家卫生健康委. 胃癌诊疗指南[EB/OL]. 2022[2024－01－05]. http://www.nhc.gov.cn/yzygj/s7659/202204/

中国知网 https://www.cnki.net

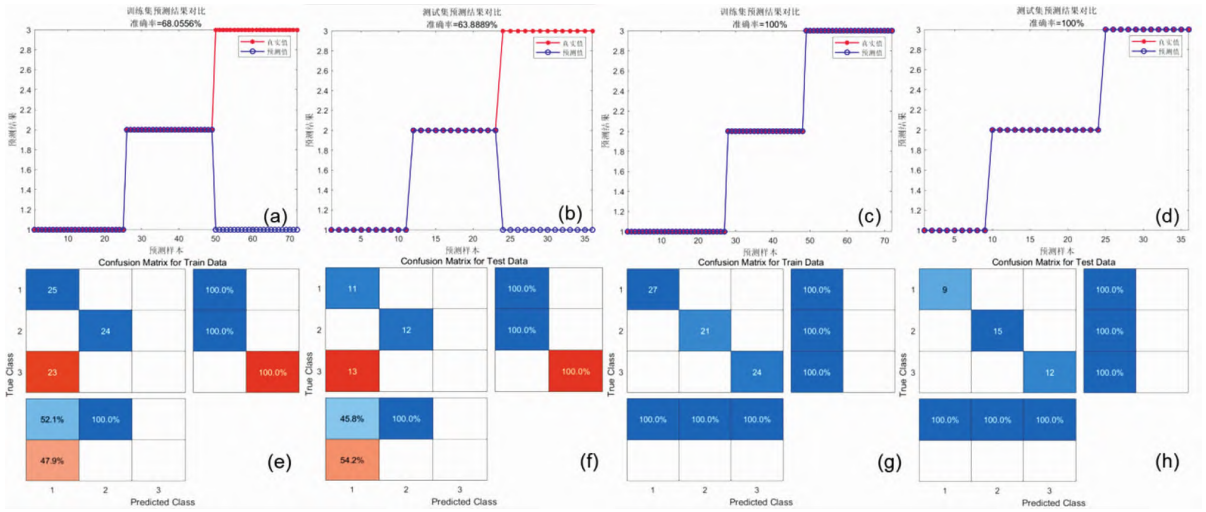
a0e67177df1f439898683e1333957c74. shtml.
[10] 国家消化系疾病临床医学研究中心(上海),国家消化道早癌防治中心联盟,中华医学会消化病学分会幽门螺杆菌学组,等. 中国胃黏膜癌前状态和癌前病变的处理策略专家共识(2020年)[J]. 中华消化杂志,2020,40(11):731-741.
[11] AGRéUS L,KUIPERS E J,KUPCINSKAS L,et al. Rationale in diagnosis and screening of atrophic gastritis with stomach-specific plasma biomarkers[J]. Scandinavian journal of gastroenterology,2012,47(2):136-147.
[12] TU H,SUN L,DONG X,et al. A serological biopsy using five stomach-specific circulating biomarkers for gastric cancer risk assessment: a multi-phase study[J]. Official journal of the American College of Gastroenterology ACG,2017,112(5):704-715.
[13] Cancer Genome Atlas Research Network Analysis Working Group: Dana-Farber Cancer Institute Bass Adam J. adam_bass@ dfci. harvard. edu 1 b,Institute for Systems Biology Thorsson Vesteinn 2 Shmulevich Ilya 2 Reynolds Sheila M. 2 Miller Michael 2 Bernard Brady 2,University of Southern California Hinoue Toshinori 3 Laird Peter W. 3 Curtis Christina 4 Shen Hui 3 Weisenberger Daniel J. 3,et al. Comprehensive molecular characterization of gastric adenocarcinoma[J]. Nature,2014,513(7517):202-209.
[14] GOLDENRING J R,MILLS J C. Cellular plasticity,reprogramming,and regeneration: metaplasia in the stomach and beyond[J]. Gastroenterology,2022,162(2):415-430.
[15] LENNERZ J K M,KIM S H,OATES E L,et al. The transcription factor MIST1 is a novel human gastric chief cell marker whose expression is lost in metaplasia,dysplasia,and carcinoma[J]. The American journal of pathology,2010,177(3):1514-1533.
[16] ADKINS-THREATS M,MILLS J C. Cell plasticity in regeneration in the stomach and beyond[J]. Current Opinion in Genetics & Development,2022,75:101948.
[17] MYSURU SHIVANNA L,UROOJ A. A review on dietary and non-dietary risk factors associated with gastrointestinal cancer[J]. Journal of gastrointestinal cancer,2016,47:247-254.
[18] CORREA P,HOUGHTON J M. Carcinogenesis of Helicobacter pylori[J]. Gastroenterology,2007,133(2):659-672.
[19] 张声生,唐旭东,黄穗平,等. 慢性胃炎中医诊疗专家共识意见(2017)[J]. 中华中医药杂志,2017,32(7):3060-3064.
[20] 唐旭东. 董建华"通降论"学术思想整理[J]. 北京中医药大学学报,1995(2):45-48.
[21] 李军祥,陈誩,吕宾,等. 慢性萎缩性胃炎中西医结合诊疗共识意见(2017年)[J]. 中国中西医结合消化杂志,2018,26(2):121-131.
[22] 黄远程,黄超原,梁怀枫,等. 慢性非萎缩性胃炎与慢性萎缩性胃炎证候、证素分布规律文献研究[J]. 中医杂志,2019,60(16):1417-1422.
[23] 李星,樊巧玲. 胃癌中医辨证与方药应用的文献研究[J]. 中医杂志,2017,58(8):693-696.
[24] 刘北石. 胃癌中医证型症候的演变规律及其相关性的文献研究[D]. 沈阳:辽宁中医药大学,2014.
[25] GOODFELLOW I,BENGIO Y,COURVILLE A. Deep learning[M]. MIT Press,2016.
[26] JAMES G,WITTEN D,HASTIE T,et al. An introduction to statistical learning[M]. Springer Science & Business Media,2013.
[27] TANINAGA J,NISHIYAMA Y,FUJIBAYASHI K,et al. Prediction of future gastric cancer risk using a machine learning algorithm and comprehensive medical check-up data: A case-control study[J]. Scientific reports,2019,9(1):12384.
[28] SHANG Z,ZHENGGUANG D U,BEI G,et al. Correlation analysis between characteristics under gastroscope and image information of tongue in patients with chronic gastritis[J]. Journal of Traditional Chinese Medicine,2022,42(1):102.
[29] BRINDHA S K,KUMAR N S,CHENKUAL S,et al. Data mining for early gastric cancer etiological factors from diet-lifestyle characteristics[C]//2020 4th international conference on intelligent computing and control systems (ICICCS), IEEE,2020:65-73.
[30] SAFDARI R,ARPANAHI H K,LANGARIZADEH M,et al. Design a fuzzy rule-based expert system to aid earlier diagnosis of gastric cancer[J]. Acta Informatica Medica,2018,26(1):19.
[31] HUANG C R,SHEU B S,CHUNG P C,et al. Computerized diagnosis of Helicobacter pylori infection and associated gastric inflammation from endoscopic images by refined feature selection using a neural network[J]. Endoscopy,2004,36(7):601-608.
[32] LI C,LIU S,ZHANG Q,et al. Combining Raman spectroscopy and machine learning to assist early diagnosis of gastric cancer[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy,2023,287:122049.
[33] YUAN L,YANG L,ZHANG S,et al. Development of a tongue image-based machine learning tool for the diagnosis of gastric cancer: a prospective multicentre clinical cohort study[J]. EClinical Medicine,2023,57.
[34] BARADARAN REZAEI H,AMJADIAN A,SEBT M V,et al. An ensemble method of the machine learning to prognosticate the gastric cancer[J]. Annals of Operations Research,2023,328(1):151-192.
[35] ZHOU H,LIU Z,LI T,et al. Classification of precancerous lesions based on fusion of multiple hierarchical features[J]. Computer methods and programs in biomedicine,2023,229:107301.
[36] YACOB Y M,ALQURAN H,MUSTAFA W A,et al. H. pylori Related Atrophic Gastritis Detection Using Enhanced Convolution Neural Network (CNN) Learner[J]. Diagnostics,2023,13(3):336.
[37] WANG S,ZHU Y,YU L,et al. RMDL: Recalibrated multi-instance deep learning for whole slide gastric image classification[J]. Medical image analysis,2019,58:101549.
[38] ZHOU Z H,LI M. Tri-training: Exploiting unlabeled data using three classifiers[J]. IEEE Transactions on Knowledge and Data Engineering,2005,17(11):1529-1541.
[39] VASWANI A,SHAZEER N,PARMAR N,et al. Attention is all you need[J]. Advances in neural information processing systems,2017(1):30.
[40] 张姝艳,皮婷婷. 医疗领域中人工智能应用的可解释性困境与治理[J]. 医学与哲学,2023,44(3):25-29,35.
[41] RIBEIRO M T,SINGH S,GUESTRIN C. " Why should i trust you?" Explaining the predictions of any classifier[C]//Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 2016:1135-1144.
[42] LEE J W J,ZHU F,SRIVASTAVA S,et al. Severity of gastric intestinal metaplasia predicts the risk of gastric cancer: a prospective multicentre cohort study (GCEP)[J]. Gut,2022,71(5):854-863.
[43] PIAZUELO M B,BRAVO L E,MERA R M,et al. The Colombian chemoprevention trial. Twenty-year follow-up of a cohort of patients with gastric precancerous lesions[J]. Gastroenterology,2021,160(4):1106-1117. e3.
[44] 邱萍. 慢性萎缩性胃炎中医病因、病机及诊疗研究进展[J]. 内蒙古中医药,2018,37(8):105-106.

中国知网 https://www.cnki.net



注：(a)、(e)RF 训练集，(b)、(f)RF 测试集，(c)、(g)PCA + RF 训练集，(d)、(h)PCA + RF 测试集。

图 13　THz 数据的 LSTM 分类结果



注：(a)、(e)RF 训练集，(b)、(f)RF 测试集，(c)、(g)PCA + RF 训练集，(d)、(h)PCA + RF 测试集。

图 14　RS 数据的 LSTM 分类结果

基于机器学习方法预测胃癌前状态的比较研究

（正文见 31 – 35 页）

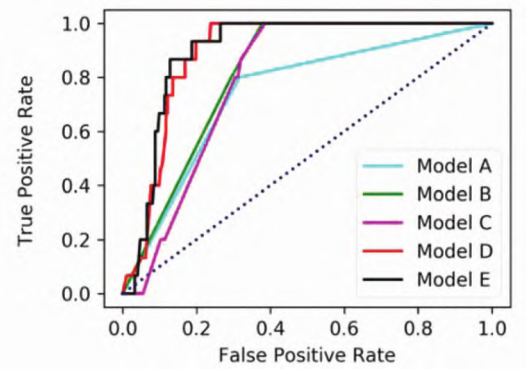


图 1　逐步添加影响因素后 XGBoost 模型表现出的 ROC 曲线

中国知网　https://www.cnki.net