·综述·

数据挖掘方法在真实世界研究中的应用进展

郭美玲[1]，柴克燕[1]，刘蕴资[1]，鲁佩颖[1]，周纪颖[1]，陈晓东[1]，孙　凤[2]，张晓朦[1]，吴嘉瑞[1]

1. 北京中医药大学中药学院（北京 102488）
2. 北京大学公共卫生学院（北京 100191）

【摘要】数据挖掘方法已被广泛应用于真实世界研究中的数据探索、规律挖掘、预测决策及偏倚控制等全过程，本文对真实世界研究常用的数据挖掘方法进行系统归纳，重点聚焦关联规则、聚类分析、因子分析、人工神经网络、倾向评分、决策树、贝叶斯网络、Logistic 回归、隐结构模型等十余种常见方法，系统梳理其定义、实现路径并结合真实世界研究的具体场景分析每种方法的适配性及应用效能，旨在通过对上述常用数据挖掘方法的系统整理与场景适配性分析，梳理不同方法在真实世界研究中的应用边界与优势，为真实世界研究中数据挖掘方法的科学选用、规范应用提供理论参考与实践指引，同时助力推动真实世界研究领域数据挖掘技术的规范化发展与创新融合，为相关研究的开展奠定基础。

【关键词】数据挖掘；真实世界研究；进展

【中图分类号】TP311.13　　【文献标识码】A

Application progress of data mining methods in real-world study

GUO Meiling[1], CHAI Keyan[1], LIU Yunzi[1], LU Peiying[1], ZHOU Jiying[1], CHEN Xiaodong[1], SUN Feng[2], ZHANG Xiaomeng[1], WU Jiarui[1]

1. School of Chinese Material Medica, Beijing University of Chinese Medicine, Beijing 102488, China
2. School of Public Health, Peking University, Beijing 100191, China
Corresponding authors: WU Jiarui, Email: exogamy@163.com; ZHANG Xiaomeng, Email: zhangxm0320@163.com

【Abstract】Data mining methods have been widely applied throughout the entire process of real-world study, including data exploration, pattern mining, predictive decision-making, and bias control. This paper presents the first systematic summary of the commonly used data mining methods in real-world study, focusing on ten common methods such as association rules, cluster analysis, factor analysis, artificial neural networks, propensity scores, decision trees, Bayesian networks, Logistic regression, and latent structure models. It systematically reviews their definitions, implementation paths, and combines them with specific scenarios of real-world study to deeply analyze the suitability and application efficiency of each method. This paper systematically summarizes the commonly used data mining methods in real-world study, clarifies the application boundaries and advantages of different methods in real-world study, provides theoretical references

DOI: 10.12173/j.issn.1005-0698.202511032
基金项目：国家自然科学基金面上项目（81473547、81673829）；北京中医药大学企业横向课题（BUCM-2024-JS-FW-104）
通信作者：吴嘉瑞，博士，教授，博士研究生导师，Email：exogamy@163.com
张晓朦，博士，讲师，Email：zhangxm0320@163.com

and practical guidance for the scientific selection and standardized application of data mining methods in real-world study, and helps promote the standardized development and innovative integration of data mining technology in the field of real-world study, laying the foundation for related research.

【Keywords】Data mining; Real-world study; Progress

真实世界研究因其能反映实际临床环境中患者的诊疗全过程，已成为支持医疗产品研发与决策的关键手段。特别是随着数字化、智能化在医药领域的应用，如电子病历、医保数据与可穿戴设备等多源信息的融合，真实世界数据已成为蕴含巨大潜力的医疗大数据载体[1–3]。国内外监管机构对真实世界数据也极为重视，相继发布相关指导原则[4–7]，推动真实世界数据在药品评价中的应用。然而，真实世界数据固有的复杂性，如质量异质性、缺失值普遍、隐私性强及混杂偏倚突出等，为深度分析带来了严峻挑战[8–9]。数据挖掘可以从海量复杂数据中提取有价值信息、识别规律并为支撑预测决策提供了强大可能[10]，但目前缺乏对常用真实世界数据处理方法的系统性梳理，还存在方法认知碎片化、实现路径不统一、场景适配逻辑不清等局限，导致临床研究者在方法选择与应用时面临困惑。

本文从500余篇相关文献中遴选出10余种常用数据挖掘方法，并进行系统分析归类，构建“方法定义–实现路径–场景应用”三位一体的整合框架，旨在清晰界定不同方法的技术边界，阐明其在真实世界研究中的适配场景与协同策略，从而为临床研究者提供清晰的方法学指引，提升真实世界研究的分析效能与证据质量。

1 真实世界研究常见数据挖掘方法

数据挖掘方法根据其解决的核心问题和应用场景，可归纳为四大类别，分别为关联性分析、隐性规律分析、偏倚控制分析和人工智能方法，各类方法的应用场景如表1和图1所示。

表1 常见数据挖掘方法分类及应用场景

Table 1. Classification and application scenarios of common data mining methods

类别	核心方法	主要应用场景
关联性分析	关联规则、聚类分析	风险因素识别、人群分层、药品处方模式分析
隐性规律分析	因子分析、隐结构模型	变量降维、病因机制探索、疾病分型
偏倚控制分析	倾向评分	不同干预措施的效果比较、真实世界疗效评价
人工智能方法	人工神经网络、Logistic回归、决策树、贝叶斯网络、LASSO回归、MGPS法	临床风险预警、疾病预后预测、药品不良反应、公共卫生干预决策

注：LASSO. 最小绝对值收缩和选择算子（least absolute shrinkage and selection operator）；MGPS. 多项伽马–泊松分布缩减（multi–item gamma Poisson shrinker）。

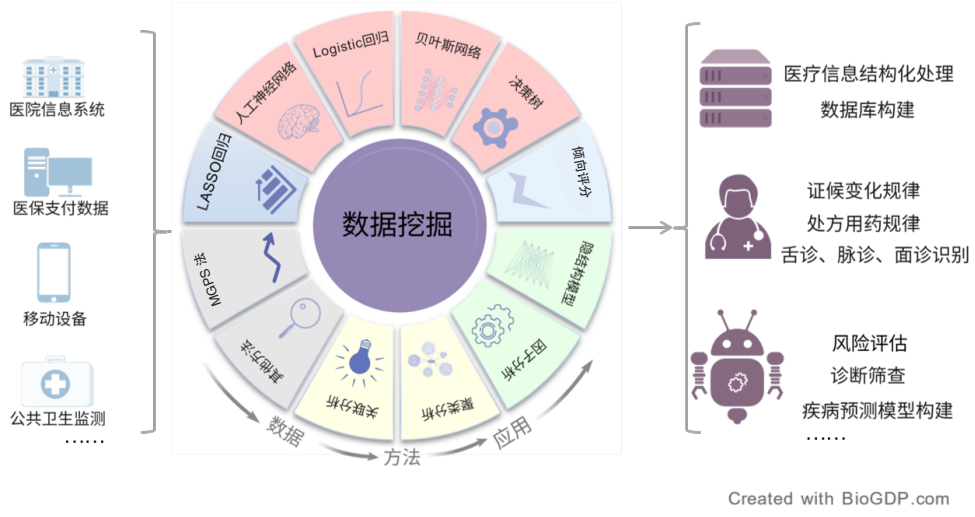

图1 真实世界研究数据来源及数据挖掘方法应用场景

Figure 1. Data sources and data mining method applications in real–world study

1.1 关联与聚类分析，探索数据内在结构

1.1.1 关联规则

（1）定义：关联规则是一种挖掘变量之间潜在依存关系的算法，通过计算支持度、置信度、提升度等指标，识别“若 A 出现，则 B 出现概率”的关联规则[11]。

（2）实现路径：根据研究目标和数据类型选择合适的算法，如 Apriori 算法、多层次关联规则算法、多值属性关联规则算法、其他关联规则算法等[12]。设定最小支持度和置信度从而限定规则范围，随后对预处理后的数据进行规则挖掘，最后从生成的规则中筛选高置信度、高支持度的关联对。

（3）应用实例：关联分析多用于探索潜在的临床关联规律，在处方配伍和疾病关联研究中价值显著。Kim 等[13]采用关联规则分析 169 959 例成年癌症患者的死亡风险与合并糖尿病、高血压的关联关系，最终证实合并糖尿病和高血压会增加癌症死亡率。施雪清等[14]以 Apriori 算法分析治疗多囊卵巢综合征处方药物的配伍关系，最终得到 3 942 条关联规则，确立以补肾为核心的常用药对组合。

1.1.2 聚类分析

（1）定义：聚类分析是一种基于数据的相似度将其自动划分为若干组，使组内对象相似度最高、组间对象相似度最低的方法[15]。

（2）实现路径：聚类分析的实现需遵循系统性流程，主要包括数据预处理、聚类算法选择、聚类结果评价、聚类结果可视化与解析 4 个部分[16]。首先根据数据特征选择聚类算法，真实世界研究中常用的是层次聚类和 K-means 聚类；然后通过轮廓系数或 V-measure 等评估以验证结果合理性[17]，最终结合专业知识确定聚类结果。

（3）应用实例：聚类分析常用于复杂数据归类，揭示治法、病机、诊治等的群体特征[18]。李家劼等[19]采用系统聚类分析归纳糖尿病周围神经病变合并血脂异常患者的证型分布规律，分别是痰瘀阻络证、阴虚血瘀证、脾肾阳虚证、肝肾亏虚证和气虚血瘀证。Min 等[20]采用 K-means 聚类分析探究中老年人群血浆动脉粥样硬化指数与心血管疾病发生率的关系，将血浆致动脉粥样硬化指数对照水平分为 5 类，发现降低该指数可显著降低心血管事件的发生率。

关联规则与聚类分析作为无监督学习的经典描述性数据挖掘方法，可实现对真实世界临床数据规律的初步探索，能够有效识别处方配伍模式、疾病表型分群等直观关联特征与群体聚类属性，为后续深层数据挖掘确定核心分析要点、明确研究方向。但需明确的是，真实世界临床数据的核心规律常隐藏于多变量的复杂关系之中，前述关联分析与群体分类结果仅能反映数据表层的关联特征，难以直接揭示其背后的潜在因素。因此，需进一步引入因子分析、隐结构模型等隐性规律挖掘技术，实现数据分析维度的递进，通过挖掘数据中无法直接观测的核心关联与结构特征，为阐释疾病证候特征、临床用药配伍等核心问题提供更具针对性的方法学依据。

1.2 隐性规律分析，揭示潜在结构与规律

1.2.1 因子分析

（1）定义：因子分析是一种用于降维与潜在结构挖掘的多元统计方法[21]，核心是从多个观测变量中提取少数几个互不相关的“公共因子”，揭示变量背后隐藏的、无法直接观测的潜在规律。因于分析是一种数据挖掘的中间手段，可为后续其他数据分析奠定基础。

（2）实现路径：因子分析需遵循数据适配、因子提取并命名、结果解释的标准化流程，具体步骤如图 2 所示。首先对原始数据进行预处理，然后通过 KMO 检验和 Bartlett 球形检验判断数据是否适合因子分析；根据载荷系数和因子的归属对应关系来命名公共因子；利用回归法或者 Bartlett 法计算因子得分，将其作为新变量用于后续分析。

（3）应用实例：因子分析更多适用于处理多症状-单证候、多成分-单功效等复杂数据，可实现症状特征规律总结、证候分类标准化、中药核心功效群挖掘等目标。吕咪等[22]对 800 例非糜烂性反流病和上腹痛综合征胃肠症状重叠患者的体征数据进行因子分析，得到 18 个公因子变量，发现病位证素肝、胃、脾最常出现，病性证素气滞、气虚、阴虚和阳虚、湿最常出现。江雯婷等[23]在中药饮片治疗 IgA 肾病的研究中纳入 585 例患者，针对使用频率前 60 位的中药饮片进行因子分析，最终得到 17 个公因子，其中半数公因子均有疏风解表药、祛风药，证实风类中药在延缓肾病进展中的重要作用。

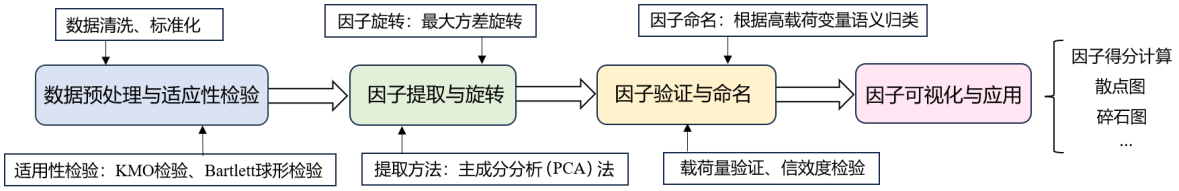

图2 因子分析实现流程

Figure 2. Process of factor analysis implementation

1.2.2 隐结构模型

（1）定义：隐结构模型是指利用隐类分析使用的统计模型，它是由隐变量和显变量组成，通过数据分析确定隐变量的个数、隐变量包含的各子变量对应的概率、隐变量之间的关系等。

（2）实现路径：构建隐结构模型首先对纳入分析的变量进行频次统计，结合研究需求保留高频变量；其次，基于筛选后的变量进一步构建隐树模型，采用贝叶斯信息准则（Bayesian information criterion，BIC）作为模型质量的核心评价指标；随后，对模型输出的隐变量开展综合聚类与专业诠释；在此基础上，结合研究核心目标将隐变量聚类结果转化为具有专业理论内涵的分类体系[24-26]。

（3）应用实例：隐结构模型适用于高维、异质性强的真实世界数据挖掘，多用于识别疾病潜在亚型或症状群，为精准分型、优化治疗方案提供依据。李伟柯等[27]利用隐结构模型量化慢性阻塞性肺疾病肺气虚证的严重程度，构建了轻、中、重度分级的诊断模型，实现了证候的量化评估。李俊雅等[28]运用隐结构模型结合关联规则分析中医药治疗非小细胞肺癌的用药规律，发现中药处方多用甘温之品，并挖掘出陈皮、茯苓、党参、黄芪等临床高频用药，为临床肺癌治疗的遣方用药提供了参考。

上述因子分析与隐结构模型，从复杂变量中挖掘出了潜在数据结构与核心关联规律，为阐释疾病证候特征、用药模式提供了深层的数据支撑。但真实世界临床数据普遍存在混杂因素、选择偏倚、数据缺失等偏倚问题，此类偏倚可能导致挖掘的隐性规律出现失真，进而影响后续结论的科学性。因此，需进一步引入偏倚控制方法，对数据偏倚进行识别、量化与校正，以保障隐性规律挖掘结果的可靠性。

1.3 偏倚控制分析，还原数据真实价值

1.3.1 倾向评分

（1）定义：倾向评分是一种用于控制混杂偏倚的统计方法[29]，具体定义为在其他混杂因素存在的条件下，研究对象进入处理组的条件概率。通过控制混杂偏倚，提升干预措施有效性评价的可靠性。

（2）实现路径：首先要明确暴露组（如接受中药治疗）和非暴露组（如接受常规治疗）；然后采用逻辑回归模型，通过最大似然估计求解模型参数，进而计算每个研究对象的倾向评分[30]；继而实施组间匹配，常用的算法包括1∶1最近邻匹配、卡钳匹配、马氏矩阵匹配等方法，使两组评分分布趋于一致[31]；最后通过标准化差异或 t 检验验证协变量平衡性，确保组间可比性后，再分析干预措施与结局的关联。

（3）应用实例：倾向评分广泛应用于中西医结合疗效评价对比、药品不良反应监测和处方分析等领域。袁博寒等[32]采用倾向评分匹配法控制组间基线差异，评价百令片联用西药对肾保护的真实作用。此外还有扶正散结方治疗乙肝相关肝癌、参附注射液治疗心力衰竭的疗效评价等研究[33-34]也采用了倾向评分方法。谢雁鸣团队[35]应用该方法探究脉络疏通丸的胃肠道安全性，结果显示该药物与疑似不良反应之间无显著关联，为其安全应用提供了支持性证据。

偏倚控制方法通过对临床数据中混杂、选择偏倚等干扰因素的识别与校正，有效还原了数据的真实价值，为后续深度数据挖掘奠定了基础。但是经过偏倚校正后的高质量临床数据仍蕴含高维、非线性的复杂交互关系，传统数据挖掘方法难以充分挖掘其在临床结局预测、个体化诊疗方案优化等方面的深层价值，也难以充分满足精准临床决策支持的需求。然而人工智能技术凭借其强大的特征自主学习与预测识别能力，可在可靠数据基础上进一步发挥数据的决策支撑的价值。因此，后续将聚焦人工智能技术类数据挖掘方法，探讨其在辅助临床决策中的应用路径与实践价值。

1.4 人工智能技术，辅助实现临床决策

1.4.1 决策树算法

（1）定义：决策树算法是机器学习、归纳学习与数据挖掘领域的核心树状预测及分类模型构建方法，以树状结构模拟人类分步决策逻辑，将复杂决策拆解为基于数据属性的简单选择，形成可解释的预测或分类规则[36]。

（2）实现路径：首先利用训练数据集建立决策树模型，然后根据此决策树模型对输入数据进行分类。决策树算法种类较多，最常用的是 ID3 算法、C4.5 算法、C5.0 算法、CART 算法和 CHAID 算法[37]。

（3）应用实例：决策树算法广泛应用于疾病的风险评估预测、病证诊断，以及药物经济学研究等。Nielsen 等[38]应用决策树和混合模型预测早期银屑病患者是否有患银屑病关节炎的风险。石玉琳等[39]采用 C5.0 决策树算法构建了基于舌脉象数据的非小细胞肺癌气虚证与阴虚证的证候分类模型，模型分类准确率为 80.37%。辛雅雯等[40]运用周期为 14 d 的决策树模型，发现以美罗培南为基础的抗菌药物方案治疗耐碳青霉烯病原菌更具经济学优势，但此研究存在局限性，属于回顾性单中心研究，病例资料质量有差异、不良反应描述不充分，且样本量较少，结果外推性及可信度有待进一步验证。

1.4.2 Logistic回归分析

（1）定义：Logistic 回归是一种广义的线性回归分析模型，是研究分类型因变量与某些影响因素之间关系的一种回归分析方法[41]。在复杂、非随机的真实临床场景中，其可量化变量关联、控制混杂、辅助预测与评价。

（2）实现路径：首先需要根据因变量类型选择对应的回归形式，然后进行模型建立、检验与评价，解读各自变量的偏回归系数、比值比（odds ratio，OR）及其置信区间，明确影响研究结局的因素及影响程度，最终综合模型显著性、拟合评价结果与变量影响分析，总结呈现研究结论。

（3）应用实例：该方法常用于识别独立危险因素并控制混杂、预测个体的结局发生概率等。韩文杰等[42]利用单因素和多因素 Logistic 回归对数据进行分析，构建冠心病患者再次入院的风险预测模型，提供了各因素数字化的精准风险预测。刘兴兴等[43]采用 3 种不同类型的 Logistic 回归对 3 071 例骨关节炎患者是否使用鹿瓜多肽注射液的疗效差异进行分析，为临床该疾病患者合理用药方案提供参考依据。

1.4.3 贝叶斯网络

（1）定义：贝叶斯网络[44]是一种用于计算复杂逻辑、推理因果关系中模糊概率的图模型，能够表示随机变量以及变量间的依赖关系。

（2）实现路径：首先对数据进行预处理，确保数据格式适配模型需求；然后进入结构学习阶段构建有向无环图，若数据量充足，还可采用极大似然估计等方法辅助参数优化；随后进行参数学习，在确定的有向无环图结构下估计各变量的条件概率分布[45]；最终，从统计性能与实际意义双维度进行模型验证与迭代优化。

（3）应用实例：贝叶斯网络应用于临床试验、诊断筛查、风险预测等。Qi 等[46]利用贝叶斯网络分析重度抑郁障碍和代谢综合征的关系，发现了“精神疾病家族史–复发性抑郁症–代谢综合征”的关键影响路径。陈玲等[47]利用贝叶斯网络对 5 825 例老年人失能样本数据构建风险预测模型，该模型可以直观描述老年人失能与影响因素之间的复杂关系，对于发生失能风险较高的因素可以尽早采取针对性的干预措施，降低老年人失能率。

1.4.4 人工神经网络

（1）定义：人工神经网络是一种模拟生物神经系统结构的机器学习模型，由输入层、隐藏层、输出层构成，通过多层非线性变换来学习数据特征与结局的映射关系[48]，适用于处理非线性、多因素交互的复杂问题。

（2）实现路径：人工神经网络的实现首先根据研究目标设计网络结构，输入层为特征变量，隐藏层实现非线性映射，输出层为预测结果；然后使用反向传播算法调整各层权重，通过随机失活神经元、L2 正则化等技术防止过拟合；最后采用交叉验证评估模型泛化能力，通过调整隐藏层节点数、学习率等超参数优化模型性能[49]，实现路径流程图见图 3。

（3）应用实例：人工神经网络在疗效预测、证候分类识别、脉象识别分析、舌象图像处理等方面应用较多[50–52]。叶桦等[53]通过卷积与反向传播神经网络构建 2 型糖尿病中医证候预测模型，

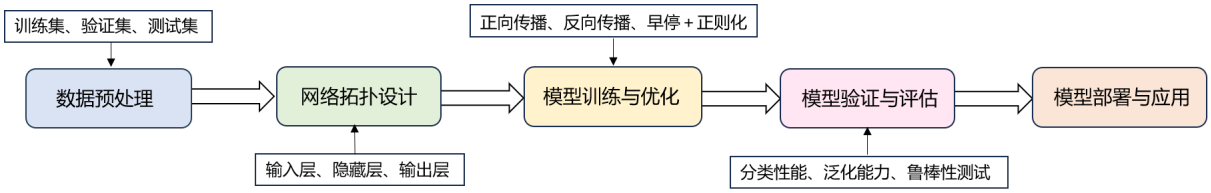

图3 人工神经网络实现流程

Figure 3. Process of implementing artificial neural network

结果显示，反向传播神经网络模型 82.53%、卷积神经网络模型 82.75%，说明该基于卷积神经网络的集成模型能有效提升 2 型糖尿病证候诊断的准确率，为辨证论治提供新思路。王吉庆等 [54] 利用人工神经网络模型验证了真实世界中健康人的舌象特征与年龄、性别的相关性，随着年龄变化，舌象特征会随之改变，性别差异对舌色与苔色也有影响。

1.4.5 其他方法

在真实世界研究中还涉及 LASSO 回归、MGPS 算法等多种研究方法，LASSO 回归是一种在多元线性回归中，通过添加惩罚函数，不断压缩系数，从而达到精简模型，以避免共线性和过拟合的数据挖掘方法，适用于高维数据分析，能够有效处理多重共线性问题。在数据分析过程中，LASSO 回归经常与 Logistic 回归结合使用，如刘丽娜等 [55] 基于两者联用建立并验证了住院患者直接使用口服抗凝药相关问题影响因素的风险预测模型，研究结果显示预测模型效果良好，为临床合理用药提供了参考。刘晓涵等 [56] 在探究糖尿病肾病的危险因素时，采用 LASSO–Logistic 回归构建预测模型，最终识别出糖尿病肾病的关键影响因素。MGPS 算法是一种基于贝叶斯统计的数据挖掘方法。在真实世界研究中多用于药物警戒领域对不良事件的信号挖掘，如 Liu 等 [57] 在甲氨蝶呤相关不良反应研究中，采用包括 MGPS 法在内的多种数据挖掘方法检测不良反应信号，为临床安全用药提供了参考依据。

2 真实世界数据挖掘方法的应用发展与思考

2.1 真实世界常用数据挖掘方法的特点与协同应用

数据挖掘构建了真实世界数据价值转化的技术支撑体系，其方法体系覆盖真实世界数据探索、规律挖掘、预测决策、偏倚控制全流程。在传统统计方法中，描述性统计通过中心趋势、离散程度等指标完成数据概览，为多中心真实世界研究提供基线分析基础，是经验医学向数据医学转化的初始环节。深度挖掘层面，关联规则通过支持度、置信度量化变量之间的隐性关系，为其提供算法支撑；聚类分析基于相似度聚类实现患者亚型划分与处方模式识别，推动真实世界研究从群体分析向精准分型进化；Logistic 回归通过 OR 量化变量与结局的关联强度，是疾病危险因素识别、干预效应初步分析的经典工具；因子分析通过提取公共因子实现高维变量降维，可高效挖掘潜在关联信息，简化复杂数据的解读逻辑；隐结构模型则依托隐变量与显变量的概率映射，解析临床数据中难以直接观测的复杂分型，为中医辨证等领域提供客观化分析工具；人工神经网络凭借多层非线性变换能力，可处理图像、文本等非结构化数据例如舌象特征提取、病历语义解析等，同时在非线性关系预测中展现优势，填补传统统计方法的技术空白；决策树以可解释性树状结构构建临床风险模型，兼顾预测效能与临床可读性；倾向评分通过估算“处理组分配概率”模拟随机化，有效控制观察性数据的混杂偏倚，填补了真实世界干预效应评价的方法学空白；贝叶斯网络则以概率图模型刻画变量间的依赖关系，在小样本、缺失数据场景中仍能实现稳健的因果推理与风险预测，为复杂临床决策提供概率化参考。

在真实世界应用中，单一方法往往难以满足复杂研究需求，大多数情况下需通过多种方法的整合应用发挥协同作用。例如郭春彦等 [58] 利用关联规则和聚类分析对 1 459 例青春期异常子宫出血患儿的病例资料进行数据挖掘，结果显示肾阴亏虚、血热内扰为常见证型，治疗先以止血为主，重在调补脾肾，组方注重止血、补虚、收涩和清热等药物的配伍应用。杨绚如等 [59] 利用倾向评分和决策树模型对丹白颗粒治疗盆腔炎性疾病后遗

症的成本-效果进行分析，结果发现丹白颗粒治疗比常规抗菌药物治疗更具有经济学优势。

2.2 数据挖掘方法在真实世界研究中应用的局限性

尽管上述数据挖掘方法已在药物流行病学导向的真实世界研究中广泛应用，为药物疗效验证、不良反应信号挖掘及个体化用药方案优化提供了技术支撑。但这些方法在真实世界研究中仍面临核心瓶颈。

①方法适配性不足：十余种方法中存在功能交叉，如贝叶斯网络和人工神经网络都是有向图，且模型训练方式相似，但两者技术逻辑存在本质差异，前者是生成模型，结构更加灵活，后者是判别模型，结构是经过标准化的[60]，当前真实世界研究领域尚未建立统一的方法学选择规范，因此导致临床研究者在高维数据处理中容易出现方法和目标错配；②数据保密性和伦理约束：真实世界数据涉及患者个人隐私，《中华人民共和国个人信息保护法》等法规对数据采集、共享、使用的严格限制，使得跨机构数据整合面临阻碍[61]，导致需要多源数据支撑的挖掘方法实现成本较高、推进缓慢[62]；③非结构化数据挖掘深度不足：尽管人工神经网络可通过深度学习处理图像、文本类非结构化数据，例如舌象特征提取、病历实体识别等，但现有研究多停留在表层信息提取，未能深入解析，制约真实世界数据隐性信息激活；④模型可解释性与临床适配性失衡：例如深度神经网络因“黑箱”特性无法提供决策依据[63]，难以满足临床可追溯性需求。所以，为了突破这些局限性，需要从数据标准化、算法改良、因果推断方法创新以及伦理框架的完善多维度协同，才能使数据挖掘真正服务于发现临床规律与生成决策依据。

2.3 关于真实世界研究中数据挖掘方法应用的进一步展望

随着真实世界研究需求的深化，数据挖掘方法在真实世界研究的应用呈现三大趋势：首先是数据挖掘和因果推断的融合[64-65]，当前多数数据挖掘方法聚焦于相关性挖掘，但真实世界研究需回答“因果关系”，例如某干预措施是否真能改善患者预后，未来需将因果推断框架与数据挖掘模型结合，实现“从关联发现到因果验证”的跨越，提升研究结论的临床决策价值。其次是个性化挖掘的落地，针对真实世界中患者个体差异大的特点，开发面向个性化医疗的挖掘方法——例如基于患者的基因数据、生活习惯数据、临床数据构建个体预测模型，通过实时数据更新动态调整模型参数，实现对个体疾病风险、治疗结果的精准预测，支撑个性化干预方案的制定[66-67]。最后是数据多类型结合实时挖掘的拓展，随着可穿戴设备、物联网设备的普及，真实世界数据将进一步向“多类型、实时化”方向发展[68]，未来需研发支持多类型数据实时处理的挖掘技术，例如边缘计算与深度学习结合的实时分析模型，实现对疾病的“动态监测-早期预警-及时干预”，例如通过实时分析心电信号数据，提前预测心血管疾病发作风险[69]，推动真实世界研究从回顾性分析向前瞻性干预转型。

3 结语

本文梳理了真实世界研究中十余种数据挖掘方法，涵盖传统统计与深度挖掘，明确其覆盖数据探索至模型预测全过程，并通过临床案例印证多方法整合的协同作用。同时也需正视方法落地瓶颈：方法适配性不足、数据伦理约束、非结构化数据挖掘深度有限、模型可解释性与临床需求失衡，这些问题制约了数据价值的充分实现。数据挖掘方法在真实世界研究中的应用趋势包括：融合因果推断实现“关联到因果”的跨越、落地个性化挖掘支撑精准医疗、拓展多类型数据实时分析推动前瞻性干预。唯有从数据标准化、算法改良、伦理框架完善等多维度协同突破，才能让数据挖掘真正服务于临床规律发现与真实世界证据生成，为医药卫生领域发展提供更强技术支撑。

利益冲突声明：作者声明本研究不存在任何经济或非经济利益冲突。

参考文献

1 郭方达，李京儒，胡洁，等. 基于真实世界数据的可穿戴监护设备在门急诊患者中的临床价值研究 [J]. 中国医学装备，2025, 22(8): 11-16. [Guo FD, Li JR, Hu J, et al. A study on the clinical value of a wearable monitoring device based on real-world data in patients of outpatient and emergent department[J]. China Medical Equipment, 2025, 22(8): 11-16.] DOI: 10.3969/j.issn.1672-8270.2025.08.003.

2 江旻, 袁延楠, 刘晓红, 等. 基于电子病历的抗肿瘤药物真实世界外对照臂研究适用性评价框架的考量与建立 [J]. 中国医院药学杂志, 2024, 44(12): 1457-1462. [Jiang M, Yuan YN, Liu XH, et al. Practices and considerations for feasibility evaluations of real-world external control arm for antineoplas-tic drugs based upon electronic medical records[J]. Chinese Journal of Hospital Pharmacy, 2024, 44 (12): 1457-1462.] DOI: 10.13286/j.1001-5213.2024.12.15.

3 Lavertu A, Vora B, Giacomini KM, et al. A new era in pharmacovigilance: toward real-world data and digital monitoring[J]. Clin Pharmacol Ther, 2021, 109(5): 1197-1202. DOI: 10.1002/cpt.2172.

4 U.S. Congress. H.R.34-21st century cures act.114th congress (2015-2016)[S/OL]. (2016-12-13) [2025-10-05]. https://www.congress.gov/bill/114th-congress/house-bill/34.

5 FDA. Use of electronic health record data in clinical investigations[R/OL]. (2018-07-19) [2025-10-05]. https://www.fda.gov/media/97567/download.

6 U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Oncology Center for Excellence. Real-world data: assessing electronic health records and medical claims data to support regulatory decision-making for drug and biological products: guidance for industry[R/OL]. (2024-07-01) [2025-10-05]. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/real-world-data-assessing-electronic-health-records-and-medical-claims-data-support-regulatory.

7 国家药品监督管理局. 真实世界证据支持药物研发与审评的指导原则（试行）[EB/OL]. (2020-01-07) [2021-12-16]. https://www.nmpa.gov.cn/xxgk/ggtg/ypggtg/ypqtggtg/20200107151901190.html.

8 Little RJ, D'Agostino R, Cohen ML, et al. The prevention and treatment of missing data in clinical trials[J]. N Engl J Med, 2012, 367(14): 1355-1360. DOI: 10.1056/NEJMsr1203730.

9 韩梅, 郭蓉娟, 夏芸. 真实世界注册研究设计要素及伦理的研究进展 [J]. 重庆医学, 2025, 54(10): 2431-2436. [Han M, Guo RJ, Xia Y. Research progress on the design elements and ethics of real-world registry studies[J]. Chongqing Medicine, 2025, 54(10): 2431-2436.] DOI: 10.3969/j.issn.1671-8348.2025.10.034.

10 朱凌云, 吴宝明, 曹长修. 医学数据挖掘的技术、方法及应用 [J]. 生物医学工程学杂志, 2003, 20(3): 559-562. [Zhu LY, Wu BM, Cao CX. Introduction to medical data mining[J]. Journal of Biomedical Engineering, 2003, 20(3): 559-562.] DOI: 10.3321/j.issn:1001-5515.2003.03.047.

11 Tseng MH, Wu HC. Investigating health equity and healthcare needs among immigrant women using the association rule mining method[J]. Healthcare (Basel), 2021, 9(2): 195. DOI: 10.3390/healthcare9020195.

12 毕建欣, 张岐山. 关联规则挖掘算法综述 [J]. 中国工程科学, 2005, 7(4): 88-94. [Bi JX, Zhang QS. Survey of the algorithms on association rule mining[J]. Engineering Science, 2005, 7(4): 88-94.] DOI: 10.3969/j.issn.1009-1742.2005.04.016.

13 Kim SS, Kim HS. The impact of the association between cancer and diabetes mellitus on mortality[J]. Pers Med, 2022, 12(7): 1099. DOI: 10.3390/jpm12071099.

14 施雪清, 沈锡容, 熊梦欣, 等. 基于医院信息系统数据挖掘真实世界多囊卵巢综合征中药用药规律 [J]. 新中医, 2024, 56(20): 15-20. [Shi XQ, Shen XR, Xiong MX, et al. Study on Chinese medicinal medication rules of treating polycystic ovary syndrome based on real-world data mining of hospital information system[J]. Journal of New Chinese Medicine, 2024, 56 (20): 15-20.] DOI: 10.13457/j.cnki.jncm.2024.20.003.

15 Oyewole GJ, Thopil GA. Data clustering: application and trends[J]. Artif Intell Rev, 2023, 56(7): 6439-6475. DOI: 10.1007/s10462-022-10325-y.

16 章永来, 周耀鉴. 聚类算法综述 [J]. 计算机应用, 2019, 39(7): 1869-1882. [Zhang YL, Zhou YJ. Review of clustering algorithms[J]. Journal of Computer Applications, 2019, 39(7): 1869-1882.] DOI: 10.11772/j.issn.1001-9081.2019010174.

17 董芷欣. 基于 Python 聚类分析的聚类数确定方法对比 [J]. 微型电脑应用, 2023, 39(12): 220-223. [Dong ZX. Comparative analysis of cluster number determination based on python cluster analysis[J]. Microcomputer Applications, 2023, 39(12): 220-223.] DOI: 10.3969/j.issn.1007-757X.2023.12.057.

18 张迪, 陈艺幻, 张伟娜. 常用关联与聚类分析方法对中医处方数据的适用性探讨 [J]. 中草药, 2025, 56(11): 3974-3984. [Zhang D, Chen YH, Zhang WN. Discussion on applicability of common association and cluster analysis methods to traditional Chinese medicine prescription data[J]. Chinese Traditional and Herbal Drugs, 2025, 56(11): 3974-3984.] DOI: 10.7501/j.issn.0253-2670.2025.11.018.

19 李家劼, 申国明, 刘金星, 等. 基于因子分析与聚类分析的糖尿病周围神经病变合并血脂异常中医证候学特征研究 [J]. 中华中医药杂志, 2024, 39(11): 6132-6139. [Li JJ, Shen GM, Liu JX, et al. Research on traditional Chinese medicine syndrome characteristics of diabetes peripheral neuropathy with dyslipidemia based on factor analysis and cluster analysis[J]. China Journal of Traditional Chinese Medicine and Pharmacy, 2024, 39(11): 6132-6139.] https://d.wanfangdata.com.cn/periodical/CiBQZXJpb2RpY2FsQ0hJU29scjkyMDI2MDMwNjE2NTI1NxIPemd5eXhiMjAyNDExMDk1Ggh1NzVobnd0NQ%3D%3D.

20 Min Q, Wu Z, Yao J, et al. Association between atherogenic index of plasma control level and incident cardiovascular disease in middle-aged and elderly Chinese individuals with abnormal glucose metabolism[J]. Cardiovasc Diabetol, 2024, 23(1): 54. DOI: 10.1186/s12933-024-02144-y.

21 （美）金在温, 查尔斯·W·米勒著, 叶华译. 因子分析统计方法与应用问题 [M]. 上海: 格致出版社, 2023: 1-5.

22 吕咪, 车慧, 周秉舵, 等. 基于因子分析和聚类分析的 800 例非糜烂性反流病与上腹痛综合征胃肠症状重叠患者中医证型横断面研究 [J]. 中国中医药信息杂志, 2025, 32(9): 141-148. [Lyu M, Che H, Zhou BD, et al. Cross-sectional study on TCM syndromes of 800 patients with overlapping

gastrointestinal symptoms of NERD and EPS based on factor analysis and clustering analysis[J]. Chinese Journal of Information on Traditional Chinese Medicine, 2025, 32(9): 141–148.] DOI: 10.19879/j.cnki.1005–5304.202503321.

23 江雯婷 , 黄嬿 , 常昕楠 , 等 . 基于药性理论及因子分析法分析中药饮片治疗 IgA– 肾病的用药规律 [J]. 中国医药导刊 , 2022, 24(9): 843–849. [Jiang WT, Huang Y, Chang XN, et al. Analysis of the medication rule of Chinese herbal decoction pieces in the treatment of IgA nephropathy based on Chinese herbal property theory and factor analysis[J]. Chinese Journal of Medical Guide, 2022, 24(9): 843–849.] DOI: 10.3969/j.issn.1009–0959.2022.09.002.

24 香港科技大学 . 隐结构模型与中医证候研究 [EB/OL]. (2014–12–05) [2024–12–11]. https://www.cse.ust.hk/lzhang/tcm/resource.html.

25 李宇迪 , 丁樱 , 徐炎 , 等 . 基于隐结构模型的《幼科发挥》“药 – 证 – 方” 规律 [J]. 世界中医药 , 2025, 20(4): 606–612. [Li YD, Ding Y, Xu Y, et al. Rules of "medicine–syndrome–prescription" in Youke Fahui based on latent structure model[J]. World Chinese Medicine, 2025, 20(4): 606–612.] https://link.cnki.net/urlid/11.5529.R.20250506.1231.014.

26 袁君 , 丁霄 , 徐洁 , 等 . 基于隐结构模型和关联规则探讨非酒精性脂肪肝病的用药规律 [J]. 世界中西医结合杂志 , 2025, 20(7): 1303–1311. [Yuan J, Ding X, Xu J, et al. Medication patterns in non–alcoholic fatty liver disease based on latent structure model and association rules[J]. World Journal of Integrated Traditional and Western Medicine, 2025, 20(7): 1303–1311.] DOI: 10.13935/j.cnki.sjzx.250705.

27 李伟珂 , 伊明洋 , 倪园园 , 等 . 基于潜在类别结合隐结构模型的慢性阻塞性肺疾病肺气虚证分级量化诊断研究 [J]. 中医杂志 , 2025, 66(7): 710–716. [Li WK, Yi MY, Ni YY, et al. Study on graded quantitative diagnosis of lung qi deficiency syndrome in chronic obstructive pulmonary disease based on latent class analysis combined with hidden structure model[J]. Journal of Traditional Chinese Medicine, 2025, 66 (7): 710–716.] DOI: 10.13288/j.11–2166/r.2025.07.010.

28 李俊雅 , 史阳琳 , 杨建雅 , 等 . 基于真实世界数据探讨中医药治疗非小细胞肺癌的用药规律 [J]. 药物流行病学杂志 , 2025, 34(4): 398–409. [Li JY, Shi YL, Yang JY, et al. Exploration on medication pattern of traditional Chinese medicine treatment for non–small cell lung cancer based on real world data[J]. Chinese Journal of Pharmacoepidemiology, 2025, 34(4): 398–409.] DOI: 10.12173/j.issn.1005–0698.202408072.

29 Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects[J]. Biometrika, 1983, 70(1): 41–55. DOI: 10.1093/biomet/70.1.41.

30 岳青青 , 焦志刚 , 凡如 , 等 . 倾向性评分法简介及其SAS实现[J]. 中国卫生统计 , 2021, 38(1): 144–147. [Yue QQ, Jiao ZG, Fan R, et al. Introduction to propensity score method and its implementation in SAS[J]. Chinese Journal of Health Statistics, 2021, 38(1): 144–147.] DOI: 10.3969/j.issn.1002–3674. 2021.01.038.

31 施婷婷 , 刘振球 , 袁黄波 , 等 . 倾向性评分匹配法在非随机对照研究中的应用 [J]. 中国卫生统计 , 2021, 38(2): 312–314. [Shi TT, Liu ZQ, Yuan HB, et al. The application of propensity score matching method in non–randomized controlled studies[J]. Chinese Health Statistics, 2021, 38(2): 312–314.] DOI: 10.3969/j.issn.1002–3674.2021.02.040.

32 袁博寒 , 李亚妤 , 杨亚珍 . 基于真实世界的百令片治疗慢性肾脏病疗效评价 [J]. 中国现代应用药学 , 2024, 41(23): 3316–3321. [Yuan BH, Li YY, Yang YZ. Efficacy evaluation of Bailing tablets in the treatment of chronic kidney disease based on real–world data[J]. Chinese Journal of Modern Applied Pharmacy, 2024, 41(23): 3316–3321.] DOI: 10.13748/j.cnki.issn1007–7693.20242926.

33 朱金霞 , 刘光伟 , 张小瑞 , 等 . 基于倾向性评分的真实世界扶正散结方治疗乙肝相关肝癌的疗效评价 [J]. 中医肿瘤学杂志 , 2023, 5(6): 7–16. [Zhu JX, Liu GW, Zhang XR, et al. Evaluation of the real–world therapeutic effect of Fuzheng Sanjie prescription for hepatitis B related liver cancer based on propensity score[J]. Journal of Oncology in Chinese Medicine, 2023, 5(6): 7–16.] DOI: 10.19811/j.cnki.ISSN2096–6628.2023.11.002.

34 高洪阳 , 赵阳 , 盛松 . 基于真实世界观察参附注射液辅助治疗心力衰竭的临床疗效 [J]. 中西医结合心脑血管病杂志 , 2023, 21(20): 3696–3702. [Gao HY, Zhao Y, Sheng S. The clinical effect of shenfu injection in the adjuvant treatment of heart failure based on real world[J]. Chinese Journal of Integrative Medicine on Cardio/Cerebrovascular Disease, 2023, 21(20): 3696–3702.] DOI: 10.12102/j.issn.1672–1349.2023.20.004.

35 成冯镜茗 , 谢雁鸣 , 厉将斌 , 等 . 基于医疗电子数据的脉络舒通丸对疑似胃肠道不良反应影响的真实世界研究 [J]. 中国中医基础医学杂志 , 2022, 28(9): 1474–1479. [Cheng FJM, Xie YM, Li JB, et al. Real–world study on the effect of Mailuo Shutong pill on suspected gastrointestinal adverse reactions based on medical electronic data[J]. Chinese Journal of Basic Medicine in Traditional Chinese Medicine, 2022, 28(9): 1474–1479.] DOI: 10.19945/j.cnki.issn.1006–3250.2022.09.015.

36 张棪 , 曹健 . 面向大数据分析的决策树算法 [J]. 计算机科学 , 2016, 43(S1): 374–379, 383. [Zhang Y, Cao J. Decision tree algorithm for big data analysis[J]. Computer Science, 2016, 43(S1): 374–379, 383.] https://d.wanfangdata.com.cn/periodical/CiBQZXJpb2RpY2FsQ0hJU29scjkyMDI2MDMwNjE2NTI1NxIOanNqa3gyMDE2ejEwODkaCHl3ajQzZndh.

37 马红丽 , 徐长英 , 杨新鸣 . 决策树模型在中医药领域的应用现状 [J]. 世界中医药 , 2021, 16(17): 2648–2651, 2656. [Ma HL, Xu CY, Yang XM. Application status of decision tree model in the field of traditional Chinese medicine[J]. World Chinese Medicine, 2021, 16(17): 2648–2651, 2656.] DOI: 10.3969/j.issn.1673–7202.2021.17.025.

38 Nielsen ML, Petersen TC, Maul LV, et al. Predicting psoriatic arthritis in psoriasis patients–a Swiss registry study[J]. J Psoriasis Psoriatic Arthritis, 2024, 9(2): 41–50. DOI: 10.1177/24755303231217492.

39 石玉琳 , 刘嘉懿 , 胡晓娟 , 等 . 基于舌脉象数据的决策树算法的非小细胞肺癌证候分类方法 [J]. 世界科学技术 – 中医药

现代化 , 2022, 24(7): 2766–2775. [Shi YL, Liu JY, Hu XJ, et al. Classification of syndromes of non–small cell lung cancer based on decision tree algorithm based on tongue data and pulse data[J]. World Science and Technology–Modernization of Traditional Chinese Medicine, 2022, 24(7): 2766–2775.] DOI: 10.11842/wst.20210705003.

40 辛雅雯 , 卓玛层 , 师永兰 , 等 . 基于真实世界的治疗耐碳青霉烯病原菌抗菌药物方案的药物经济学评价 [J]. 中国药物经济学 , 2024, 19(3): 5–10. [Xin YW, Zhuo MC, SHI YL, et al. Pharmacoeconomic evaluation of antimicrobial treatment for carbapenem–resistant organism according to real world data[J]. China Journal of Pharmaceutical Economics, 2024, 19(3): 5–10.] DOI: 10.12010/j.issn.1673–5846.2024.03.001.

41 陈炳为 , 主编 . 医学统计学 , 第 4 版 [M]. 南京 : 南京东南大学出版社 , 2023: 268.

42 韩文杰 , 朱明军 , 王新陆 , 等 . 冠心病稳定型心绞痛患者再入院中西医风险预测模型构建与应用评估——基于真实世界临床数据的前瞻性研究 [J]. 中医杂志 , 2025, 66(6): 604–611. [Han WJ, Zhu MJ, Wang XL, et al. Construction and application evaluation of integrated traditional Chinese and Western medicine risk prediction model for readmission of patients with stable angina pectoris of coronary heart disease–a prospective study based on real–world clinical data[J]. Journal of Traditional Chinese Medicine, 2025, 66(6): 604–611.] DOI: 10.13288/j.11–2166/r.2025.06.010.

43 刘兴兴 , 黎元元 , 魏戌 , 等 . 鹿瓜多肽注射液治疗 3 071 例骨关节炎真实世界疗效分析 [J]. 世界中医药 , 2021, 16(7): 1126–1133. [Liu XX, Li YY, Wei X, et al. Efficacy analysis of Lugua polypeptide injection in the treatment of 3,071 cases of osteoarthritis in real world[J]. World Chinese Medicine, 2021, 16(7): 1126–1133.] DOI: 10.3969/j.issn.1673–7202.2021.07.020.

44 Kitson NK, Constantinou AC, Guo Z, et al. A survey of Bayesian network structure learning[J]. Artif Intell Rev, 2023, 56: 8721–8814. DOI: 10.1007/s10462–022–10351–w.

45 Beresniak A, Bertherat E, Perea W, et al. A Bayesian network approach to the study of historical epidemiological databases: modelling meningitis outbreaks in the Niger[J]. Bull World Health Organ, 2012, 90(6): 412–417A. DOI: 10.2471/BLT.11.086009.

46 Qi H, Liu R, Dong CC, et al. Identifying influencing factors of metabolic syndrome in patients with major depressive disorder: a real–world study with Bayesian network modeling[J]. J Affect Disord, 2024, 362: 308–316. DOI: 10.1016/j.jad.2024.07.004.

47 陈玲 , 郝志梅 , 魏霞霞 , 等 . 基于贝叶斯网络的老年人失能风险预测模型构建 [J]. 中国老年学杂志 , 2023, 43(22): 5596–5600. [Chen L, Hao ZM, Wei XX, et al. Construction of disability risk prediction model for the elderly based on Bayesian network[J]. Chinese Journal of Gerontology, 2023, 43(22): 5596–5600.] DOI: 10.3969/j.issn.1005–9202.2023.22.059.

48 常强 , 赵伟 , 赵仰杰 . 基于神经网络的数据分类预测与实现[J]. 软件, 2018, 39(12): 207–209. [Chang Q, Zhao W, Zhao YJ. Prediction and implementation of data classification based on neural network[J]. Computer Engineering & Software, 2018, 39(12): 207–209.] DOI: 10.3969/j.issn.1003–6970.2018.12.047.

49 许朝霞 , 王忆勤 , 颜建军 , 等 . 基于支持向量机和人工神经网络的心血管疾病中医证候分类识别研究 [J]. 北京中医药大学学报 , 2011, 34(8): 539–543. [Xu CX, Wang YQ, Yan JJ, et al. Study on classification and identification of TCM syndromes of cardiovascular diseases based on support vector machine and artificial neural network[J]. Journal of Beijing University of Chinese Medicine, 2011, 34(8): 539–543.] https://d.wanfangdata.com.cn/periodical/CiBQZXJpb2RpY2FsQ0hJU29scjkyMDI2MDMwNjE2NTI1NxISYmp6eXlkeHhiMjAxMTA4MDA5GghpY2I4ZDc5Ng%3D%3D.

50 郭红霞 , 师义民 . 中医脉象的 BP 神经网络分类方法研究 [J]. 计算机工程与应用 , 2005, 41(32): 187–189. [Guo HX, Shi YM. Research on BP neural network classification method for TCM pulse conditions[J]. Computer Engineering and Applications, 2005, 41(32): 187–189.] DOI: 10.3321/j.issn:1002–8331.2005.32.059.

51 李秋华 , 史国峰 , 李玥博 , 等 . 基于卷积神经网络的“舌边白涎”舌象识别研究 [J]. 湖南中医药大学学报 , 2024, 44(7): 1254–1260. [Li QH, Shi GF, Li YB, et al. Research on tongue image recognition of "white saliva on tongue edge" based on convolutional neural network[J]. Journal of Hunan University of Chinese Medicine, 2024, 44(7): 1254–1260.] DOI: 10.3969/j.issn.1674–070X.2024.07.016.

52 罗爱静 , 王哲轩 , 谢文照 , 等 . 基于神经网络的甲状腺肿瘤复发风险评估模型 [J]. 中国医学物理学杂志 , 2025, 42(7): 974–980. [Luo AJ, Wang ZX, Xie WZ, et al. Risk assessment model for thyroid tumor recurrence based on neural network[J]. Chinese Journal of Medical Physics, 2025, 42(7): 974–980.] DOI: 10.3969/j.issn.1005–202X.2025.07.020.

53 叶桦 , 何黎 , 胡远樟 , 等 . 基于卷积神经网络的 2 型糖尿病证候分布演化规律研究 [J]. 时珍国医国药 , 2021, 32(6): 1522–1524. [Ye H, He L, Hu YZ, et al. Study on the evolution law of TCM syndrome distribution in type 2 diabetes mellitus based on convolutional neural network[J]. Lishizhen Medicine and Materia Medica Research, 2021, 32(6): 1522–1524.] DOI: 10.3969/j.issn.1008–0805.2021.06.70.

54 王吉庆 , 张蕾 , 徐世芬 , 等 . 基于机器学习的真实世界健康人舌象与年龄及性别相关性研究 [J]. 世界科学技术 – 中医药现代化 , 2024, 26(11): 2806–2814. [Wang JQ, Zhang L, Xu SF, et al. Study on the correlation between machine learning–based tongue features of healthy individuals in the real world and age and gender[J]. World Science and Technology–Modernization of Traditional Chinese Medicine, 2024, 26(11): 2806–2814.] DOI: 10.11842/wst.20240115012.

55 刘丽娜 , 宋奇修 , 张伦 , 等 . 真实世界下直接口服抗凝药药物相关问题的 Lasso–Logistic 回归分析及列线图预测模型构建 [J]. 实用药物与临床 , 2025, 28(6): 406–412. [Liu LN, Song QX, Zhang L, et al. Lasso–Logistic regression analysis of drug–related problems of direct oral anticoagulants in real world and construction of nomogram prediction model[J]. Practical Pharmacy and Clinical Remedies, 2025, 28(6): 406–412.] DOI: 10.14053/j.cnki.ppcr.202506002.

56 刘晓涵 , 郑宇钰 , 李瑞雪 , 等 . 基于 LASSO 回归分析的糖尿病肾病风险预测模型的构建 [J]. 中国社会医学杂志 , 2025, 42(3): 363–368. [Liu XH, Zheng YY, Li RX, et al. Construction of risk prediction model for diabetic nephropathy based on LASSO regression analysis[J]. Chinese Journal of Social Medicine, 2025, 42(3): 363–368.] DOI: 10.3969/j.issn.1673–5625.2025.03.023.

57 Liu S, Yuan Z, Rao S, et al. Adverse drug reactions related to methotrexate: a real–world pharmacovigilance study using the FAERS database from 2004 to 2024[J]. Front Immunol, 2025, 16: 1586361. DOI: 10.3389/fimmu.2025.1586361.

58 郭春彦 , 柳静 , 张萌 , 等 . 基于真实世界数据的青春期异常子宫出血用药规律研究 [J]. 世界中医药 , 2024, 19(10): 1469–1475, 1485. [Guo CY, Liu J, Zhang M, et al. Medication rules of TCM for abnormal uterine bleeding during puberty based on real world data mining[J]. World Chinese Medicine, 2024, 19(10): 1469–1475, 1485.] DOI: 10.3969/j.issn.1673–7202.2024.10.016.

59 杨绚如 , 薛晓鸥 , 朱玉莹 , 等 . 丹白颗粒治疗盆腔炎性疾病后遗症的成本 – 效果分析 [J]. 世界中医药 , 2024, 19(12): 1820–1825. [Yang XR, Xue XO, Zhu YY, et al. Cost–effectiveness analysis of Danbai granules in the treatment of sequelae of pelvic inflammatory disease[J]. World Chinese Medicine, 2024, 19(12): 1820–1825.] DOI: 10.3969/j.issn.1673–7202.2024.12.018.

60 李妍 . 基于贝叶斯网络和 BP 神经网络预测用户使用 App 行为的分析与研究 [D]. 北京 : 北京邮电大学 , 2018.

61 葛永彬 , 董剑平 . 利用医疗大数据开展真实世界临床研究的合规性要求 [J]. 中国食品药品监管 , 2023(10): 86–94.[Ge YB, Dong JP. Compliance requirements for real–world clinical research using medical big data[J]. China Food and Drug Administration Magazine, 2023(10): 86–94.] DOI: 10.3969/j.issn.1673–5390.2023.10.010.

62 弓孟春 , 陆亮 . 医学大数据研究进展及应用前景 [J]. 医学信息学杂志 , 2016, 37(2): 9–15. [Gong MC, Lu L. Research progress and application prospect of medical big data[J]. Journal of Medical Informatics, 2016, 37(2): 9–15.] DOI: 10.3969/j.issn.1673–6036.2016.02.002.

63 袁培江 , 苏峰 . 两个黑箱问题——深度神经网络和脑神经网络 [J]. 科技导报 , 2017, 35(18): 12. [Yuan PJ, Su F. Two black box problems: deep neural network and brain neural network[J]. Science & Technology Review, 2017, 35(18): 12.] https://d.wanfangdata.com.cn/periodical/CiBQZXJpb2RpY2FsQ0hJU29scjkyMDI2MDMwNjE2NTI1NxINa2pkYjIwMTcxODAwNhoIcml4Z2h1MjY%3D.

64 卢存存 , 陈子佳 , 张强 , 等 . 基于真实世界数据的目标试验模拟研究 : 现状与展望 [J]. 中国循证医学杂志 , 2023, 23(4): 492–496. [Lu CC, Chen ZJ, Zhang Q, et al. Target trial emulation based on real–world data: current status and prospects[J]. Chinese Journal of Evidence–Based Medicine, 2023, 23(4): 492–496.] https://kns.cnki.net/kcms2/article/abstract?v=W–y7fKBVJCzVfuhoDu5QtfJMfeMFXKoylM1sR3vZrLEzXogOJp7yGLqjQfF4BD9uZHq_hcQQfKtoIgI7450rXbJ0CPoCkJc1vy0WJoOZgIoEz528tKhMpDQpwfWP6QDHiGpvGFXdxk8glM30U_ffC8k8BcwoqO5vdViHnrAiFdcLyL0wI73R4g==&uniplatform=NZKPT&language=CHS.

65 Schuler MS, Rose S. Targeted maximum likelihood estimation for causal inference in observational studies[J]. Am J Epidemiol, 2017, 185(1): 65–73. DOI: 10.1093/aje/kww165.

66 宋雨昕 , 叶倩 , 赵盟生 , 等 . 疾病风险动态预测模型方法前沿进展与精准预防 [J]. 科技导报 , 2024, 42(12): 75–91. [Song YX, Ye Q, Zhao MS, et al. Frontier progress and precise prevention of dynamic disease risk prediction model methods[J]. Science & Technology Review, 2024, 42(12): 75–91.] DOI: 10.3981/j.issn.1000–7857.2024.05.00543.

67 Yoo SK, Fitzgerald CW, Cho BA, et al. Prediction of checkpoint inhibitor immunotherapy efficacy for cancer using routine blood tests and clinical data[J]. Nat Med, 2025, 31(3): 869–880. DOI: 10.1038/s41591–024–03398–5.

68 Jafleh EA, Alnaqbi FA, Almaeeni HA, et al. The role of wearable devices in chronic disease monitoring and patient care: a comprehensive review[J]. Cureus, 2024, 16(9): e68921. DOI: 10.7759/cureus.68921.

69 余新艳 , 赵旭东 , 赵晓晔 , 等 . 基于社区移动医疗的心律失常筛查方案真实世界研究 [J]. 中国全科医学 , 2023, 26(2): 192–200, 209. [Yu XY, Zhao XD, Zhao XY, et al. Real–world study on arrhythmia screening program based on community mobile health[J]. Chinese General Practice, 2023, 26(2): 192–200, 209.] DOI: 10.12114/j.issn.1007–9572.2022.0511.

收稿日期：2025 年 11 月 07 日 修回日期：2026 年 03 月 05 日
本文编辑：冼静怡 杨 燕