

# 数据挖掘技术在医学领域的应用综述

阿依波塔·沙尔木哈买提<sup>1</sup>, 安冬青<sup>2\*</sup>

<sup>1</sup>新疆医科大学, 中医学院, 新疆 乌鲁木齐

<sup>2</sup>新疆医科大学, 新疆名医名方与特色方剂学实验室, 新疆 乌鲁木齐

收稿日期: 2023年11月27日; 录用日期: 2023年12月21日; 发布日期: 2023年12月28日

## 摘要

医疗大数据的复杂性源于疾病及其合并症的多样性; 治疗和疗效的异质性; 医疗数据获取、处理、解读以及研究设计和分析方法的复杂性。由于DM具有处理海量、非线性复杂信息以及发现庞大的数据背后隐藏的巨大潜在价值的优势, 应用于医疗数据挖掘有望提高医疗质量、优化医疗流程和管理策略。本文概述数据挖掘各个算法的同时从现代医学和祖国传统医学等方面入手综述数据挖掘技术在医疗领域的应用。

## 关键词

数据挖掘, 医疗领域的应用, 综述

# A Review of the Application of Data Mining Technology in Medical Field

Ayibota·Shaermuhamaiti<sup>1</sup>, Dongqing An<sup>2\*</sup>

<sup>1</sup>Institute of TCM, Xinjiang Medical University, Urumqi Xinjiang

<sup>2</sup>Xinjiang Famous Medicine and Characteristic Formulary Laboratory, Xinjiang Medical University, Urumqi Xinjiang

Received: Nov. 27<sup>th</sup>, 2023; accepted: Dec. 21<sup>st</sup>, 2023; published: Dec. 28<sup>th</sup>, 2023

## Abstract

The complexity of medical big data stems from the diversity of diseases and their comorbidities; heterogeneity of treatment and efficacy; complexity of medical data acquisition, processing, interpretation, research design and analysis methods. Since DM has the advantages of processing massive, non-linear and complex information and discovering huge potential value behind huge

\*通讯作者。

**data, the application of medical data mining is expected to improve medical quality, optimize medical process and management strategy. This paper summarizes the application of data mining technology in the medical field from the aspects of modern medicine and traditional medicine of the motherland.**

## Keywords

**Data Mining, Applications in the Medical Field, Review**

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 数据挖掘概念与技术

数据挖掘(Data Mining, DM)是从大型数据中挖掘有趣模式和知识的过程, 又称之为知识发现(Knowledge Discovery from Database, KDD)。提取的知识一般可表示为概念、规则、规律、模式等形式, 是一门涉及到统计学、机器学习、模式识别、数据库和数据仓库、信息检索、可视化等不同领域的交叉性学科[1]。数据挖掘有描述性和预测性两种任务, 预测性任务就是根据其他属性的值来预测特定属性的值, 分类属于预测任务; 描述性任务就是概括数据中潜在的联系模式(如相关性、趋势、聚类、轨迹和异常等) [2]。简而言之, 数据挖掘技术是通过关联、分类、聚类等方法实现对目标数据的描述与预测任务, 常见算法包括 Apriori 算法、K-means 算法、决策树, 朴素贝叶斯, 支持向量机(SVM), 以及人工神经网络、遗传算法等。下面简单介绍几种常见算法原理及应用[3] [4]。

关联规则(Association Rules)——关联规则分析是从大量数据中发现项集之间的有趣关联, 其中支持度和置信度是较为重要的概念, 其中支持度反应规则的普遍性, 置信度反应规则的可靠性[5]。Apriori 算法是其重要的算法。作为关联规则的来源, 销售领域是关联规则应用最重要的领域之一, 在销售的相关性分析中为提高零售企业产品配置合理性起到重要作用[6], 其原理是挖掘顾客购买的商品之间各种有趣联系, 帮助商家制定营销策略, 最终向每个顾客提供一组正确的推荐信息, 从而改善客户的购物体验, 增加总的销售额。[7]。关联规则在销售领域应用不断更新, 如在需求关联性对库存分配影响性研究中来为库存管理提供了一定的参考等[8] [9]。同样被应用于信息技术、医药、教育等各个领域。如关联规则应用于中医病因病机、病症、证候及药物之间关系, 名老中医经验传承, 古今医案类方研究以及针灸处方、中药药性分析等方面的应用[10]。如关联规则在图书馆信息服务中的应用[11]; 如应用改进 K-means 和引入兴趣度的 Apriori 的学生课程成绩分析为教学方案的设计和进步提供参考, 还有助于提高教学质量和学生的学习学习质量[12]。

聚类分析(Clustering Analysis)——聚类通过静态分类的过程将相似的对象放入不同的类别或子集。因此, 同一子集内的对象具有相似的性质。聚类分析常用分割聚类(K-means 算法)、分层群聚(BIRCH、CURE、ROCK)、密度算法、网络聚类算法[13]。应用于经济与金融、医药、公共管理等领域。例如农业保险对农民收入的影响的区域分析; 采用聚类分析方法, 对中国 47 个城市的竞争力进行聚类分析, 并根据结果把城市分成五大阵营, 同时总结出相应的基本特点[14]; 中药黄芪属植物物种鉴定及亲缘关系研究[15]。

决策树(Decision Tree)——是一种常用于预测模型的算法, 它通过将大量数据有目的地分类, 从中找到一些具有价值的, 潜在的信息。同时挖掘中一项重要的任务是对数据进行分类, 而决策树是分类算

法中一个最主要而且应用最广泛的算法分支[16][17]。目前广泛应用于经济学、生态学、企业信用评估、医疗等很多领域。如随机森林算法用于分析肿瘤相关氨基酸序列、提取肿瘤代谢物的特征信息、医学影像学图像分割、预测判断蛋白质作用位点等方面[18]; C5.0 算法用于建立预测模型、预测疾病发生风险及危重疾病生存预测等方面[19]。如银行信用卡评估[20]。

人工神经网络(Neural Network)——人工神经网络最初由大脑中的神经网络启发而来, 由层层相互连接的计算单元(神经元)组成[21]。2006 年 Hinton 等提出了深度学习的概念, 利用多层神经网络从大量训练数据中自动学习高层特征。同时掀起了人工神经网络的又一热潮。深度学习因较好的自学习[22]、建模能力较强[23]的鲁棒性能, 受到学界的广泛关注[24][25]。在目标检测和计算机视觉、自然语言处理、语音识别和语义分析等领域成效卓然。深度学习技术在医学图像分析的许多领域, 如在脑部 MRI 分割、辅助检测乳腺病变, 以及组织病理学诊断等方面越来越受欢迎[26]。在其他领域应用于各类风险评估。

遗传算法(Genetic Algorithm)是模拟自然界生物进化过程与机制求解问题的一类自组织与自适应的人工智能技术[27]。遗传算法提供了一种求解复杂系统优化问题的通用框架, 它不依赖于问题的具体领域, 广泛应用于多种学科领域。遗传算法广泛应用于自动控制、计算科学、模式识别、工程设计、智能故障诊断、管理科学和社会科学等领域, 适用于解决复杂的非线性和多维空间寻优问题。如在水污染控制规划、水库(群)调度、水环境模型参数估计中的应用。在故障诊断中的应用主要为优化, 并且在消除噪音, 提高故障诊断的效率和识别率方面有广阔的前景[28]。一项研究基于中医证候的疗效评价信息, 通过 GA 进行数据挖掘核心处方, 并进行了动物实验的验证, 证实 GA 可用于中医药治疗肺癌 CEF 的数据挖掘研究[29]。

综上所述 DM 最核心的功能其实就是分类及预测, 通过上述各种算法对目标数据进行分类、预测而实现对海量信息的有效分析, 不管是在哪个领域都起着至关重要的作用。信息爆发的大数据时代, 以惊人的速度增长的海量数据已经超出了计算早期使用的简单数据库和数据处理架构的范围, 以快速更新的高容量、高维度数据为特征的数据集, 代表了数字时代已经出现的一种现象或特征。在计算机科学领域, 大数据是指无法利用传统 IT 和硬件工具在可容忍的时间内感知、获取、管理、处理或服务的数据集[30]。然而 DM 具有处理海量、非线性复杂信息以及发现庞大的数据背后隐藏的巨大潜在价值的优势。因此有望在这大数据时代 DM 能为科学研究及企业发展提供决策依据。

## 2. 数据挖掘应用在医药领域的应用

### 2.1. 数据挖掘技术在现代医学的应用

医疗大数据的复杂性源于疾病及其合并症的多样性; 治疗和疗效的异质性; 医疗数据获取、处理、解读以及研究设计和分析方法的复杂性[31]。由于 DM 具有处理海量、非线性复杂信息以及发现庞大的数据背后隐藏的巨大潜在价值的优势, 应用于医疗数据挖掘有望提高医疗质量、优化医疗流程和管理策略。当前广泛应用于建立预测模型和临床决策支持、疾病或安全监控、公共卫生和研究等方面。如常见的医学诊断、疾病风险及影响因素评估、药物临床试验、药物研发、预防、护理及基础实验研究等。临床决策支持一般指的是疾病预测和疾病智能诊断两大方面, 并在临床研究中取得一定得研究成果[32]。

首先是疾病预测与分类, 其主要原理是通过已知的临床数据利用数据挖掘各类算法建立预测模型, 再通过该模型对新的临床数据进行预测分析。包括疾病检测, 疾病分类, 疾病风险及影响因素评估等。疾病分类, 如支持向量机建模在常见疾病预测中的应用, 即利用支持向量机(SVM)技术来分类患有常见疾病的人[33]; Pranab Haldar、Ian D. Pavord 等探讨一种多变量数学技术 k-means 聚类分析在鉴别哮喘不同表型组的应用[34]; 疾病检测, 如 Aisha Mashraqi、Hanan Halawani 等研究 Covid-19 ICU 患者肝功能不良反应预测模型, 研究结果表明 SVM 和 DT 分类器的准确率、精确率和召回率均优于其他分类器, 因此

认为 SVM 和 DT 可能是检测肝损伤风险的最佳算法[35];

其次是在医学诊断中的应用, 说道医学诊断不得不提疾病智能诊断, 尤其是医学影像学在疾病辅助诊断中的应用。临床上大量的医学图像的人工判读是繁琐又耗时的, 并且容易造成偏见和错误, 因此计算机辅助诊断系统的应用更受关注[36]。利用卷积神经网络快速高效地对医学影像数据进行分析和处理可以实现医学影像数据快速分类、定位等操作, 提高医学诊疗的效率[37]。如深度卷积神经网络在心电信号自动检测心肌梗死中的应用[38]; 基于深度特征和贝叶斯优化的新型 COVID-19 感染检测医疗诊断模型, 研究提出一种能够自动检测 COVID-19 阳性病例的智能计算机辅助模型, 以支持日常临床应用。该模型基于卷积神经网络(CNN)架构[39]; 癌症早期检测在癌症诊断中起着重要的作用, 可以提高生存率。医学影像学应用于早期癌症检测、监测和治疗及随访, 如基于深度卷积神经网络 CNN 方法的 MR 脑图像中有有效的肿瘤检测[40]。

以数据挖掘技术在心血管领域应用为例, 众所周知, 心血管疾病对人们健康造成很大的威胁, 预测心血管疾病风险、影响因素及早发现、早诊断意味着拯救生命, 目前大数据时代通过数据挖掘技术可以更好的识别心血管疾病风险, 以及早检测疾病帮助诊断。

如基于决策树数据挖掘的冠心病事件危险因素评估研究, 使用 C4.5 决策树算法评估心血管事件发生危险因素, 以减少心血管事件发生[41]; 如应用 TreeNet 算法建立原发性高血压(EH)早期预测模型, 为早期监测 EH 提供预测方法[42]; 使用分类回归树(CART)给急性心力衰竭 1 年死亡率的危险分层研究结果表明在 360 天内死亡的低、中和高风险的 AHF 患者可以很容易地通过患者的人口统计学和实验室数据来识别。应用这种简单的危险分层算法可能有助于改善这些患者的管理[43]; 一项深度卷积神经网络在心电信号自动检测心肌梗死 MI 中的应用研究提出了一种使用 11 层深度 CNN 自动诊断 MI 的新方法, 经研究表明提出的系统的整体性能足够好, 可以在临床环境中引入辅助诊断[44]。

其他如公共卫生及预防医学、药物临床试验、新药开发以及科研等方面应用。如王维等 31 利用 1995~2003 年数据预测 2004~2005 年出生缺陷患病率, 并表示与传统统计方法相比, BP 人工神经网络不仅具有更好的预测精度, 而且对相关数据的类型和分布没有限制, 为流行病学预测提供了强有力的方法。

## 2.2. 数据挖掘技术在中医领域应用

### 2.2.1. DM 在中医诊疗中的应用

近年来数据挖掘在祖国医学的应用越开越广泛, 在中医诊断、中医治疗相关研究中起着重要的作用。在中医诊断方面有助于中医症候规范化以及辨证分型准确性; 而通过中医处方挖掘、中医药量效关系研究、中医非药物治疗研究等为中医临床治疗疾病提供依据。

中医诊断方面, 目前用于中医证候研究的数据挖掘方法主要有: 关联规则、集对分析、粗糙集理论、聚类分析、人工神经网络、决策树、支持向量机、贝叶斯网络等。如利用现有的中医胃炎病历作为样本数据, 采用关联规则的方法建立实验模型, 并给出实验分析的结果。在此基础上采用决策树方法, 构建一棵判断是否为辨证“中虚气滞”的决策树; 黄嘉韵等 DM 决策树算法探讨鼻鼾辨证规律, 结果表明决策树算法探讨辨证规律有广泛实用价值, 可用于鼻鼾中医临床辨证辅助诊断[45]; 通过算法的描述得到了较简化的区分感冒和咳嗽病的决策树模型, 使用该模型对其它历史病例进行测试, 具有较高的准确率, 基本达到诊断要求[46]; 金楠楠[47]等中医症候要素研究; 还有人工神经网络在中医诊断中的应用研究涉及到的层面较为广泛, 主要包括: 同类病中不同病的诊断、同一病中不同证的诊断、相似病与证的诊断、相似证的诊断、病情进展阶段的诊断、基于舌诊的八纲辨证诊断等; 爱明等利用 FCM 聚类算法对中医舌象自动分类, 结果表明这种方法能够有效地对舌质、舌苔的颜色进行分类[48]; 中医治疗方面, 中医处方挖掘指的是指古代金典处方挖掘、名医经验方挖掘从而发现有价值的治疗思路及配伍规律, 为临床治疗

疾病提供依据。如中医药疗效研究, 如袁南等[49]将聚类 and 模糊关联规则应用于中医药对量效分析, 符合中医自身的用药特点, 并且得到了较好的实验结果。

### 2.2.2. 中医处方数据挖掘

当前大数据时代, 数据挖掘技术为中医药数据信息处理提供了良好的技术服务平台, 主要应用于挖掘古代医家用药思路或现代医家治疗某疾病、症候的用药规律等[50] [51]。常见的分析方法包括频数统计、关联规则分析、聚类分析等[52]。

频数统计——在中医药处方数据挖掘中主要用于药物频次分析以及药物四气、五味、归经频次分析等。具体方法: 使用 WPS Office Excel 2019 软件。

关联规则——关联规则是事物或者事件之间相关性的一种说法, 事物之间关联性的大小可以用关联规则来诠释。主要用于中医药处方组方规律挖掘, 如基于关联规则的用药规律分析[53], 具体方法: 应用 WPS Office Excel 2019 软件建立中药关联分析矩阵, 将矩阵导入 SPSS Modeler 18.0 中, 以 Apriori 算法构建数据流, 进行关联规则分析及关联规则建模。

聚类分析——是以对象的基本特征为基础, 按照一定的标准对研究对象进行分类, 它可以使得组内的数据对象具有高度的相似性, 组间的差异性最大。在中医药数据挖掘中主要用于发现两味药物之间的配伍关系、多味药物之间的配伍关系、方剂之间的共性关系等[54], 还可以找到多种类方。具体方法: 将处方中高频中药整理后导入 IBM SPSS Statistics 19 进行系统聚类分析。

数据挖掘广泛应用于中医药处方研究中的同时也面临着相应的挑战, 尤其是在数据获取及数据处理中存在一定的局限, 在文献检索、中医药名称、症状、证候术语规范化、结论验证等方面有待提高。数据挖掘技术为中医药现代化研究提供了更好的机遇, 未来还需不断优化研究设计及算法应用, 从而提高研究质量[55]。

## 参考文献

- [1] 吉根林, 孙志挥. 数据挖掘技术[J]. 中国图象图形学报, 2001(8): 2-8.
- [2] 肖铮. 常用的三种分类算法及其比较分析[J]. 重庆科技学院学报(自然科学版), 2020, 22(5): 101-106.
- [3] 李燕. 海量数据与数据挖掘技术[J]. 医学信息学杂志, 2008, 29(12): 23-25.
- [4] Wu, X., Kumar, V., Ross Quinlan, J., et al. (2008) Top 10 Algorithms in Data Mining. *Knowledge and Information Systems*, 14, 1-37. <https://doi.org/10.1007/s10115-007-0114-2>
- [5] 范铁兵, 宁秋萍, 杨志旭, 顾东黎, 朱晓博, 燕嫄. 基于数据挖掘技术的中医学隐性知识显性化分析[J]. 中国实验方剂学杂志, 2017, 23(10): 221-226.
- [6] 唐敏. 关联规则挖掘算法在超市销售分析中的应用[J]. 计算机科学, 2006, 33(2): 149-150.
- [7] 李洪燕, 万新. 基于关联规则分析的产品销售推荐的应用[J]. 四川理工学院学报: 自然科学版, 2013, 26(2): 36-38.
- [8] 李昀洲, 陈璐. 考虑需求关联性的多供应商电商库存分配研究[J]. 上海理工大学学报, 2023, 45(2): 180-188.
- [9] 赵韩, 刘生强. 基于多维关联规则驱动的新产品配置研究[J]. 合肥工业大学学报: 自然科学版, 2016, 39(12): 1585-1591.
- [10] 许海柱, 张婷, 孙建立. 关联规则数据挖掘方法在中医药研究中应用进展[J]. 辽宁中医药大学学报, 2013, 15(12): 131-134. <https://doi.org/10.13194/j.issn.1673-842x.2013.12.027>
- [11] 王磊, 刘东苏. 关联规则挖掘在图书馆信息服务中的应用[J]. 情报杂志, 2008, 27(2): 154-155+158.
- [12] 郭鹏, 蔡骋. 基于聚类和关联算法的学生成绩挖掘与分析[J]. 计算机工程与应用, 2019, 55(17): 169-179.
- [13] Wu, W.-T., Li, Y.-J., Feng, A.-Z., et al. (2021) Data Mining in Clinical Big Data: The Frequently Used Databases, Steps, and Methodological Models. *Military Medical Research*, 8, Article No. 44. <https://doi.org/10.1186/s40779-021-00338-z>

- [14] 倪鹏飞, 刘高军, 宋璇涛. 中国城市竞争力聚类分析[J]. 中国工业经济, 2003(7): 34-39.  
<https://doi.org/10.19581/j.cnki.ciejournal.2003.07.005>
- [15] 常晶茹, 姚萱航, 张雪薇, 周仪, 刘翠晶, 刘霞. 黄芪属植物 DNA 条形码与聚类分析的研究[J]. 中草药, 2022, 53(22): 7201-7206.
- [16] 唐华松, 姚耀文. 数据挖掘中决策树算法的探讨[J]. 计算机应用研究, 2001, 18(8): 18-19+22.
- [17] 张海燕, 刘岩, 马丽萌, 苑津莎, 巨汉基, 魏彤珈. 决策树算法的比较与应用研究[J]. 华北电力技术, 2017(6): 42-47. <https://doi.org/10.16308/j.cnki.issn1003-9171.2017.06.008>
- [18] 林文怡, 宛小燕, 刘元元. 常见新近决策树算法及其在卫生领域中的应用[J]. 现代预防医学, 2019, 46(23): 4233-4237+4242.
- [19] Song, Y.-Y., Lu, Y. 用于分类与预测的决策树分析(英文) [J]. 上海精神医学, 2015, 27(2): 130-135.
- [20] 何珊, 刘振东, 马小林. 信用评分模型比较综述——基于传统方法与数据挖掘的对比[J]. 征信, 2019, 37(2): 57-61.
- [21] Angermueller, C., Pärnamaa, T., Parts, L., et al. (2016) Deep Learning for Computational Biology. *Molecular Systems Biology*, **12**, 878. <https://doi.org/10.15252/msb.20156651>
- [22] 李炳臻, 刘克, 顾佼佼, 姜文志. 卷积神经网络研究综述[J]. 计算机时代, 2021(4): 8-12+17.
- [23] 张驰, 郭媛, 黎明. 人工神经网络模型发展及应用综述[J]. 计算机工程与应用, 2021, 57(11): 57-69.
- [24] 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述[J]. 计算机学报, 2017, 40(6): 1229-1251.
- [25] 孙肇阳, 吕桂娇, 郭义, 陈波, 陈勇, 马铭泽. 中医诊疗复杂巨系统的人工神经网络建模原理[J]. 中华中医药杂志, 2022, 37(10): 5841-5844.
- [26] Akkus, Z., Galimzianova, A., Hoogi, A., et al. (2017) Deep Learning for Brain MRI Segmentation: State of the Art and Future Directions. *Journal of Digital Imaging*, **30**, 449-459. <https://doi.org/10.1007/s10278-017-9983-4>
- [27] 张文修, 梁怡. 遗传算法的数学基础[J]. 西安交通大学学报, 2000, 34(12): 5.
- [28] 骆志高, 田海泉, 仇学青. 遗传算法在故障诊断中的应用研究综述[J]. 煤矿机械, 2006(1): 169-172.  
<https://doi.org/10.13436/j.mkjx.2006.01.076>
- [29] 李嘉旗, 杨铭, 焦丽静, 黄默菲, 许玲. 基于遗传算法的中医药治疗肺癌核心有效处方的发现及实验研究[J]. 中华中医药杂志, 2018, 33(9): 4143-4146.
- [30] Wu, W.-T., Li, Y.-J., Feng, A.-Z., et al. (2021) Data Mining in Clinical Big Data: The Frequently Used Databases, Steps, and Methodological Models. *Military Medical Research*, **8**, Article No. 44.  
<https://doi.org/10.1186/s40779-021-00338-z>
- [31] Lee, C.H. and Yoon, H.J. (2017) Medical Big Data: Promise and Challenges. *Kidney Research and Clinical Practice*, **36**, 3-11. <https://doi.org/10.23876/j.krep.2017.36.1.3>
- [32] 梁书彤, 郭茂祖, 赵玲玲. 基于机器学习的医疗决策支持系统综述[J]. 计算机工程与应用, 2019, 55(19): 1-11.
- [33] Yu, W., Liu, T., Valdez, R., et al. (2010) Application of Support Vector Machine Modeling for Prediction of Common Diseases: The Case of Diabetes and Pre-Diabetes. *BMC Medical Informatics and Decision Making*, **10**, Article No. 16.  
<https://doi.org/10.1186/1472-6947-10-16>
- [34] Haldar, P., Pavord, I.D., Shaw, D.E., et al. (2008) Cluster Analysis and Clinical Asthma Phenotypes. *American Journal of Respiratory and Critical Care Medicine*, **178**, 218-224. <https://doi.org/10.1164/rccm.200711-1754OC>
- [35] Mashraqi, A., Halawani, H., Alelyani, T., et al. (2022) Prediction Model of Adverse Effects on Liver Functions of COVID-19 ICU Patients. *Journal of Healthcare Engineering*, **2022**, Article ID: 4584965.  
<https://doi.org/10.1155/2022/4584965>
- [36] Hu, Z., Tang, J., Wang, Z., et al. (2018) Deep Learning for Image-Based Cancer Detection and Diagnosis—A Survey. *Pattern Recognition*, **83**, 134-149. <https://doi.org/10.1016/j.patcog.2018.05.014>
- [37] 王霄, 朱恩照, 艾自胜. 卷积神经网络的原理及其在医学影像诊断中的应用[J]. 中国医学物理学杂志, 2022, 39(12): 1485-1489.
- [38] Acharya, U.R., Fujita, H., Oh, S.L., et al. (2017) Application of Deep Convolutional Neural Network for Automated Detection of Myocardial Infarction Using ECG Signals. *Information Sciences*, **415-416**, 190-198.  
<https://doi.org/10.1016/j.ins.2017.06.027>
- [39] Nour, M., Cömert, Z. and Polat, K. (2020) A Novel Medical Diagnosis Model for COVID-19 Infection Detection Based on Deep Features and Bayesian Optimization. *Applied Soft Computing*, **97**, Article 106580.  
<https://doi.org/10.1016/j.asoc.2020.106580>

- [40] Arunachalam, S. and Sethumathavan, G. (2022) An Effective Tumor Detection in MR Brain Images Based on Deep CNN Approach: *i-YOLOV5*. *Applied Artificial Intelligence*, **36**, Article 2151180. <https://doi.org/10.1080/08839514.2022.2151180>
- [41] Karaolis, M., Moutiris, J.A., Hadjipanayi, D., *et al.* (2010) Assessment of the Risk Factors of Coronary Heart Events Based on Data Mining with Decision Trees. *IEEE Transactions on Information Technology in Biomedicine*, **14**, 559-566. <https://doi.org/10.1109/TITB.2009.2038906>
- [42] 郁小红, 钱棧梅, 周晨洁, 马越, 唐艳超, 邹玲莉. 应用 TreeNet 算法建立原发性高血压早期预测模型[J]. 预防医学, 2022, 34(9): 923-927. <https://doi.org/10.19485/j.cnki.issn2096-5087.2022.09.012>
- [43] Arenja, N., Breidthardt, T., Socrates, T., *et al.* (2011) Risk Stratification for 1-Year Mortality in Acute Heart Failure: Classification and Regression Tree Analysis. *Swiss Medical Weekly*, **141**, w13259. <https://doi.org/10.4414/smw.2011.13259>
- [44] Acharya, U.R., Fujita, H., Oh, S.L., *et al.* (2017) Application of Deep Convolutional Neural Network for Automated Detection of Myocardial Infarction Using ECG Signals. *Information Sciences*, **415-416**, 190-198. <https://doi.org/10.1016/j.ins.2017.06.027>
- [45] 黄嘉韵, 郭宏, 邝艳萍. 基于决策树算法的鼻鼾辨证规律初步研究[J]. 中华中医药杂志, 2016, 31(11): 4770-4773.
- [46] 王星, 刘晓燕. 医疗大数据环境下的疾病预测模型研究[J]. 制造业自动化, 2022, 44(7): 24-27.
- [47] 金楠楠, 马堃. 基于聚类及因子分析的功能失调性子宫出血中医证候要素研究[J]. 中国中药杂志, 2008, 33(13): 1622-1625.
- [48] 王爱民, 沈兰荪, 赵忠旭. 监督 FCM 聚类算法及其在中医舌像自动分类中的应用[J]. 模式识别与人工智能, 1999, 12(4): 480-485.
- [49] 袁楠, 金晖, 田玲, 蒋永光, 于中华. 基于聚类和模糊关联规则的中医药对量效分析[J]. 计算机应用研究, 2009, 26(1): 59-61.
- [50] 段展辉, 刘桂荣. 基于数据挖掘探讨温补学派治疗郁证处方及用药规律[J]. 中药药理与临床, 2022, 38(4): 174-179.
- [51] 蔡铃英, 傅丽芳. 以数据挖掘为基础的中医治疗冠心病处方用药规律研究[J]. 中国全科医学, 2021, 24(S1): 160-162.
- [52] 冯梅, 王颖, 柏冬, 王俊宏. 基于数据挖掘的中医处方分析方法研究进展[J]. 世界中医药, 2022, 17(23): 3411-3416.
- [53] 赵雪茹, 马利, 李淑琪, 李慧, 苏晓川, 赵移畛. 基于关联规则的骨质疏松性骨折中药熏洗处方用药规律研究[J]. 中国骨质疏松杂志, 2022, 28(1): 84-88.
- [54] 芦锰, 周雨慧, 李晓宁, 孙昇丹, 郭金栋, 武博文, 武密山. 基于数据挖掘中医药治疗阿尔茨海默病用药规律研究[J]. 中国中药杂志, 2021, 46(6): 1558-1563.
- [55] 但文超, 赵国桢, 何庆勇, 张辉, 李博, 张广中. 中医药处方数据挖掘的常见问题辨析与展望[J]. 中国中药杂志, 2023, 48(17): 4812-4818.