

项目编号:

# 暨南大学大学生创新创业训练计划 项目申报书

项目名称	<u>遵循中医思维的退化性骨关节病智能辅助诊断</u>
项目类型	<u>创新训练项目</u>
项目类别	<u>重点支持领域项目</u>
项目负责人	<u>阮炜霖</u>
学 号	<u>2020101603</u>
院（部）专业	<u>信息科学技术网络工程</u>
入学年份	<u>2020</u>
联系电话	<u>13672546637</u>
E-mail	<u>1264532114@qq.com</u>
指导教师	<u>张佳</u>
申报时间	<u>2023年3月7日</u>

暨南大学教务处制

二〇二二年十二月

# 填 写 须 知

## 一、项目类型分类说明：

1. 创新训练项目是本科生个人或团队，在导师指导下，自主完成创新性实验方法的设计、研究条件准备和项目实施、研究报告撰写、成果（学术）交流等工作。

2. 创业训练项目是本科生团队，在导师指导下，团队中每个学生在项目实施过程中承担一个或多个具体的角色，通过编制商业计划书、开展可行性研究、模拟企业运行、参加企业实践、撰写创业报告等工作。

3. 创业实践项目是学生团队，在学校导师和企业导师共同指导下，采用前期创新训练项目（或创新性实验）的成果，提出一项具有市场前景的创新性产品或者服务，以此为基础开展创业实践活动。

二、项目类别分为一般项目和重点支持领域项目两类。其中，重点支持领域项目旨在鼓励引导大学生根据国家经济社会发展和重大战略需求，结合创新创业教育发展趋势，在重点领域和关键环节取得突出创新创业成果。

三、申报书请按顺序逐项填写，填写内容必须实事求是，表达明确严谨。空缺项要填“无”。

四、填写时可以改变字体大小等，但要确保表格的样式没有被改变；填写完后用 A4 纸张双面打印，不得随意涂改。

项目名称	遵循中医思维的退化性骨关节病智能辅助诊断					
项目类型	创新训练项目		项目类别	重点支持领域项目	项目研究周期	12
负责人情况	姓名	阮炜霖	性别	男	平均绩点	3.60
	学院	信息科学技术学院	专业	网络工程	入学年份	2020
	联系电话	13672546637	QQ号	1264532114	Email	rwlino@gmail.com
项目组成员 (不含负责人)	姓名	性别	学号	院(部)	专业	入学年份
	刘桢江	男	2019051393	信息科学技术学院	智能科学与技术	2019
	周旭林	男	2019051398	信息科学技术学院	智能科学与技术	2019
	陈彦亨	男	2020101602	信息科学技术学院	网络工程	2020
	黄崇铎	男	2020101604	信息科学技术学院	网络工程	2020
指导老师情况 (第一导师)	姓名	张佳	学科专业	智能科学与技术	所在单位	暨南大学信息科学技术学院计算机科学系
	职称/职务	中级	学历学位	博士	联系电话	13602213851
校外导师情况 (创业实践项目必填)	姓名		所在单位			
	职称/职务		学历学位		联系电话	

## 一、项目简介

退化性骨关节病又称骨关节炎，是一种退化性病变，系由于增龄、肥胖、劳损、创伤、关节先天性异常、关节畸形等诸多因素引起的关节软骨退化损伤、关节边缘和软骨下骨反应性增生。临床表现为缓慢发展的关节疼痛、压痛、僵硬、关节肿胀、活动受限和关节畸形等。

作为中医优势病种，骨关节病中医数据在统计特征选择中起到了关键作用。相比西医的 X 线、CT 等光学诊断数据，中医的脉诊、舌色等“望闻问切”的数据能较好的反映部分骨关节病症的隐含细节，提供更可靠、更快捷且低成本的诊断结果，具有重要的分析价值和挖掘潜力。

基于近年来中医在治疗骨关节病症方面的研究成果，该项目首先梳理了近年来有关退化性骨关节病的研究工作，然后从遵循中医整体思维的角度，提出了一种客观化、规范化、智能化的退化性骨关节病辅助诊断手段，即**结合人工智能技术，结合特征选择、深度学习的方法，从当地医院收集骨关节病患者的性状特征数据来训练神经网络，构建精准高效的骨关节病预测模型，并最终投入应用在小程序或网站平台等用户界面上。**该项目能为骨关节病人的症状评估提供参考，方便群众对骨关节炎的早期预防和诊断。

## 二、立项依据（研究意义、国内外研究现状及发展动态分析等）

### 1. 研究意义

遵循中医思维研究人工智能辅助诊断退化性骨关节病的方法具有重大的意义。

其研究意义概括表现在：（一）行业发展方面，推进中医药创新发展并支撑健康中国建设是国家重大战略布局，该项目积极响应国家政策和社会需求，瞄准中医药现代化关键科学问题及智能化发展目标，利用新一代信息技术为中医药事业赋能，引领中医药现代化发展，助力中医药产业和健康服务体系转型升级。（二）社会价值层面，优化预测骨关节炎方法，有利于病患及早发现并治疗，减轻百姓家庭和社会的经济负担。西医诊断条件苛刻，成本较高，西方国家患者人数较我国少，而中医诊断便捷快速且经济成本低，结合我国患病人数基数大和消费水平的基本情况，相比起西医，中医更适合作为骨关节炎的优先诊断方法。（三）创新科研层面，探讨传统的特征选择方法与深度学习方法在特征选择上的性能对比，深度学习学习方法学习出来的无物理含义无解释性的特征，性能是否会比基于统计学方法有物理含义的机器学习方法更好。

#### 1.1 中医健康行业发展的重大需求

中医的整体思维是中医学的灵魂，中医的思维特色和优势集中体现在：一是把人看作一个有机整体，强调形神统一以及人与自然的统一；二是把生命的全过程当作一个整体，涵盖出生、生长、发育、疾病、衰老、死亡；三是整体观念指导下的辨证论治。伴随着健康观念的转变，现代医学也已从单纯的“生物-医学模式”转向“环境—社会—心理—生物医学模式”，已从“预防为主”转变为“促进健康为主”，人们对健康的需求也在不断增加。早在几年前，健康产业就已被认为是信息产业之后下一个超万亿美元的行业。面对这一重大需求，中医药的优势日益凸显出来，正如《中医药健康服务发展规划（2015-2020年）》指出：中医药（含民族医药）强调整体把握健康状态，注重个体化，突出治未病，临床疗效确切，治疗方式灵活，养生保健作用突出，是我国独具特色的健康服务资源，必将在健康产业的发展中发挥重要的基础和引领作用。但是，目前能够真正满足中医健康服务需求的中医诊疗手段还十分匮乏，也无法满足中医健康大数据对数据源的可靠性的需求。因此，要加强智能化中医诊疗系统的研发，重点研发中医健康识别系统、智能中医体检系统等中医健康辨识、干预设备，以实现中医诊疗的规范化、可测量、可评价。

本项目计划立足中医思维，以中医健康状态辨识理论为核心，充分应用现代类人智能技术，研发中医诊疗系统，可以与移动互联网融合，充分发挥其自动化、智能化的优势，在大数据支撑下，融合中医诊疗的共性技术，构建起中医健康服务的通用技术平台，满足中医健康产业发展的需求。

#### 1.2 提高中医服务能力的社会需要

我国患退化性骨关节病人数多，致残率高，对社会和家庭均造成巨大经济负担。随着我国人口老龄化程度加深，发病率还有逐渐上升的趋势。目前国内退化性骨关节病患者依赖西医诊断治疗手段，治疗条件苛刻，价格昂贵，相比之下中医四诊便捷快速且更节省费用，因此运用中医方法对退化性骨关节病的早期诊断更有优势。

可与此同时，我国发展中医药健康服务面临重大问题：中医从业人员数较少，部分基层中医的服务能力参差不齐，服务品质不高，四诊技能不规范，辩证思维不熟练，临床诊断准确性不高，这

种现象不但不能满足人民群众日常诊断的健康需要，也限制了中医药对各个病种整体优势的发挥。因此，发展中医人工智能诊疗系统势在必行。我国《中医药发展战略规划纲要（2016-2030年）》指出：到2030年，大力发展中医远程医疗、移动医疗、智慧医疗等新型医疗服务模式。综合运用现代科技手段，开发一批基于中医理论的诊疗仪器与设备，使中医药健康服务能力显著增强，使中医在治未病中的主导作用、在重大疾病治疗中的协同作用得到充分发挥。

### 1.3 人工智能结合中医的科学需要

现代科学技术的发展给中医技术带来了革命性的进步，但依然存在以下几个问题：（1）研究比较分散，缺乏统一的理论支撑，与健康疾病相关的“天、时、地”等因素关注不够，难以体现中医的整体思维特征；与人工智能结合不够，多学科交叉的优势难以发挥。（2）由于过分强调指标的特异性，四诊信息采集仍立足于单一信息，如脉诊仪、舌诊仪、经络仪等，虽然传感器技术已经十分发达，但仍存在结果不一致甚至相互矛盾的现象，无法体现四诊合参。（3）由于对中医健康认知核心缺乏深刻的认识，导致中医思维的偏差，加上辨证方法不一致，结果难以形成共识。在信息分类、模型识别方面缺乏统一的标准，成为制约中医药诊疗设备研发与健康大数据接轨的瓶颈。

通过人工智能技术构建智能辅助诊断模型是一项极具挑战而意义重大的工作。其原因在于：其一，中医健康数据的不确定性。诊断有赖于未来高质量、大规模的临床研究和观察结果，而且中医健康数据不可避免地存在大量缺失，要求开发的学习算法适用于极端稀疏数据的处理；其二，中医健康数据多模态多标记的表示形式对学习算法的设计提出了更高的要求。这意味着采集的病人信息具有多源异构的特性，故多模态数据的融合是数据分析有效性的重要环节。最后，由于需要辨识病人的病位证素、病性证素与证型，辨识过程可看作一个典型的多标记学习问题，而基于标记之间的关联信息来构建模型，是获取兼具准确性与鲁棒性的辨识结果的关键。因此，以健康状态为核心，应用人工智能实现四诊合参和构建中医智能诊疗模型，紧扣数据不确定性、多模态、多标记等科学问题对大数据分析的影响，运用特征工程、深度学习等方法解决此类问题是实现退化性骨关节病智能辅助诊断的必由之路。

总而言之，将人工智能技术应用到中医辅助智能诊疗，为中医的理论和临床研究提供科学基础和方法学平台，为中医数据库建立和健康产业发展提供技术平台，以及实现中医诊疗的客观化、规范化和智能化，对于推进中医现代化发展具有重要意义。

## 2. 国内外研究现状

目前国内外普遍根据患者病史、症状、体征和影像学表现对退化性骨关节病做出临床诊断（参考美国风湿病学会标准、欧洲抗风湿联盟的诊断建议以及国际骨关节炎研究学会的MRI诊断研究）。在西医方面，对KOA(knee osteoarthritis, 膝骨关节炎)分期有多种方法，主要有Kellgren-Lawrence。而现阶段中医也在膝骨关节炎研究方面得到进展。中医将KOA分为气滞血瘀、风寒湿痹、肝肾亏虚和湿热蕴结四型，中医药治疗历史悠久，中药、手法、针灸和传统功能锻炼等是常用的中医非手术特色疗法。近年来研究已证实针刺治疗KOA具有良好效果，经典名方独活寄生汤以及中成药也能够有效改善KOA患者膝关节疼痛、功能障碍等症状体征，且安全性相对较好。

中医学具有非常悠久的历史，几千年来对保障中国百姓的健康和生活方式发挥了不可替代的作用。进入二十一世纪以来，人类医学的目的正在从“疾病医学”向“健康医学”转变，中医学整体观念和辩证论治的优势得到了世界医学和各级政府的高度重视。尤其是在信息技术高速发展的今天，大数据、云计算和人工智能给医学带来了颠覆性的革命，应用人工智能技术为中医诊疗服务，

成为业内外高度关注的问题。但是，由于对中医的思维特点及其复杂性、模糊性缺乏有效的把握，成为制约本领域发展的瓶颈，也限制了中医药的时代表达和优势的发挥。

### 3. 发展动态分析

传统中医药学具有悠久的历史，其整体观和辩证观的独特优势得到了世界医学领域的高度重视。尤其是在大数据和人工智能正在得到广泛应用的今天，如何应用人工智能技术去挖掘数据潜能，并结合中医传统文化的深厚理念去进行诊疗，是一个前沿且关键的交叉学科问题。

中医诊疗系统由四诊信息采集、健康状态辨识（诊断）、干预（治疗）及效果评价等几个部分构成。上世纪 50 年代以来，部分中医药工作者和医药工程专家开始了中医诊疗仪器的研究与探索。特别是近 10 年来，随着人工智能、大数据、云计算等技术的飞速发展，大量的新技术、新方法开始被广泛应用于中医诊疗系统的研究，推动了本领域的发展。

随着科学技术的不断发展，很多新技术、新材料也不断被运用于中医诊断设备的研制中，形成多学科交叉的态势，使得传统的中医四诊也有了新的发展。信息技术的迅猛发展，特别是信息获取技术与物理信息系统等技术的不断突破，引发了网上的数据规模正以前所未有的速度增长。人工智能与中医诊断信息处理相结合，对于提高中医诊断辨证的客观性、规范性和准确性，降低误诊率有着积极的现实意义。

机器学习领域的经典和前沿算法也被大量用于中医领域中。通过将肤色识别、状态评估等特定问题抽象成监督学习中的经典分类和回归等问题，可以采用支持向量机、随机森林、神经网络等经典学习算法和神经网络为智能诊断提供帮助。此外，为了处理中医领域由丰富的语义信息所呈现的多样化的标签特征信息，多标记学习等研究方法也被引入，以解决为一个病例同时诊断多种类别的症状的复杂情况。总的来说，机器学习算法在中医领域的应用是跨学科知识融合和应用的典范。

在大数据时代，以健康状态为核心，应用人工智能技术，能够实现表征参数的采集、存储、分类、融合、状态辨别、干预方案匹配和效果评价的客观化、标准化和智能化，为中医诊疗系统的研发、中医数据库的建立、治未病健康工程和中医养生保健服务产业的发展提供方法学支持。

总的来说，人工智能技术正处于蓬勃发展的阶段，中医专家系统和中医四诊仪器研究近年来也积累了一定的成果，运用人工智能方法结合中医思维辅助诊断，将成为传统中医医学发展革新中的一个契机，也正是这个项目致力实现的目标。

### 4. 主要参考文献

- [1] 国务院关于印发中医药发展战略规划纲要（2016-2030 年）的通知[J]. 中华人民共和国国务院公报, 2016, (8): 21-29.
- [2] 国务院办公厅关于印发中医药健康服务发展规划（2015-2020 年）的通知[J]. 中华人民共和国国务院公报, 2015, (14): 32-39.
- [3] 甘富文. (2020). 多特征融合的肝细胞癌分化等级术前预测方法研究 (Master's thesis, 郑州大学). 11-13
- [4] 胡艳梅, 杨波, & 多滨. (2021). 基于网络结构的正则化逻辑回归. 计算机科学, 48(7), 281-291.
- [5] 何馨, 邹绮蕾, 卞禾, & 何诗思. (2014). 三种分类预测模型在医学中的应用研究. 怀化学院学报, 33(11), 29-32.
- [6] 计智伟, 胡珉, & 尹建新. (2011). 特征选择算法综述. 电子设计工程, 19(9), 46-51.
- [7] 徐蕾, 贺佳, 孟虹, 贺宪民, & 范思昌. (2004). 决策树技术及其在医学中的应用. 数理医药学杂

志, 17(2), 161-164.

- [8] 许家佗. 基于四诊信息决策支持的中医健康评价体系研究述评与展望[J]. 中国中西医结合杂志, 2012, 32(3): 307-310.
- [9] 邸丹, 周敏, 秦鹏飞, 等. 中医舌诊、面诊客观化研究进展[J]. 上海中医药杂志, 2012, 46(4): 89-92.
- [10] 余江维, 马利庄, 杨华元. 中医智能化诊断的研究现状与展望[J]. 辽宁中医杂志, 2010, 37(1): 50-53.
- [11] 褚娜. 基于混合智能的中医辨证系统研究[D]. 上海交通大学: 2012.
- [12] 石英杰. 基于病机模型的胸痹病中医智能辅助诊断方法研究[D]. 中国中医科学院, 2021.
- [13] 任丽娜. 基于病案数据挖掘的中风病痰热腑实特点, 证效转归和证候预测研究[D]. 北京中医药大学, 2017.
- [14] 陈德鑫. 基于深度学习的在线医疗信息抽取研究[D]. 武汉大学, 2017.
- [15] 胡杨. 面向智能中医辅助诊疗的多注意力和知识辅助神经网络设计研究[D]. 华南理工大学, 2020.
- [16] 刘旭龙, 洪文学, 张涛, 等. 基于形式概念分析的中医辨证可视化方法[J]. 燕山大学学报, 2010, 34(2): 162-164
- [17] C.L. Philip Chen, Chunyang Zhang. Data-intensive applications, challenges, techniques and technologies: A survey on big data [J]. Information Sciences, 2014, 275(11): 314-347.
- [18] Peiqin Gu, Huajun Chen n. Modern bioinformatics meets traditional Chinese medicine [J]. Briefings in Bioinformatics, 2014, 15(6): 984-1003.
- [19] LI, B. K., CONG, Y., TIAN, Z. Y., & XUE, Y. (2014). Predicting and virtually screening the selective inhibitors of MMP-13 over MMP-1 by molecular descriptors and machine learning methods. Acta Physico-Chimica Sinica, 30(1), 171-182.
- [20] Wenhsiang Wu, Jiayou Liu, Henhong Chang. Chang. Latent class model based diagnostic system utilizing traditional Chinese medicine for patients with systemic lupus erythematosus [J]. Expert Systems with Applications, 2011, 38(1): 281-287.
- [21] M. Wang, W. J. Fu, X. N. He, S. J. Hao, X. D. Wu. A survey on large-scale machine learning[J]. IEEE Transactions on Knowledge and Data Engineering, doi: 10.1109/TKDE.2020.3015777.
- [22] R. Agrawal, A. Gupta, Y. Prabhu, M. Varma. Multi-label learning with millions of labels: recommending advertiser bid phrases for web pages[C]. In Proceedings of the 22nd International World Wide Web Conference, Rio de Janeiro, Brazil, 2013, pp. 13-24.
- [23] 李灿东. 中医状态学[M]. 北京: 中国中医药出版社, 2016: 序一, 序二, 60.
- [24] 李灿东, 纪立金, 鲁玉辉, 等. 论中医健康认知理论的逻辑起点. 中华中医药杂志, 2011, 26(1): 109-111
- [25] 李建中, 刘显敏. 大数据的一个重要方面: 数据可用性[J]. 计算机研究与发展, 2013, 50(6): 1147-1162
- [26] 王元卓, 靳小龙, 程学旗. 网络大数据: 现状与展望[J]. 计算机学报, 2013, 36(6): 1125-1138. [21] 李国杰, 程学旗. 大数据研究: 未来科技及经济社会发展的重大战略领域——大数据的研究现状与科学思考[J]. 中国科学院院刊, 2012, 27(6): 647-657.
- [27] 程学旗, 靳小龙, 王元卓, 等. 大数据系统和分析技术综述[J]. 软件学报, 2014, 25(9): 1889-1908
- [28] 白丽娜. 基于 BP 神经网络的中医体质辨识研究[D]. 天津理工大学: 2014
- [29] 徐玮斐, 刘国萍, 王忆勤, 燕海霞, 郭睿. 近 5 年中医证候诊断客观化研究述评[J]. 中医杂志, 2016, 57(5): 442-445.
- [30] 李灿东, 杨雪梅, 甘慧娟, 赖新梅, 周常恩, 陈梅妹. 健康状态辨识模型算法的探讨[J]. 中华中医药杂志, 2011, 26(6): 1351-1355.
- [31] 刘慧婷, 冷新杨, 王利利, 赵鹏. 联合嵌入式多标签分类算法[J]. 自动化学报, 2019, 45(10): 1969-1982.
- [32] 刘海洋, 王志海, 张志东. 基于 ReliefF 剪枝的多标记分类算法[J]. 计算机学报, 2019, 42(3): 483-496.



[33] 刘海洋, 王志海, 张志东. 基于 ReliefF 剪枝的多标记分类算法[J]. 计算机学报, 2019, 42(3): 483-496.

[34] M.L. Zhang, Z.H. Zhou. A review on multi-label learning algorithms[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26 (8): 1819 - 1837.

[35] Y. Zhu, J.T. Kwok, Z.H. Zhou. Multi-label learning with global and local label correlation[J]. IEEE Transactions on Knowledge and Data Engineering, 2018, 30(6): 1081-1094.

[36] K. Bhatia, H. Jain, P. Kar, M. Varma, P. Jain. Sparse local embeddings for extreme multi-label classification[C]. In Advances in Neural Information Processing Systems 28, Montreal, Canada, 2015, pp. 730-738.

[37] C. Xu, D.C. Tao, C. Xu. Robust extreme multi-label learning[C]. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, 2016, pp. 1275-1284.

[38] Y. Tagami. AnnexML: Approximate nearest neighbor search for extreme multi-label classification[C]. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, Canada, 2017, pp. 455-464.

### 三、项目方案（具体研究内容、研究目标、计划、技术路线、人员分工等）

围绕本项目的研究内容和目标，绘制总体技术路线图如下所示：

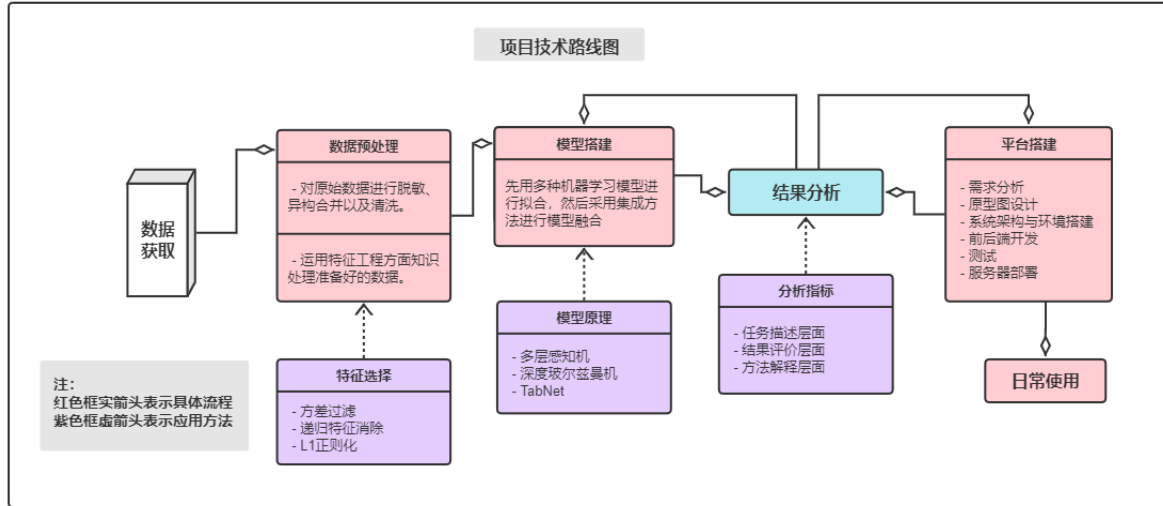


图 1：总体技术路线图

#### 1. 研究目标

首先，基于对退化性骨关节病的风险性因素来研究收集退化性骨关节病患者的数据。然后，为了适应退化性骨关节病人的数据特点，进一步发掘特征选择的隐含信息，需要合理进行模型设计，得出关键风险性因素。接下来，基于中医诊断数据的原始信息及特征选择得出的关键风险性要素，结合数据科学的处理流程和深度神经网络等关键技术，建立用于骨关节炎患者早期诊断的机器学习模型，并根据模型的具体实施结果，结合中医理论体系进行客观的解释和分析。最后，在不断改进完善模型的同时，搭建网络平台雏形，将研究成果部署到小程序或者网站之类能应用于人们日常诊断的平台。在项目研究的后期工作中，不仅需要跟进完善当前模型，也需要更新维护平台。

#### 2. 具体研究内容和技术路线

要在中医诊断领域应用人工智能方法，可将步骤依次分为数据准备、模型构建、结果分析和平台搭建四个部分。对应的研究内容主要包括：

(1) 收集并合理利用退化性骨关节病人的原始数据信息，对数据进行批量汇总和预处理，并通过特征选择方法筛选出关键风险性因素；

(2) 根据退化性骨关节病的风险性因素和诊断结果类型，结合智能辅助诊断的任务特点，设计并训练多种机器学习模型，并通过集成学习进行模型融合来提高预测效果；

(3) 从不同的评价指标出发对不同模型的输出结果进行分析，横向比较不同模型的性能，从整体上客观评价人工智能技术在退化性骨关节病上的辅助诊断效果；

(4) 将人工智能方法辅助诊断的研究成果转换，并部署到小程序、网站等平台中。需要依次进行 I. 需求分析、II. 原型图设计、III. 系统架构与环境搭建、IV. 前后端开发、V. 测试等完整的开发

生命周期，最终将服务器成功部署上线，并基于当前版本的模型进行持续跟进、测试和维护。

## 2.1 数据集介绍和预处理

原数据集为 5027 行和 289 列的表格形式，代表 5027 个潜在病例和 289 个病例特征。在潜在病例中，患病 4417 例，未患病 610 例；在病例特征中，包含了例如医疗机构名称、病史采集日期、性别、年龄、出生日期、职业、身高、体重等用户的基础信息，还有神态、面色、舌色等中医数据以及尿酸、白细胞计数、血小板计数等西医特征。

原数据集从整体上看包含的人体信息很全，但是存在部分特征大量缺失的情况。有 40% 的特征，其数据空缺超过 70%。为了对数据进行清洗，通过把数据集中例如职业、采集日期、医疗机构名称等无用特征删除，将数据集分成两个，一个是直接删除空缺超过 70% 的列，再删除空缺数量超过 10 个的行，得到一个样本较多的，但特征数量较少的数据集一；另一个是保留数据集中所有有用特征，直接删除空白数量超过 20 的行，得到数据集二，两个数据集的样本分类具体如下图：

	列数 (特征)	患病样本数	未患病样本数
原数据	289	4417	610
数据集一	179	2965	373
数据集二	434	1207	301

表 1：数据集样本分类

得到两个数据集后，需要对数据进行补全。为了尽可能地减小空缺数据对模型训练的影响，对连续的数据，采用的是取该列平均值进行补全，离散的数据，采用的是最大投票法进行补全。

为了尽可能得到每一个特征对骨关节病的影响，将具有多个含义的离散特征进行分裂，例如病人膝盖疼痛程度，原数据中只有 1 列表示三个程度，将其分为 3 列，每个表示不同的程度。这样在特征分析的时候可以找到具体是哪种程度的膝盖疼痛会对膝骨关节炎有影响。

## 2.2 数据准备和特征选择

数据准备部分拟分为数据收集和数据处理两个环节。在数据收集中，会将退化性骨关节病人的原始信息进行汇总，对信息进行脱敏处理，对异构数据进行合并，最终整理成表格类型的数据；在数据预处理中，会对原始数据进行清洗，结合实际情况合理选用替代法或舍弃法处理缺失值和异常值，同时对数据进行规范化等其他特征工程处理。在开始正式训练前，为了评估模型的泛化误差，还会将数据以特定比例拆分成训练集和验证集，采用留出法或交叉验证法训练模型并验证效果。

对于已获取的退化性骨关节病的诊断数据，由于原始数据特征较多，会引发维数灾难问题，也不利于后续智能化诊断在接下来将用特征选择来对数据进行降维筛选。为了避免维度灾难、降低学习任务的难度，并且对数据中各个特征在训练任务中的价值高低做出判断并进行分类，这里需要采用特征选择的方法。

特征选择是特征工程里的一个重要问题，其目标是寻找最优特征子集。本项目拟采用过滤式、包裹式和嵌入式的特征选择策略，借鉴主流的特征选择方法，比如：方差过滤、递归特征消除以及 L1 正则化等展开退化性骨关节病的风险性因素研究。具体如下：

其一，先对数据机进行特征选择，然后再训练学习器，特征选择的过程与后续学习器无关。基于过滤式实现特征选择是最为简单和常用的一种方法，其最大优势是不依赖于模型，仅从特征的角度来挖掘其价值高低，从而实现特征排序及选择。以方差过滤为例，在退化性骨关节病的诊断数据中，若某个特征的方差小于某个阈值，则认为该特征可区分性较低，对后续学习器的训练没有太大价值，无法作为退化性骨关节病的风险性因素，予以丢弃。

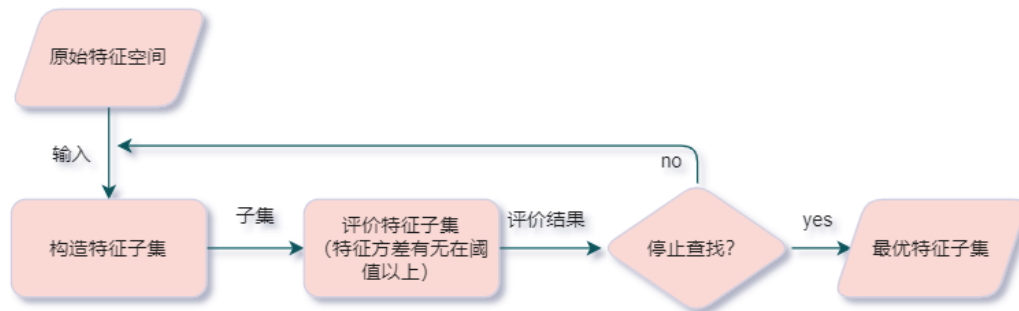


图 2：方差过滤的流程图

其二，特征选择法的特征选择过程与学习器相关，使用学习器的性能作为特征选择的评价准则，选择最有利于学习器性能的特征子集。以递归特征消除法 RFE 为例，首先选择一个外部的学习器，该学习器会对原始数据中退化性骨关节病各个特征赋予一定的权重（如线性模型的系数）作为风险性因素评分，然后将学习器在初始特征集合上训练，获得每个训练特征的风险性因素评分（特征的评分越小意味着该特征越不重要），接着剔除评分最小的特征得到一个新的特征子集，以此循环往复，不断地在新的特征子集上训练学习器，直到特征数目达到所需为止。

最后，将特征选择的过程“附着”于一个模型训练任务本身，从而依赖特定算法模型完成特征选择的过程。嵌入法是一种让算法自己决定使用哪些特征的方法，即特征选择和模型训练同时进行。以带有 L1 正则项的逻辑回归为例，直接将其在所有退化性骨关节病特征上训练，由于 L1 正则化方法具有稀疏解的惩罚特性，逻辑回归得到的特征权重（即风险性因素评分）更加稳定，且一些特征权重非常接近于零，只需将这些特征丢弃，选择相应数目的特征作为风险性因素即可。

考虑到中医诊断的高阶非线性特性与特征的不稳定因素，在退化性骨关节病的风险性因素选择过程中，需要有机结合上述各种方法，利用集成学习的思想综合各种方法的优劣进行特征选择，提高后续模型训练的鲁棒性，实现可靠的智能化诊断。

经过特征选择对数据进行处理后，提取出来的特征要符合高价值、高相关性的要求，提高下游学习模型的性能，实现更精准的诊断；冗杂的特征更少，特征维度更符合患者能够了解的程度。

### 2.3 模型构建

模型构建部分拟先构建多种合适的单一机器学习模型，然后采用集成方法进行模型融合。

在单一模型的选择中，注意到中医诊断的高维性、高阶性和非线性特征决定了其本身的复杂程度，临床症状之间会存在定量和定性混杂、线性与非线性难分的关系。因此若用简单的机器学习方法如多元线性回归、单隐层神经网络等去模拟诊断的思维难以体现中医药的诊断全貌；此外由于模型处理的是表格类型病人数据，数据存在症状复杂、标记成本高、样本较少的特点，若用大规模深

度神经网络去设计模型，则容易陷入过拟合，且有悖于智能辅助诊断对实时性的要求。综上，在模型设计上需要采用能综合考虑预测效果和运行效率的模型。本次研究中拟主要选择多层感知机、深度玻尔兹曼机与 TabNet 等模型作为单一模型分别构建。其原理及可行性论证如下：

**多层感知机：**多层感知机（Multilayer Perceptron）是经典的深度学习模型。感知机模型由若干个神经元组成，其中神经元负责接受来自输入数据或上一层网络传来的输出，经过与自身内部的权重进行运算后，得出数值结果传递给下一层网络。在多层感知机中，内部结构又可分为输入层、隐含层和输出层三层网络结构。其中输入层负责接受输入数据的特征，经过该层的加工后送入隐含层进行处理；隐含层可以包括多层网络，其意义是通过深度和宽度的堆叠，使得隐含层能充分提取输入数据的特征；隐含层的结果将经由输出层送入激活函数，最终得到预测结果。采用多层感知机构建模型，可以由隐含层的权值反映数据的非线性特征，并在训练时通过反向传播算法更新网络节点的权值以进行学习，其模型设计具有普适性，因此多层感知机是适合用于本次项目的模型。

**深度玻尔兹曼机：**深度玻尔兹曼机（Deep Boltzmann Machine）是采用多层受限玻尔兹曼机堆叠形成的深度网络。玻尔兹曼机是一种采用能量定义的网络，网络中包括隐层和显层两层布尔型神经元，网络的能量平衡状态服从 Boltzmann 分布。若仅保留隐层和显层的连接，则形成的二分图结构被称为受限玻尔兹曼机，此时显层可以理解为输入变量，隐层可以理解为输入变量的隐含特征。深度玻尔兹曼机可以通过对比散度学习算法进行训练，以使得隐层根据计算所得到的显层状态与实际状态尽量接近，以达到预测效果。其从能量函数出发的独特角度可以为中医诊断的模型设计带来新思路。

**TabNet：**TabNet 是谷歌于 2020 年提出的专门用于处理表格数据的深度学习模型，其内部网络结构在设计上模仿了决策树模型的在高维流形上的决策方式，因此具有一般神经网络所不具有的内部可解释性，可以呈现出单个特征的重要程度，以及每个特征对预测结果的影响。TabNet 能够进行端到端的预测，不像基于树结构的机器学习模型一样较为依赖输入特征的人工处理。采用 TabNet 能够从针对退化性骨关节病人信息的表格类型数据特点进行学习，其端到端的特性能够提高关键风险性因素选择的容错率，保证中医辅助诊断的鲁棒性。

在最终模型的融合中，拟结合多种单一模型的预测结果，采用集成方法对模型进行融合。这是因为单一模型的构建通常不能完全表达中医诊断的思维规律，难以获得较好的泛化性能。而采用集成方法可以降低预测结果的偏差或方差，发挥多样化模型设计的互补优势，为中医药的现代化之路提供很好的思路。

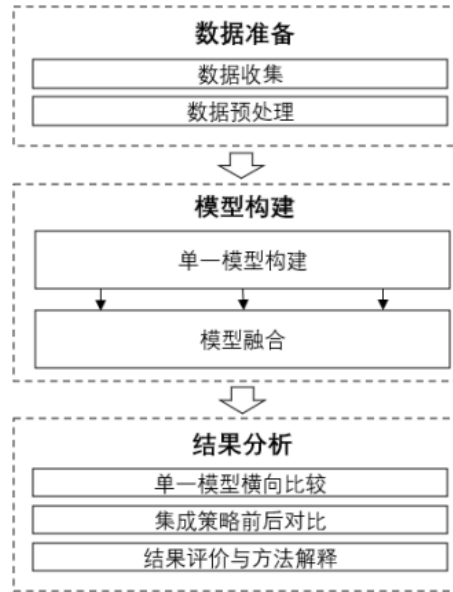


图 3：模型设计流程图

## 2.4 结果分析

要对所构建模型的预测结果进行分析，首先横向比较多层感知机、深度玻尔兹曼机与 TabNet 等多种单一模型的预测效果，分析各自的优势和劣势并验证可行性；然后将模型进行集成，比较单一模型与集成模型的预测效果，评估模型集成策略的科学性，从而发挥多模型中医智能辅助诊断的优势，提高项目的可解释性。

在任务描述上，拟将该机器学习任务定义为二分类任务，对于测试样本中的病例，模型将输出其患退化性骨关节病的概率，若概率超过某个设定阈值则将病例预测为患病，否则为健康。在后期实验中，也可实际预测结果将该任务定义为三分类任务或其他多分类任务，将病例按照患病可能性进行进一步区分，以提高模型的实用性。

在结果评价上，拟绘制混淆矩阵进行精度评价，重点关注准确率、召回率等评价指标。多样化的评价标准有助于从各个方位把握单一模型和集成模型的实际性能，也有助于模型的效果比较和特点区分。

在方法解释上，针对不同模型，拟分别对其预测错误的个别病例在中医背景下进行专门分析，并结合特征选择的实施方式，列出影响退化性骨关节病人患病的关键风险性因素并得出调研结论，以用于指导模型的构建和优化。

## 2.5 平台搭建

### 2.5.1 平台介绍

为了满足人们日常的早期诊断需要，平台将分为两个部分，一个是简单针对退化性骨关节病诊断的小程序，一个是在此基础上内容与功能更为丰富的网站。小程序可以方便对自身患病可能性的预估，而在网站平台上可以了解到更为详实的内容并进行一个长期状态的记录。之所以搭建这样一个平台，是为了使本项目的研究成果真正落实到人们生活中去，真正让群众可以从中受益。

通过将训练得到的神经网络建立一个精准高效的骨关节病预测模型，并使用 flask 框架搭建一个简单的 web 服务，将该模型封装成一个 http 接口挂载在云服务器上。根据特征选择我们会得出一系列的针对于退化性骨关节病诊断的关键性因素，我们通过搭建一个小程序，对用户提出关于这些因素的问题。用户回答后即将这些数据封装为请求参数访问服务器上发布的服务，最终经过该模型的计算得出结果后返回给用户自己患病的可能性。小程序即实现一个简单的诊断器，对数据进行计算并返回结果，并不会存储用户的数据。如果用户希望进行一个长期的跟踪诊断，并希望得到数据的变动分析和该疾病的一些相关内容，则可以选择在网站上注册一个属于自己的账号，之后即可观看到自己的所有的历史诊断结果和数据变动趋势，让用户可以通过平台简单方便地从本项目的研究成果中受益。

### 2.5.2 技术路线

在平台搭建过程中，前端将使用小程序原生、HTML5、CSS3、JavaScript、Vue 等技术并通过 Axios 封装 Ajax 与服务器进行通信，后台使用 Java 进行开发，通过 Springboot 内置 tomcat 进行服务器端的搭建，同时使用 Maven 或 Gradle 等包管理工具、mysql 以及 redis 或 MongoDB 等数据库技术和 druid 连接池、kafka 或 RabbitMQ 等中间件整合 Log4j 以及 Shiro 或 Spring Security 安全框架等辅助开发。

### 2.5.3 技术依据

前端之所以使用 Vue 和 Axios 是为了实现前后端分离，降低整个项目的耦合度，实现高内聚低耦合的设计模式，以便项目的后期迭代和维护。使用 Ajax 技术是为了能够更快地将数据更新呈现在客户端上，而不需要重载（刷新）整个页面，这使得程序能够更快地回应用户的操作，予以用户更好的使用体验。Axios 对 ajax 实现了二次封装，在提供了更丰富完善的功能之外，一定程度的防止了跨域请求伪造和重复恶意请求。

后端选用 Java 语言，利用 SpringBoot 约定大于配置的特点，解决基础 Spring+SpringMVC+Mybatis 配置文件冗杂的问题，简化开发。考虑到该项目并不会存在高并发请求的情况，所以选用了 Tomcat 作为应用服务器。Maven 或 Gradle 等工具的使用是为了解决依赖导入和依赖冲突的问题，并方便项目的构建。至于数据层除了 mysql 之外，还考虑利用 redis 的缓存机制加快数据读写的速度，将常用的数据进行存储以提高系统的性能。使用 kafka 提供的消息队列以保证数据流的顺序读写，降低连接的复杂度，同时配合 Log4j 实现日志信息的输送。最后使用 Shiro 或 Spring Security 框架实现用户认证、授权、会话管理和数据加密存储等功能以确保数据和用户登录的安全性。



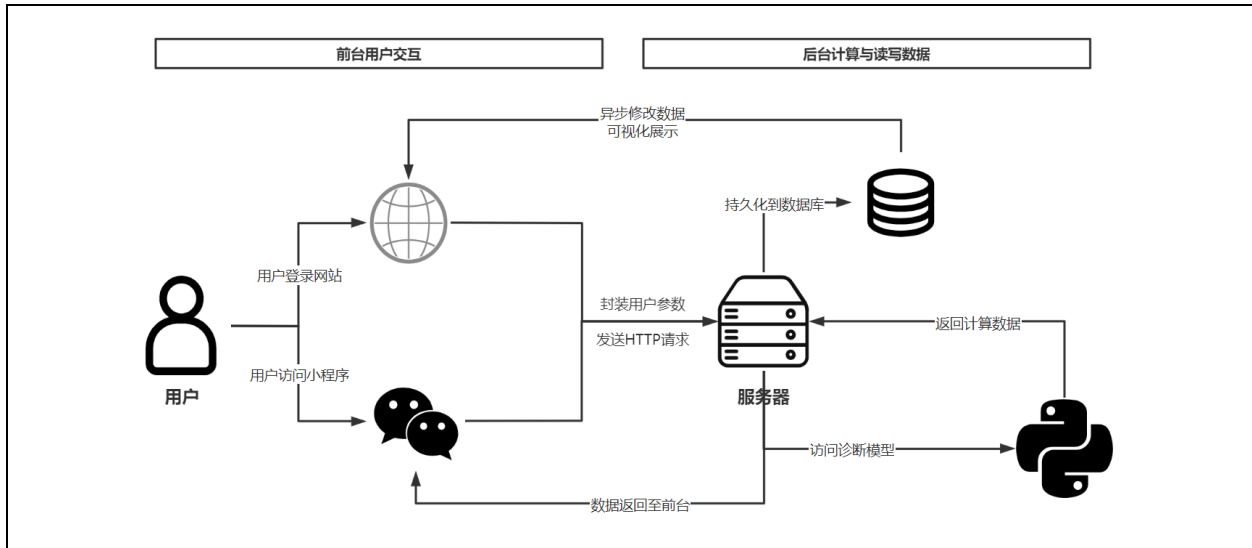


图 4：平台架构示意图

### 3. 预实验的流程和结果

为验证具体研究内容和技术路线的方案可行性，已对所收集数据采用的基础的特征选择和机器学习方法进行了预实验，取得了一定的成果。该结果可以与本次项目得出的实验结果进行对照，并支撑对项目结果的定性及定量分析。

预实验共分为五个部分，依次为特征选择的实施、实验的细节设定、实验结果的得出和讨论、基于特征选择方法得出的关键特征分析以及多标记学习运用的尝试。预实验的具体流程实施方案和获得结果描述如下：

#### 3.1 特征选择的实施步骤及方法

预实验阶段的特征选择实施方法如下：首先，对数据集进行预处理，分为训练集(Training Set)和测试集(Testing Set)。然后，使用不同的特征选择算法对训练集的特征进行打分，通过这个算法可以给每一个特征打出一个分数，分数越高说明特征与是否患病的关联度越高。接下来，取分数最高的 k 个特征作为最后选择的特征子集输出。最后，使用筛选出来的不同数量的特征子集进行分类性能预测，根据评价指标的分数高低来评价特征子集的可用性及其重要性。

在传统的机器学习方法上采用了 7 种不同的特征选择算法，尽可能的把重要的特征子集筛选出来进行分析，分别为：

- (1) 基于信息理论：条件互信息快速二分类特征选择(CMIM)、有监督的神经网络学习中的互信息特征选择(MIFS)
- (2) 基于相似度：R. O. Duda, P. E. Hart, and D. G. Stork, 模式分类(Fisher\_Score)
- (3) 基于稀疏学习：联合 L2,1 范数最小化的高效稳健的特征选择(RFS)
- (4) 基于统计特征选择：基于相关的过滤器与包装器方法的比较(CFS)、Chi2:特征选择和数值属性离散化(chi\_square)
- (5) 基于包装法：变量和特征选择简介(decision\_tree\_forward)



### 3.2 实验设定和检验指标

接下来，分别对数据集 A、B 进行训练/测试集划分，其中训练集占 70%，测试集占 30%，如下图所示：

	数据集 A 178-1508	数据集 B 433 3338
患病 0	1207	2965
未患病 1	301	373
未患病比例	19.96%	11.17%

表 2：数据集划分

在分类方法上，使用 SVM 作为分类器预测性能。对应的性能指标为以下五种：

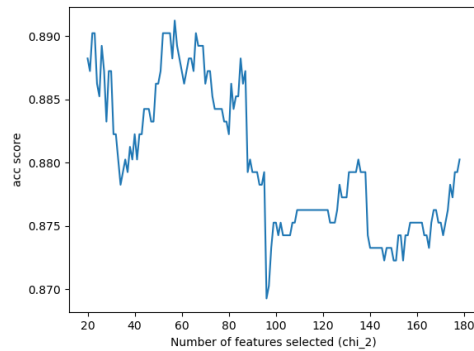
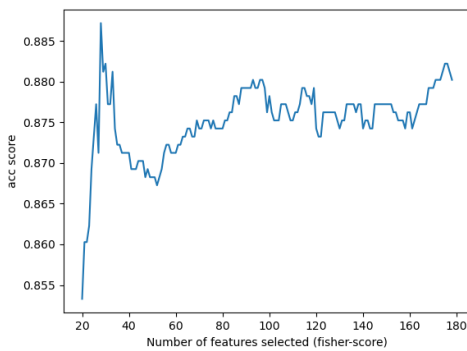
Accuracy	(准确率)
Precision	(精度)
Recall	(召回率)
f1-score	(平衡 F 分数)
AUC	(ROC 曲线下的面积)

表 3：预测性能指标

### 3.3 预实验结果和结论

对拆分出来的两类数据集分别采用上述特征选择方法、实验设定和检验指标，分别得出可视化结果及分析如下：

数据集 A



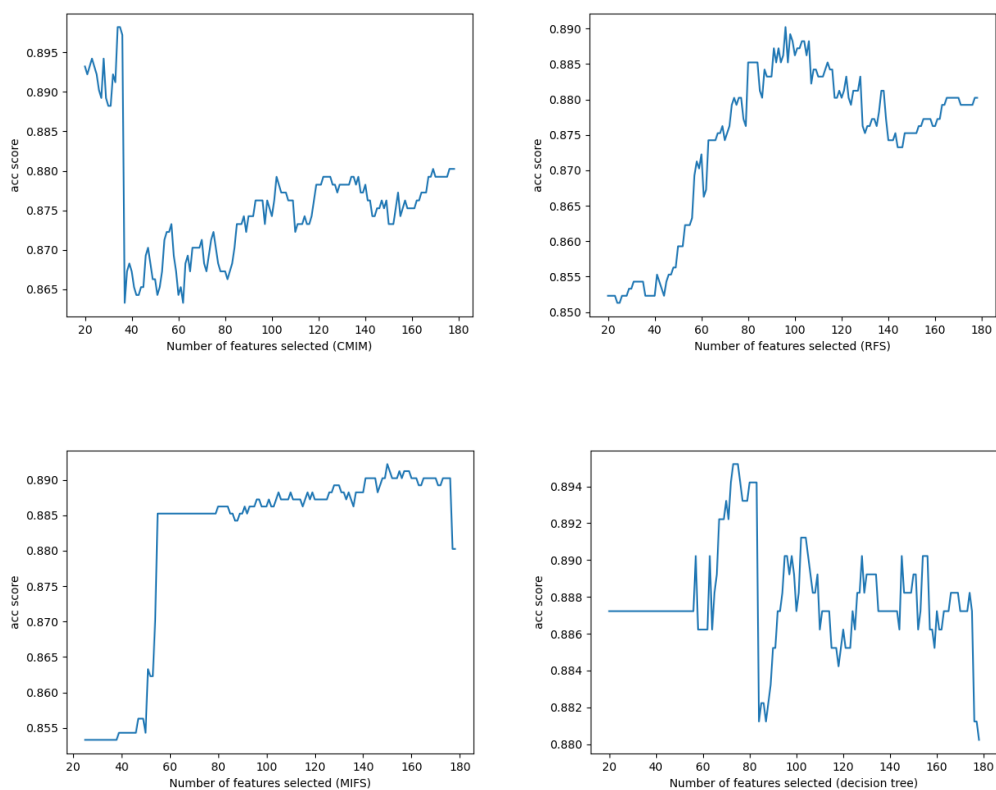


图 5：数据集 A 结果分析

	最优特征数量	acc	pre	recall	f1	auc
chi2	57	0.891	0.53	0.231	0.322	0.603
decision_tree_forward	73	0.895	0.64	0.142	0.233	0.566
Fisher-score	28	0.887	0.489	0.214	0.298	0.593
CFS	14	0.847	0.315	0.312	0.313	0.613
CMIM	34	0.898	0.625	0.223	0.328	0.603
RFS	96	0.89	0.517	0.258	0.345	0.614
MIFS	55	0.885	0.411	0.062	0.108	0.525

表 4：在数据集 A 中各算法性能指标对照表

根据实验结果得知，在数据集 A 中分别利用 7 种特征选择算法筛选出 178 维特征的重要分数，并按特征重要性从大到小排序（由于 RFS 算法比较特殊，固定输出了 14 个特征，因此不做分析）。为了找出每种算法的最优性能，遍历了特征分数排序后的的前 20 到 178 维特征，acc 平均精度都能在 89%左右，最优性能的特征数量在 14~96 之间，相比起原始的 178 维特征减少了大约 50%。其中，CMIM 算法只需要用 34 个特征，就可达到 89.8%的 acc 精度。

综上所述，在数据集 A 中不到原始数据集 50%的特征就能以高准确度的预测患者是否患病。

## 数据集 B

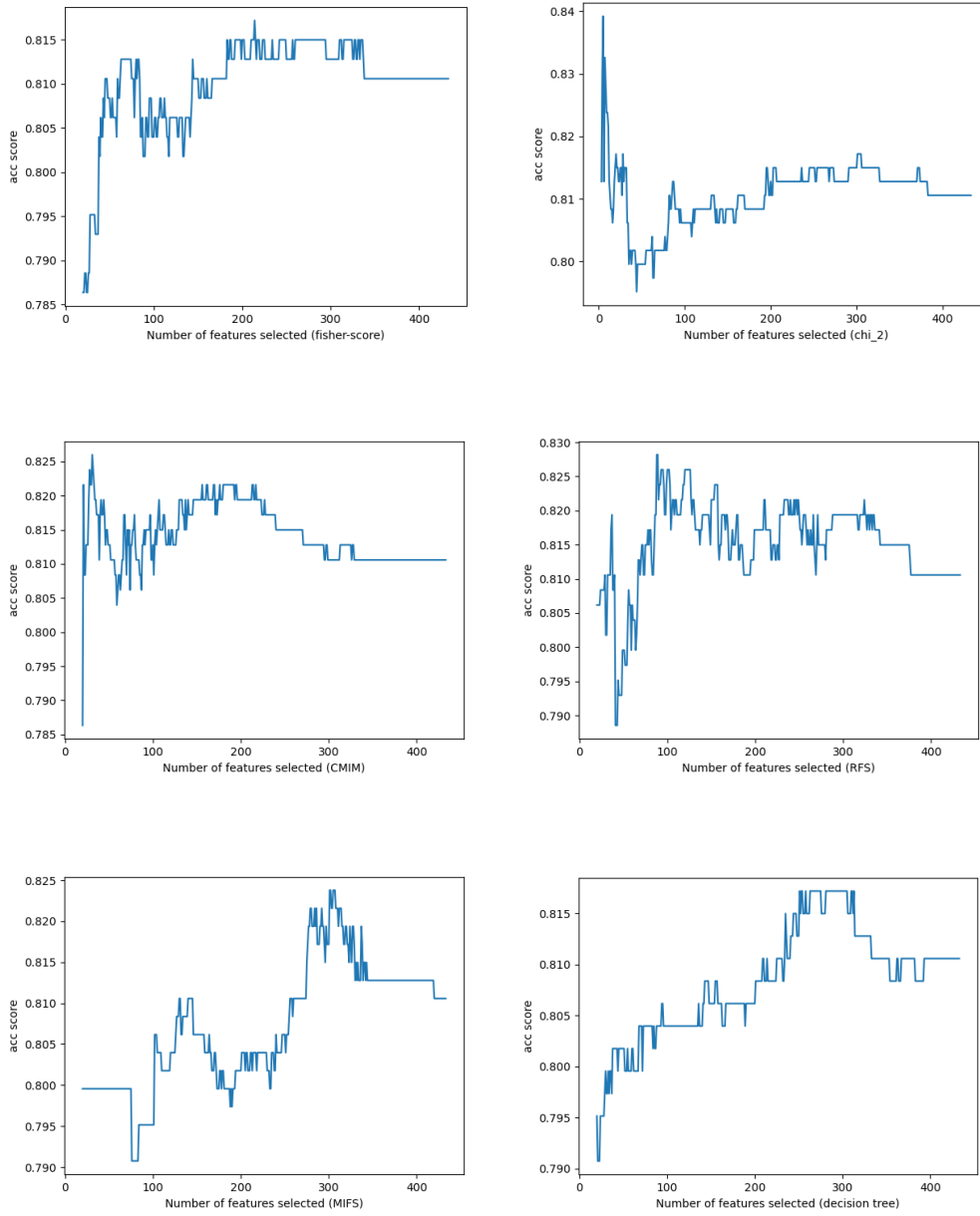


图 6: 数据集 B 结果分析

	最优特征数量	acc	pre	recall	f1	auc
chi2	5	0.839	0.821	0.252	0.386	0.619
decision_tree_forward	251	0.817	1	0.087	0.161	0.543
Fisher-score	63	0.812	0.75	0.098	0.174	0.545
CFS	27	0.817	0.666	0.175	0.278	0.576
CMIM	31	0.825	0.833	0.164	0.275	0.578
RFS	88	0.828	0.822	0.164	0.277	0.579
MIFS	301	0.823	1	0.12	0.215	0.56

表 5: 在数据集 B 中各算法性能指标对照表

在数据集 B 中的实验方法和条件与数据集 A 一致, 为了找出每种算法的最优性能, 遍历了特征分数排序后的的前 20 到 433 维特征, acc 平均精度都能在 82%左右, 最优性能的特征数量波动幅度较大, 在 5~301 之间, 大部分算法相比起原始的特征还是减少了很多。其中, 卡方 chi\_2 算法只需要用 5 个特征, 就可达到 83.9%的 acc 精度。

综上所述, 本项目所构建的特征选择方法路线基本可行, 且已在传统机器学习方法 SVM 里得到了验证。在项目后续的研究中, 将会采用深度学习算法进行多次实验, 并对预实验的结果进行比较和分析。

### 3.4 特征分析

为了对 7 种特征选择算法筛选出的特征子集中做特征分析, 统计出出现频率最高的特征并对其进行排序, 结果见下表:

数据集 A 特征	物理含义	重复次数
commonMz11	诊脉 11	7
commonMs0	诊脉 0	7
ggjbTtxz12	疼痛性质 12: 痛处固定	6
commonJzqk0	居住情况 0: 不祥	6
commonSs3	舌色 3	6
commonWn1	胃纳	6
ggjbXx4	X 线 K&L 分级 4:0 级	6
ggjbJbjzhjl2	居住楼层: 无电梯五楼以上	6
ggjbSfyzz0		5
commonSttz5	苔质 5	5
commonSxmlxt1	舌下脉络形态 1	5
commonZxlxbfb	中性粒细胞百分比	5

表 6: 数据 A 中特征出现频率统计表

数据集 B 特征	物理含义	重复次数
commonMz11	诊脉 11	7
commonSs3	舌色 3	7
ggjbTtxz12	疼痛性质 12: 痛处固定	6
ggjbjzhjqh0	居住环境 0: 气候干燥	5
commonTzFbfmsr3		5
ggjbZz6		5
commonNsh1		4
commonTzFbfmsr0		4
commonTzFbfmsr4		4
commonMz3	诊脉 3	4
commonWn1	胃纳 1	4
commonDmdzdbc	低密度脂蛋白 C	4
commonTzGjxmb10		4
commonDbcs3		4
commonTzMbhby3		4
commonTzGdkgsz4		4
commonTzTsz0	痰湿质 0	4

表 7: 数据 B 中特征出现频率统计表

在数据集 A 中, 脉诊 11 和脉诊 0 在 7 种算法中都曾出现, 而在数据集 B 中则是脉诊 11 和舌色 3, 由此可见脉诊 11 是一个比较重要的患病因素之一, 因为它同时高频的出现在不同的算法中。

交集特征	物理含义	重复次数
ggjbTtxz12	疼痛性质 12: 痛处固定	2
commonMz11	诊脉 11	2
commonSs3	舌色 3	2
commonWn1	胃纳 1	2

表 8: 数据集 A 和 B 的交集特征

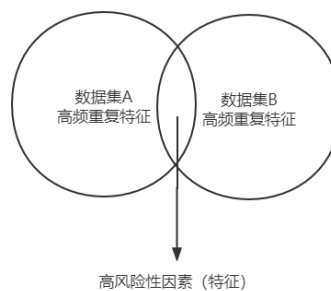


图 7: 高风险因素示意图

最后取出数据集 A 和数据集 B 的高频重复特征作交集, 得出了 4 个关键性特征, 分别是疼痛性质 12、脉诊 11、舌色 3 和胃纳 1, 这 4 个特征相比起单个数据集的重复特征具有更好的泛化性和可靠性, 并可以作为诊断膝骨关节病的重要因素。在本项目接下来的工作中, 将基于具体研究内容和技术路线部分的特征选择方案和机器学习模型, 对以上 4 个特征的重要性进行进一步量化评估。

### 3.5 多标记学习应用尝试

为了实现对骨关节炎的平和质、气虚质、阳虚质、阴虚质、痰湿质、湿热质、血瘀质、气郁质、特禀质等中医特征的多标记预测。多标记算法是为了解决真实世界中对象的描述与及概念标签之间的一对多关系而提出的。它从输出空间，也就是概念标记上来考察对象的歧义性。对骨关节数据采用多标记学习有助解决为一个病例同时诊断多种类别的症状的复杂情况提供解决方案。

在本次预实验的实施中，尝试采用 BRknn、MLknn、BinaryRelevance、ClassifierChain 等多种不同的多标记算法进行模型构建。这些算法封装在机器学习库 sklearn 中，可直接调用来处理数据以为多标记学习的探究提供便利。实验的性能指标采用 Hamming Loss、One-error、Coverage、Ranking Loss、Average Precision 等评价标准，其中除了 Average Precision 是结果越大越好，其余的指标都是结果越小越好。最终结果如图所示：

	Hamming Loss	One-error	Coverage	Ranking Loss	Average Precision
BRkNN	0.05593220338983051	0.36949152542372876	3.91864406779661	0.16572571966639763	0.6455954394785808
MLkNN	0.05694915254237288	0.3728813559322034	3.7254237288135594	0.150363196125908	0.6322005091791821
BR_SVM	0.03694915254237288	0.29491525423728815	3.647457627118644	0.133425477535647	0.7215251889904695
CC_SVM	0.03728813559322034	0.3016949152542373	3.6779661016949152	0.1351203927898843	0.7196580430922062
LP_rf	0.05898305084745763	0.3627118644067797	3.9728813559322034	0.17240193704600484	0.611376351651819
RakelD	0.1840677966101695	0.9322033898305084	4.2983050847457624	0.26803981705676627	0.428849788222752
MajorityVoting	0.228135593220339	0.9627118644067797	5.647457627118644	0.39517258541834815	0.3206957835398654
matrix-based clusterer	0.04813559322033898	0.3254237288135593	3.7254237288135594	0.15483360236750066	0.6332616096448452

表 8：各多标记算法性能指标对照表

由结果可知，嵌入 SVM 的多标签算法，在结果上都会优于其他算法，其平均精度可达到 70%以上，证实了多标记学习运用在骨关节炎数据中的可行性。本项目后续的研究会基于该结果作为参照，进一步挖掘不同算法运用于骨关节炎中医辅助诊断的优劣势，结合量化标准从理论上和实际表现上进行比较和评估。

## 4. 人员分工

刘桢江：机器学习模型训练与调优。

周旭林：探究技术路线，模型设计和完善。

陈彦亨：平台开发，研究成果转化。

阮炜霖：项目申报，撰写项目文档，汇报项目进度。

黄崇铎：探索关键发病要素，对数据进行特征选择等处理。

## 四、项目特色与创新点

人工智能技术与智能辅助诊断结合，可以降低中医诊断领域的误诊率，提高中医诊断的准确度。将传统中医“望闻问切”的方法与先进的特征选择、深度学习等方法结合，是推动中医现代化发展的重大举措，也是该项目的主要特色和创新点。开展面向中医行业的人工智能模型和算法的研发，推进在退化性骨关节病的早期诊断场景形成应用创新和应用方案，是国家重点支持领域项目申报的范畴，也是本项目团队致力研究的目标。

在理论方面，该项目遵循了中医思维规律，关注患者健康状况的各种风险性因素，以个性化诊断与评估为核心，深度融合人工智能、云计算、大数据等技术，在系统整理中医的科学知识和经验知识，建立符合计算机语言的数据结构及知识规则库的基础上，构建一个基于深度学习和特征工程的网络模型，并运用大数据、云计算的理论搭建退化性骨关节病的中医智能辅助早期诊断平台。依赖互联网实现中医远程诊疗，从而有效弥补中医技术短板，充分发挥中医简便廉验的特点。

在技术方面，该项目结合大数据、云计算和人工智能技术，采用特征工程、深度学习的原理实现从原数据采集降维，到模型构建和完善、再到成果转换和部署的一系列过程，实现中医诊疗客观化、智能化，为中医治疗退化性骨关节病的理论和临床研究提供技术支撑。

## 五、项目进度安排

### 前期准备 2021.12.16—2022.2.20

参考往届优秀项目的经验，做好立项计划，了解项目申报的流程。准备与学习与项目相关的知识，做到能够在项目中熟练的运用。

### 项目申报 2022.2.21—2022.3.12

收集汇总退化性骨关节病的相关数据，调研中医辅助诊断的模型设计方法，形成初步模型设计方案。完成并改进项目申报书，确定项目的研究方向以及实施方案，描述各部分的具体计划，完成项目申报工作。

### 开题报告

根据初步设计方案和数据特点，按照技术路线对收集到的数据进行预处理，提炼出反映骨关节病的特征，然后由此构建深度学习模型雏形以及搭建网络平台的雏形。

### 中期检查

保证数据的质量，并对已经写好的程序进行反复调试；确保提取出来的特征更能精确的符合预期要求；确保构建的网络模型能准确判别退化性骨关节病的风险性因素；确保搭建的网络平台能够稳定、快捷地使用；制作 ppt，准备答辩材料，将项目成果的各项数据展示出来。

### 结项报告

对数据预处理方式、模型构建方式、模型参数和超参数等进行微调，对最终成果不断进行测试，继续完善和巩固项目成果，逐渐将项目成果投入到应用当中；整理、汇总所构建的模型和分析结果，形成结项成果，对项目的完成情况进行总结，写好项目结项书。

## 六、经费预算及依据

本项目申请经费共计 3000 元，主要通途在于科学调研、论文发表等，具体使用情况如下：

经费支出项目	金额（元）	说明
图书资料费	500	用于购买资料查阅
平台搭建费	500	用于平台开始部署
项目维护费	500	用于模型和服务器维护
咨询费	300	用于技术咨询
数据采集费	200	用于数据收集和质量保证
其他费用	1000	
总计	3000	主要用于项目设计和研发

## 七、项目预期成果

1. 将完成对退化性骨关节病患者的数据收集和处理，并通过特征选择、深度学习等技术构建精简高效的网络模型，并且完成平台雏形的搭建。
2. 将持续进行对患者数据的采集、模型的设计和更新、诊断平台的搭建和维护，将机器学习模型对数据的分析结果用于跟进模型以及维护平台两方面。
3. 将完成关于该项目课题在研究生命周期里的相关研究报告，包括开题报告、中期检查报告和结项报告等。

## 八、申请人承诺

我保证上述填报内容的真实性。如果获得资助，我与本项目组成员将严格遵守有关规定，切实保障研究工作的进行，按时报送有关材料。

项目负责人（签名）：



**九、指导教师意见：**

本项目关注中医智能化，能够结合中医优势病种开展研究，具有重要的现实意义。对于所研究的内容调研详尽，技术路线清晰，预期成果可期，总的看来具有较强的创新性，可行性，以及独特性。望得到资助。

签 名：



**十、学院（部）意见：**

签名盖章：

**十一、学校意见：**

签名盖章：