

## • 研究报告 •

# 基于多标记深度森林的膝骨关节炎 智能辅助诊断方法

龙锦益<sup>1,2,3</sup>, 杨宇<sup>1</sup>, 张子龙<sup>1</sup>, 叶倩云<sup>2,4</sup>, 吴汉瑞<sup>1</sup>, 张荣华<sup>2,5,6</sup>, 张佳<sup>1,2</sup>

(<sup>1</sup>暨南大学信息科学技术学院, 广州 510632; <sup>2</sup>暨南大学广东省中医药信息技术重点实验室, 广州 510632; <sup>3</sup>广州琶洲实验室, 广州 510335; <sup>4</sup>暨南大学中医学院, 广州 510632; <sup>5</sup>暨南大学药学院, 广州 510632; <sup>6</sup>暨南大学癌症研究所, 广州 510632)

**摘要:** 目的: 利用多标记深度森林 (MLDF) 算法构建膝骨关节炎 (KOA) 智能辅助诊断模型, 并探索多标记方法在中医数据集上的优势。方法: 基于1 421例临床样本, 使用MLDF算法构建分类模型, 在6个评价指标上与其他5种多标记算法进行对比; 使用多标记算法Rank-SVM、ML-kNN和单标记算法SVM、kNN建模并对比。结果: 使用MLDF构建的分类模型在6个评价指标上均优于其他5种对比算法, 并且在KOA标记上的AUC为0.8122, 远高于其他对比算法; Rank-SVM与ML-kNN分类准确率 (0.7746、0.7787) 高于其对应的单标记算法 (0.7641、0.7570), 且在大多数评价指标上均优于其对应的单标记算法。结论: 在多证兼夹的中医数据集上, 多标记分类算法性能优于其对应的单标记算法, MLDF算法在KOA的诊断结果上与真实诊断结果的一致性较好, 具有较好的推广和应用前景。

**关键词:** 膝骨关节炎; 人工智能; 机器学习; 多标记学习; 多标记深度森林; 智能辅助诊断模型

## Intelligent assisted diagnosis method of knee osteoarthritis based on multi-label deep forest

LONG Jinyi<sup>1,2,3</sup>, YANG Yu<sup>1</sup>, ZHANG Zilong<sup>1</sup>, YE Qianyun<sup>2,4</sup>, WU Hanrui<sup>1</sup>,  
ZHANG Ronghua<sup>2,5,6</sup>, ZHANG Jia<sup>1,2</sup>

(<sup>1</sup>College of Information Science and Technology, Jinan University, Guangzhou 510632, China; <sup>2</sup>Guangdong Key Laboratory of Traditional Chinese Medicine Information Technology, Jinan University, Guangzhou 510632, China; <sup>3</sup>Pazhou Laboratory, Guangzhou 510335, China; <sup>4</sup>College of Traditional Chinese Medicine, Jinan University, Guangzhou 510632, China; <sup>5</sup>College of Pharmacy, Jinan University, Guangzhou 510632, China; <sup>6</sup>Cancer Research Institution, Jinan University, Guangzhou 510632, China)

**Abstract:** Objective: To construct an intelligent assisted diagnosis model for knee osteoarthritis (KOA) using a multi-label learning with deep forest (MLDF) algorithm and explore the advantages of the multi-label method in the traditional Chinese medicine (TCM) data set. Methods: Based on 1 421 clinical samples, a classification model was constructed using a MLDF algorithm and compared with five other multi-marker algorithms on six evaluation indexes. Multi-label algorithms Rank-SVM and ML-kNN and single-label algorithms SVM and kNN were used for modeling and comparison. Results: The classification model built with the MLDF algorithm outperformed the comparison algorithms on all six evaluation metrics. The AUC for KOA was 0.8122, significantly higher than other comparison algorithms. Rank-SVM and ML-kNN showed higher classification accuracies (0.7446, 0.7787) compared to their corresponding single-label algorithm (0.7641, 0.7570), and they also outperformed these single-label methods in most evaluation metrics. Conclusion: On the TCM data set with multiple syndromes and clips, the performance of the multi-label classification algorithm is better than that of its corresponding single-label algorithm. The diagnosis results of the MLDF algorithm in KOA are in good agreement with the confirmed diagnosis, and it has good promotion and application prospects.

**Keywords:** Knee osteoarthritis (KOA); Artificial intelligence; Machine learning; Multi-label learning; Multi-label learning with deep forest (MLDF); Intelligent assisted diagnosis model

基金资助: 国家自然科学基金项目 (No.62106084, No.61773179), 广东省自然科学基金项目 (No.2022A1515010468, No.2019A1515012175), 广州市科技计划项目 (No.202201010498), 广东省中医药信息化重点实验室 (No.2021B1212040007)

通信作者: 张佳, 广东省广州市天河区黄埔大道西601号暨南大学信息科学技术学院, 邮编: 510632, 电话: 020-85220010

E-mail: jiazhang@jnu.edu.cn

膝关节关节炎(knee osteoarthritis, KOA)是退行性骨关节病中常见的类型,是老年人常见疾病,其中又以女性、肥胖者居多,其临床的主要表现为关节疼痛、肿胀、活动受限等,严重时会使患者致残<sup>[1]</sup>。随着我国人口老龄化加剧,KOA对人体健康和社会经济带来巨大负担。随着信息技术高速发展,人工智能广泛应用于医疗领域,为许多疾病诊治提供新的解决方法<sup>[2]</sup>。近10年来,机器学习算法开始应用于KOA预测领域<sup>[3]</sup>,但大部分都侧重于影像学或者生物指标,而二者往往因为价格昂贵,采集困难,导致临床转化应用受限。中医认为“有诸内者,必形诸外”,通过中医四诊捕获的临床信息对于预测疾病的发生、发展和预后具有重要价值,且中医指标易于采集,成本低。因此本文提出结合患者中医四诊信息,利用辨证论治理论,使用多标记机器学习算法构建KOA辅助诊断模型。

KOA的临床诊断通常基于患者的既往病史、体征和症状以及影像学检查。磁共振成像、关节镜检查、高频彩超和热构图是接受程度最高的4种影像学检查方式<sup>[4]</sup>。然而,这些基于图像的KOA诊断方式对老年人来说既麻烦又昂贵。既往研究表明,中医四诊信息有利于各类疾病的诊断和预防,如慢性乙型肝炎综合征<sup>[5]</sup>、冠心病<sup>[6]</sup>等。中医治疗KOA具有疗效可靠、价格合理、不良反应少等优点,是中医诊治的优势病种<sup>[7]</sup>,对老年人来说更方便、经济和安全。

多标记分类是机器学习中一个重要研究方向,广泛应用于多语义分割问题。在中医临床中,患者常表现出多证型兼夹的情况,这种复杂的中医诊断过程正是典型的多标记学习问题<sup>[8]</sup>。多标记学习又分为低阶和高阶策略。如一个患者可能同时存在气虚、气滞血瘀和心脾不交等多个中医证型,低阶策略把这种多证型兼夹的情况简化为单个证型或者两两组合进行处理,忽略了其他证型对目标证型的影响,而高阶策略则能充分考虑标记间的相关性,更符合中医临床的多证型兼夹情况。

基于此,笔者提出了一种基于多标记深度森林(multi-label learning with deep forest, MLDF)<sup>[9]</sup>的KOA智能辅助诊断方法。MLDF是一种基于多标记树的非神经网络风格的深层模型,底层使用高阶策略实现,其标记相关性则是通过逐层训练,重用上层表现好的特征实现。相较传统的神经网络分类算法,MLDF在训练过程中以依赖数据结果的方式控制模型增长,这使得算法能适用于不同规模的训练数据。故MLDF特有的结构以及在训练过程中对数据的处理方式可以有效挖掘中医数据中复杂的证型兼夹关系,从而得到更优的分类结果。

## 资料与方法

### 1. 数据集

1.1 一般资料 本研究收集2011—2021年暨南大学附属第一医院、佛山市中医院、五邑中医院、湖南中医药大学附属第一医院的门诊和住院患者5 013例,其中非KOA患者601例,KOA患者4 412例。所有患者均知晓并签署知情同意书。数据集包含

556个变量,其中KOA、平和质、气虚质、阳虚质、阴虚质、痰湿质、湿热质、血瘀质、气郁质以及特禀质10个变量作为标记。

1.2 诊断标准 KOA诊断标准参照《膝关节关节炎中医诊疗指南(2020年版)》<sup>[10]</sup>中Kellgren-Lawrence影像分级。

1.3 数据预处理 为了降低数据的稀疏性,控制正负样本比例,加快模型收敛速度,对数据集进行降采样、空缺值填补、特征分解、对频数为零的特征进行剔除以及归一化处理。最后得到1 421例样本作为训练集,其中KOA患者样本1 111例,男性309例,女性802例,平均年龄为 $(66.40 \pm 11.58)$ 岁,平均体重 $(59.51 \pm 10.64)$ kg。非KOA患者样本310例,男性122例,女性188例,平均年龄 $(62.48 \pm 13.84)$ 岁,平均体重 $(60.26 \pm 11.57)$ kg。

2. 方法 由多层级联森林组成的MLDF结构如图1所示,其每一层级联森林由基于预测聚类树的随机森林(蓝色)和完全随机森林(黑色)两类森林构成,每一个森林模块都有100棵预测聚类树<sup>[11]</sup>。MLDF的深度由其内部度量机制决定,在模型训练前设置最大层数max,在每一层训练结束时模型内部评价当前层效果是否提升,如果连续3层没有提升则模型停止增长,最后选取效果最好的层作为MLDF的输出层,也就是说模型最终输出层数不会大于初始设定的max。

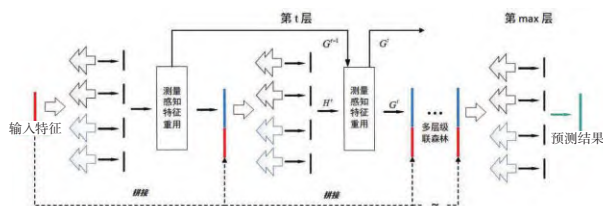


图1 MLDF结构图

MLDF训练时,患者数据经过第 $t$ 层级联森林得到特征表示 $H_t$ ,测量感知特征重用部分根据其内部的性能评价机制使用上一层的表征 $G^{t-1}$ 来更新 $H_t$ ,得到新的表征 $G^t$ ,再与原始的患者数据拼接在一起作为下一层的输入,直到模型收敛停止增长。

本研究由两部分组成:①使用MLDF构建辅助诊断模型,与排序支持向量机(rank support vector machines, Rank-SVM)、多标记K最近邻(multilabel learning k nearest neighbor, ML-kNN)、多标记神经网络(multilabel neural networks, BP-MLL)、多标记自适应共振关联映射神经网络(multilabel adaptive resonance associative map neural networks, ML-ARAM)、二元相关性(binary relevance, BR)5种低阶多标记学习算法<sup>[12]</sup>做对比,在6个评价指标上评估MLDF构建的辅助诊断模型的有效性。②探究单标记算法与多标记算法在中医多证型兼夹数据集上的表现,选取单标记算法kNN、SVM与其对应的多标记算法ML-kNN和Rank-SVM进行比较。

为了保证实验结果的无偏性,本研究训练时采用五折交叉验证。即把数据集随机分为5等份,每次选取1份作为测试集,

其余部分作为训练集,重复该过程5次,最终得到5次结果的均值,需要说明的是,测试集之间是互斥的,所以能有效覆盖整个数据集。

3. 评价指标 使用汉明损失、1-错误率、排序损失、覆盖率、平均精度等5个多标记学习中常用评价指标以及准确率来评估其效果<sup>[12]</sup>。其中,准确率和平均精度数值越大,评价效果越好;汉明损失、1-错误率、覆盖率、排序损失数值越小,评价效果越好。

结果

1. 单标记算法和其对应的多标记算法在10个标记上的准确率比较 见图2。Rank-SVM和ML-kNN的分类准确率分别为0.7746、0.7787,均高于其对应的单标记算法SVM(0.7641)和kNN(0.7570),并且在其他9种中医体质的分类预测中多标记算法普遍优于其对应的单标记算法。

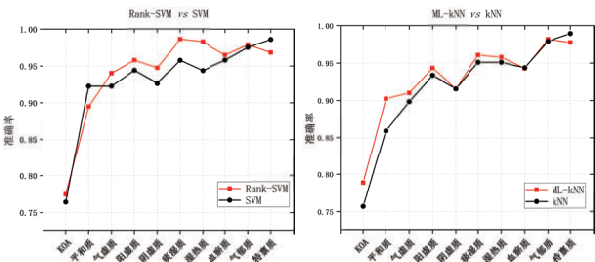


图2 多标记算法与单标记算法在10个标记上的对比

2. 多标记算法与其对应的单标记算法在6个评价指标的比较 见表1。ML-kNN在1-错误率、覆盖率上的性能表现略差于kNN, Rank-SVM则在所有指标上的性能表现均优于SVM。故基于单标记算法改进的多标记算法总体性能更优。

表1 单标记算法与多标记算法在6个评价指标上的对比

| 评价指标  | kNN    | ML-kNN | SVM    | Rank-SVM |
|-------|--------|--------|--------|----------|
| 准确率   | 0.9176 | 0.9279 | 0.9299 | 0.9394   |
| 汉明损失  | 0.0736 | 0.0720 | 0.0730 | 0.0605   |
| 1-错误率 | 0.4084 | 0.4295 | 0.4514 | 0.4084   |
| 覆盖率   | 4.2971 | 4.3330 | 5.0450 | 4.4330   |
| 排序损失  | 0.2087 | 0.2063 | 0.2720 | 0.2020   |
| 平均精度  | 0.6113 | 0.6167 | 0.5312 | 0.7143   |

3. MLDF与BP-MLL、ML-ARAM、BR、ML-kNN和Rank-SVM 5种多标记算法比较 见表2。使用高阶策略的MLDF算法在6个评价指标上均优于其他算法,可见MLDF能更有效地挖掘出KOA与其他标记间的数据关系,从而提高模型分类效果。

4. MLDF与BP-MLL、ML-ARAM、BR、ML-kNN、Rank-SVM 5种算法在10个标记下的ROC曲线下与坐标轴围成的面积(AUC)对比 见表3。MLDF在各个标记下的AUC结果均高于

表2 6种多标记分类算法在6个评价指标上的对比

| 评价指标  | MLDF   | BP-MLL | ML-ARAM | BR     | ML-kNN | Rank-SVM |
|-------|--------|--------|---------|--------|--------|----------|
| 准确率   | 0.9542 | 0.5354 | 0.9191  | 0.9263 | 0.9279 | 0.9394   |
| 汉明损失  | 0.0457 | 0.4645 | 0.0808  | 0.0736 | 0.0720 | 0.0605   |
| 1-错误率 | 0.3521 | 0.8732 | 0.4429  | 0.4084 | 0.4295 | 0.4084   |
| 覆盖率   | 4.0774 | 7.6739 | 4.5021  | 4.2971 | 4.3330 | 4.4330   |
| 排序损失  | 0.1582 | 0.6455 | 0.2248  | 0.2087 | 0.2063 | 0.2020   |
| 平均精度  | 0.7394 | 0.2937 | 0.5883  | 0.6113 | 0.6167 | 0.7143   |

表3 6种多标记分类算法在10个标记上的AUC对比

| 标记  | MLDF   | BP-MLL | ML-ARAM | BR     | ML-kNN | Rank-SVM |
|-----|--------|--------|---------|--------|--------|----------|
| KOA | 0.8122 | 0.5526 | 0.7241  | 0.7556 | 0.6889 | 0.7746   |
| 平和质 | 0.9685 | 0.4857 | 0.9267  | 0.9473 | 0.9485 | 0.8943   |
| 气虚质 | 0.9859 | 0.5288 | 0.9220  | 0.9435 | 0.9485 | 0.9401   |
| 阳虚质 | 0.9991 | 0.6268 | 0.9597  | 0.9677 | 0.9642 | 0.9577   |
| 阴虚质 | 0.9957 | 0.5104 | 0.9172  | 0.9513 | 0.9523 | 0.9471   |
| 痰湿质 | 0.9981 | 0.5552 | 0.8371  | 0.8981 | 0.9757 | 0.9859   |
| 湿热质 | 0.9978 | 0.5239 | 0.9096  | 0.9263 | 0.9762 | 0.9823   |
| 血瘀质 | 0.9926 | 0.5378 | 0.8596  | 0.9028 | 0.9652 | 0.9647   |
| 气郁质 | 0.9934 | 0.4338 | 0.7587  | 0.8518 | 0.9561 | 0.9788   |
| 特禀质 | 0.9800 | 0.6499 | 0.7272  | 0.7410 | 0.7011 | 0.9683   |
| 平均值 | 0.9723 | 0.5405 | 0.8542  | 0.8885 | 0.9077 | 0.9394   |

0.8, 在阳虚质的分类上达到最大值0.9991, 10个标记平均AUC为0.9723, 其表现全面优于其他5种算法。

5. MLDF和其他5种算法在6个评价指标上配对t检验P值比较 见表4。结果P值均<0.05, 表明MLDF算法的优势在统计上是可信的, 验证了MLDF算法的性能优势。从已有的几个对比算法来看, BR算法的P值相对较小, 这说明MLDF算法相对BR算法具有更好的性能, 而BR算法的基本思想是将多标记问题转化为多个独立的二分类问题, 每个问题对应一个标记。进一步验证了考虑标记间相关性的高阶多标记算法会比不考虑标记相关性的算法有着更好的分类效果。

表4 MLDF与其他5种分类算法配对t检验P值比较

| 评价指标  | BP-MLL | ML-ARAM | BR     | ML-kNN | Rank-SVM |
|-------|--------|---------|--------|--------|----------|
| 准确率   | 0.0149 | 0.0022  | 0.0037 | 0.0169 | 0.0394   |
| 汉明损失  | 0.0170 | 0.0016  | 0.0014 | 0.0058 | 0.0220   |
| 1-错误率 | 0.0194 | 0.0099  | 0.0095 | 0.0204 | 0.0250   |
| 覆盖率   | 0.0299 | 0.0049  | 0.0022 | 0.0319 | 0.0202   |
| 排序损失  | 0.0234 | 0.0162  | 0.0093 | 0.0348 | 0.0451   |
| 平均精度  | 0.0286 | 0.0119  | 0.0037 | 0.0233 | 0.0352   |

## 讨论

中医药在防治重大慢性疑难疾病中具有独特优势。人工智能和中医药的交叉融合为中医药数字化防控重大疑难疾病和慢性疾病提供了突破口。多标记学习是人工智能的一个重要研究方向, 符合中医辨证论治思维, 有利于解决多证兼夹问题。在kNN和SVM两个单标记算法中, 将KOA和9种中医体质分解为多个单标记问题进行预测, 然后将结果进行组合。在这个过程中丢失了标记间的潜在关系。然而, 在ML-kNN和Rank-SVM两种多标记算法中, 保留了标记间的潜在关系, 因此, ML-kNN和Rank-SVM的预测精度理论会优于kNN和SVM, 研究结果也证实了这一点。同时, 通过单标记算法与其对应的多标记算法比较, 揭示了使用多标记算法考虑KOA与体质间相关性以及症状兼夹性, 可以提高对KOA的诊断精度。

不同多标记算法具有各自不同的底层策略。MLDF作为一种较新的高阶多标记算法, 充分利用层级结构来挖掘KOA与其他症状之间的内在关系, 从而提高整体预测效果。本研究基于无创、低成本、易采集的中医学指标和危险因素指标, 使用MLDF算法构建了KOA辅助诊断模型, 并与其他5种低阶多标记算法进行了比较。结果显示, MLDF算法在KOA的诊断上与真实

诊断结果高度一致, 同时冗余诊断较少, 能够有效解决兼夹证型诊断问题。该模型可作为评估患病风险和治疗效益的量化工具, 为医生、患者和卫生行政人员的决策提供直观和理性的依据, 具有重要的卫生经济学意义。同时体现了多标记学习算法与中医诊疗思维的高度契合, 为挖掘“有诸内者, 必行诸外”和“治未病”的理论内涵提供了新思路。

本课题组在此基础上使用机器学习方法构建了用于KOA辅助诊断的小程序, 可以预测患有KOA的可能性。可通过<https://ihealthapp.jnu.edu.cn/ms/#/tool-ggjb>在线获取。

## 参 考 文 献

- [1] 徐颖鹏, 谢利民, 王文岳. 中药外治与西药治疗膝骨关节炎疗效及安全性Meta分析. 中国中药杂志, 2012, 37(19): 2977-2984
- [2] 李灿东, 辛基梁, 雷黄伟, 等. 中医健康管理与人工智能. 中华中医药杂志, 2019, 34(8): 3586-3588
- [3] JAMSHIDI A, PELLETIER J P, MARTEL-PELLETIER J. Machine-learning-based patient-specific prediction models for knee osteoarthritis. Nat Rev Rheumatol, 2019, 15(1): 49-60
- [4] XU H, ZHAO G, XIA F, et al. The diagnosis and treatment of knee osteoarthritis: A literature review. Int J Clin Exp Med, 2019, 12(5): 4589-4599
- [5] 於仁俊, 杜斌, 尤松鑫. 辨治慢性乙型肝炎经验. 中华中医药杂志, 2022, 37(2): 874-878
- [6] 沈绍功, 姚金城. 中医心病之心绞痛诊断与疗效标准. 中国医药学报, 2004, 19(10): 581-583
- [7] 辛基梁, 张佳, 李绍滋, 等. 中医健康状态辨识中的多标记分类方法研究. 中华中医药杂志, 2019, 34(9): 3952-3955
- [8] XIA S, ZHANG J, DU G, et al. A microcosmic syndrome differentiation model for metabolic syndrome with multilabel learning. Evid Based Complement Alternat Med, 2020, 2020: 9081641
- [9] YANG L, WU X Z, JIANG Y, et al. Multi-label learning with deep forest//24th European Conference on Artificial Intelligence, 2020: 1634-1641
- [10] 陈卫衡. 膝骨关节炎中医诊疗指南(2020年版). 中医正骨, 2020, 32(10): 1-14
- [11] KOCEV D, VENS C, STRUYF J, et al. Tree ensembles for predicting structured outputs. Pattern Recogn, 2013, 46(3): 817-833
- [12] GIBAJA E, VENTURA S. A tutorial on multilabel learning. ACM Computing Surveys(CSUR), 2015, 47(3): 1-38

(收稿日期: 2023年3月29日)