

中医证型分类模型的构建

姓名: 胡屹莹 时间: 2026.1.31

Abstract

本文主要针对筛选后的临床数据构建中医证型分类模型，

关键词：智能中医，机器学习，深度学习，

1 介绍 (Introduction)

1.1 背景 (Background)

1.2 任务 (Main Task)

针对 259 个病例样本，分别使用传统机器学习以及神经网络进行模型构建，首先是传统机器学习，使用随机森林和 KNN 算法建模，构建中医证型分类模型，并使用五折交叉验证；神经网络部分，鉴于数据量较小，考虑使用简单神经网络

2 方法 (Method)

2.1 随机森林和 KNN 模型的搭建

本研究的调参逻辑采用分层交叉验证框架，完成基线模型评估以识别潜在过拟合、欠拟合或类别不平衡问题后，为了进一步分析最重要的指标并减少数据集划分而产生的偏移，本研究设计了重复抽样验证策略：通过固定随机种子、执行 100 次独立的数据集分层划分，每次划分均重新计算并记录两种独立算法（随机森林与神经网络）导出的特征权重排序。在此基础上，构建累积积分评价体系：对每次划分结果中排名前 10 位的特征，按排名顺序分别赋予 10 至 1 分的递减权重（即排名第 1 得 10 分，第 10 名得 1 分），最终通过对 100 次划分结果的得分进行累加，遴选出累计得分最高的 10 个核心特征，作为最具判别力且稳定性强的关键生物标志物。

2.2 神经网络模型的构建

本研究基于中医临床数据构建了一个深度神经网络模型，用于中医证型的多分类任务。研究从数据预处理开始，首先加载包含症状与证型标签的 Excel 数据，进行类别分布分析和可视化；模型架构采用三层全连接神经网络（128-64-32 个神经元），并集成了 L2 正则化、批标准化和 Dropout 等强正则化组件以防止过拟合，输出层使用 Softmax 函数实

现多分类。训练过程采用五折分层交叉验证确保模型稳健性，通过 Adam 优化器（学习率 0.001）和类别权重平衡解决数据不平衡问题，并结合早停法、学习率衰减等回调函数优化训练效率。模型评估综合了准确率、平衡准确率、加权 F1 分数、Cohen’s Kappa 系数等多维度指标，并通过混淆矩阵和类别性能对比进行可视化分析。特征重要性分析采用梯度方法识别各证型的关键症状，为中医辨证提供可解释性依据。最终模型保存为可部署格式，并封装了预测函数接口，能够输入症状字典输出证型预测及概率分布。整个流程强调可复现性与临床实用性，既可作为中医辅助诊断工具，也为证候规律研究提供数据驱动的方法支持。

3 结果 (Results)

3.1 数据分析与预处理

在 259 个病例样本数据中，共有 78 个特征指标，其中面诊指标 24 个，舌诊指标 25 个，脉诊指标 29 个；分类指标为中医证型，其中包含 8 个类别，分别是气虚血瘀证，心肾阳虚证，痰浊闭阻证，心血瘀阻证，气滞血瘀证，气阴两虚证，心肾阴虚证，寒凝心脉证，8 种证型占比均在 12% 左右，数据数量分布比较均匀。

中医证型	样本数量	百分比 (%)
气虚血瘀证	35	13.5
心肾阳虚证	34	13.1
痰浊闭阻证	34	13.1
心血瘀阻证	33	12.7
气滞血瘀证	32	12.4
气阴两虚证	31	12.0
心肾阴虚证	30	11.6
寒凝心脉证	30	11.6

Table 1: 中医证型样本分布统计

本研究通过卡方检验对 8 种冠心病中医证型与 30 项临床特征进行关联性分析，发现其中 10 项特征具有极显著的统计学差异 ($p_{adj} < 0.001$)。其中，“五心烦热”与“形寒肢冷”的鉴别能力最为突出，卡方值分别为 174.44 和 154.71，Cramer’s V 值分别为 0.821 和 0.773，表明二者与证型之间存在强关联，可作为区分阴虚证型与阳虚证型的关键指标。此外，舌象特征如“腻苔” ($V = 0.511$)、“涩脉” ($V = 0.496$) 及“舌色红” ($V = 0.442$) 亦表现出中等程度的鉴别价值，提示舌诊在证型辨别中具有重要作用。症状方面，“瘀点瘀斑” ($V = 0.468$)、“乏力” ($V = 0.440$) 等特征虽效应量稍弱，但仍能提供一定的辨证参考。以上结果经 Benjamini-Hochberg 多重检验校正后仍保持显著，说明本研究的发现具有较好的统计可靠性。

排名	特征	卡方值	p 值	p_{adj}	Cramer's V
1	五心烦热	174.4441	2.90×10^{-34}	1.77×10^{-32}	0.8207
2	形寒肢冷	154.7093	4.16×10^{-30}	1.27×10^{-28}	0.7729
3	腻苔	67.5523	4.61×10^{-12}	9.37×10^{-11}	0.5107
4	涩	63.7917	2.63×10^{-11}	4.01×10^{-10}	0.4963
5	瘀点瘀斑	56.6626	6.98×10^{-10}	7.09×10^{-9}	0.4677
6	细	56.6827	6.91×10^{-10}	7.09×10^{-9}	0.4678
7	数	54.2367	2.11×10^{-9}	1.84×10^{-8}	0.4576
8	舌色红	50.7076	1.05×10^{-8}	8.00×10^{-8}	0.4425
9	乏力	50.0352	1.42×10^{-8}	9.64×10^{-8}	0.4395
10	紧	46.6683	6.48×10^{-8}	3.95×10^{-7}	0.4245

Table 2: 各特征在 8 种中医证型间的卡方检验结果（前 10 项）

3.2 随机森林与神经网络模型的结果

基于本研究构建的随机森林与神经网络交叉验证框架，对 259 例包含八种中医证候（寒凝心脉证、心肾阳虚证、心肾阴虚证、心血瘀阻证、气滞血瘀证、气虚血瘀证、气阴两虚证、痰浊闭阻证）的样本数据进行 100 次分层随机划分与特征重要性分析，结果显示各基线模型的加权 F1 分数分别为随机森林 0.4830 ± 0.0180 、决策树 0.3777 ± 0.0163 、神经网络 0.4244 ± 0.0405 、支持向量机 0.4269 ± 0.0568 、K 近邻 0.3852 ± 0.0428 ，整体分类性能中等且存在一定过拟合风险与维度灾难可能。

模型名称	加权 F1 分数（均值 \pm 标准差）	性能评价
随机森林	0.4830 ± 0.0180	最佳性能
支持向量机	0.4269 ± 0.0568	次优性能，波动较大
神经网络	0.4244 ± 0.0405	中等性能
K 近邻	0.3852 ± 0.0428	性能一般
决策树	0.3777 ± 0.0163	性能最低

Table 3: 五种模型的验证性能比较

经累计积分评价体系筛选，获得总得分排名前十的关键辨证指标依次为形寒肢冷（1961 分）、五心烦热（1740 分）、乏力（1395 分）、腻苔（868 分）、涩脉（755 分）、细脉（722 分）、心悸失眠（558 分）、气短懒言（427 分）、舌色暗红（388 分）及胸闷心慌（266 分）

排名	特征名称	总得分	算法贡献 (RF/ANN)
1	形寒肢冷	1961	1000/961
2	五心烦热	1740	839/901
3	乏力	1395	832/563
4	腻苔	868	658/210
5	涩	755	235/520
6	细	722	565/157
7	心悸失眠	558	418/140
8	气短懒言	427	367/60
9	舌色暗红	388	0/388
10	胸闷心慌	266	0/266

Table 4: 特征重要性累计得分排名（前 10 项）

其中前五项特征在两种算法中均呈现高度一致的判别重要性；进一步通过特征子集验证发现，仅使用形寒肢冷、五心烦热、乏力、腻苔、涩脉这五个核心指标即可达到最佳分类性能（加权 $F1=0.4870\pm0.0535$ ），而增加更多特征并未显著提升模型判别力，说明这五项临床表征构成了八种证候分类最具稳定性与代表性的微观指标集合，为中医辨证的客观化与标准化提供了可量化的特征依据。

特征数量	加权 F1 分数（均值 \pm 标准差）	特征组合
5	0.4870 ± 0.0535	形寒肢冷、五心烦热、乏力、腻苔、涩
10	0.4264 ± 0.0537	增加细、心悸失眠、气短懒言、舌色暗红、胸闷心慌
15	0.4682 ± 0.0443	-
20	0.4546 ± 0.0389	-
30	0.4569 ± 0.0525	-

Table 5: 不同规模特征的性能比较

4 讨论 (Discussion)

文献引用：正如 Hinton 在文 [1] [2] 中所述。。。注意，文献的引用需要新建 filename.bib 文件，将文献对应的 bitex 格式的引用复制粘贴到文件中然后再引用！谷歌学术或微软学术均可生成文献引用可用的 Bibtex 格式。

References

- [1] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

- [2] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. Show, attend and tell: Neural image caption generation with visual attention. In *International conference on machine learning*, pages 2048–2057, 2015.