

# 中医证型分类模型的构建

姓名: 胡屹莹      时间: 2026.1.29

## Abstract

本研究基于 259 个病例样本数据及提取的 78 个特征数据，包括舌诊（苔质，舌色等）/脉诊（平，浮，洪等）/问诊（呕吐痰涎，脘腹痞满）等，进行中医证型分类模型的构建。

关键词：中医辨证，机器学习，

## 1 介绍 (Introduction)

### 1.1 背景 (Background)

心肾阳虚证

### 1.2 任务 (Main Task)

在 259 个病例样本数据中，使用随机森林和 KNN 算法建模，并在五折交叉验证后，进行优化调参；

## 2 方法 (Method)

### 2.1 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，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 ( $p_{\text{adj}} < 0.001$ )。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	$p_{\text{adj}}$	Cramer’s $V$
1	五心烦热	174.4441	$2.90 \times 10^{-34}$	$1.77 \times 10^{-32}$	0.8207
2	形寒肢冷	154.7093	$4.16 \times 10^{-30}$	$1.27 \times 10^{-28}$	0.7729
3	腻苔	67.5523	$4.61 \times 10^{-12}$	$9.37 \times 10^{-11}$	0.5107
4	涩	63.7917	$2.63 \times 10^{-11}$	$4.01 \times 10^{-10}$	0.4963
5	瘀点瘀斑	56.6626	$6.98 \times 10^{-10}$	$7.09 \times 10^{-9}$	0.4677
6	细	56.6827	$6.91 \times 10^{-10}$	$7.09 \times 10^{-9}$	0.4678
7	数	54.2367	$2.11 \times 10^{-9}$	$1.84 \times 10^{-8}$	0.4576
8	舌色红	50.7076	$1.05 \times 10^{-8}$	$8.00 \times 10^{-8}$	0.4425
9	乏力	50.0352	$1.42 \times 10^{-8}$	$9.64 \times 10^{-8}$	0.4395
10	紧	46.6683	$6.48 \times 10^{-8}$	$3.95 \times 10^{-7}$	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）

## 2.2 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，通过基线评估确定模型潜力后，针对二值特征为主的医学数据特性，为随机森林设计参数随机搜索策略，为 KNN 设计网格搜索策略，并结合动态参数约束机制防止过拟合，最终基于交叉验证准确率选择最优参数组合。

## 2.3 神经网络模型的构建

## 3 结果 (Results)

## 4 讨论 (Discussion)

## 5 引用

### 5.1 图片引用

图片引用：如图 1 所示，图中。。。引用对应的 label，注意图片的路径即可。latex 会根据页面的篇幅自动调整图片的位置。如果不想浮动的图片或表格越过某些部分，在该部分前加 `\FloatBarrier` 表明该部分不想被越过。

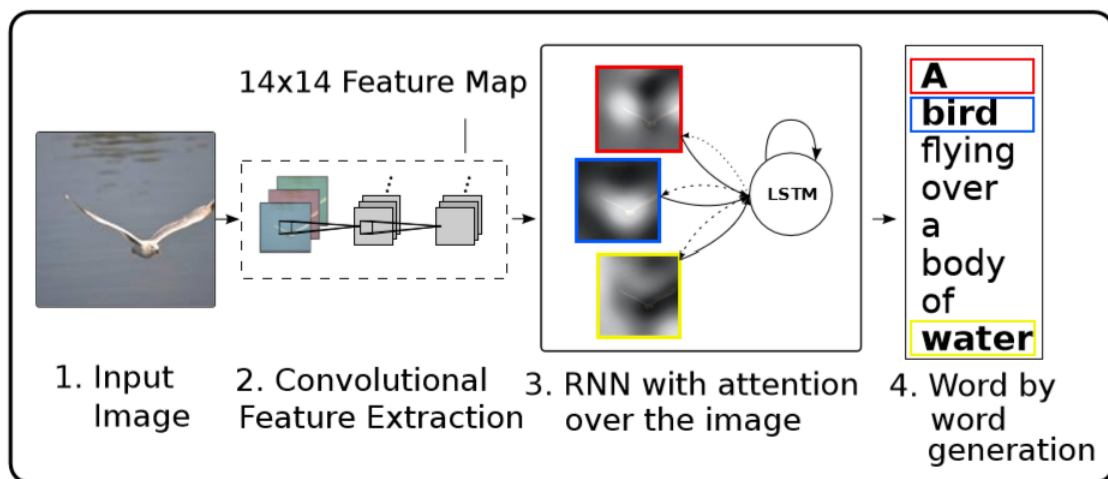


Figure 1: Encoder-decoder 结构

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。

## 6 其他

### 6.1 公式

公式等号对齐：

$$f_{att}(c_i, h_i) = v_a \tanh(W_a c_i, W_b h_i) \quad (1)$$

$$\alpha_i = f_{att}(c_i, h_i) \quad (2)$$

$$s = \text{softmax}(\alpha) \quad (3)$$

$$z = \sum s_i c_i \quad (4)$$

其中  $c_i$  表示。。。。

## 6.2 代码与伪代码

```
1  #include <iostream>
2  int main()
3  {
4      std::cout << "Hello, World!" << std::endl;
5  }
```

```
1  #include <iostream>
2  int main()
3  {
4      constexpr int MAX = 100;
5  }
```

---

### Algorithm 1 My algorithm

---

```
1: procedure MYPROCEDURE
2:    $stringlen \leftarrow \text{length of } string$ 
3:    $i \leftarrow patlen$ 
4: top:
5:   if  $i > stringlen$  then return false
6:   end if
7:    $j \leftarrow patlen$ 
8: loop:
9:   if  $string(i) = path(j)$  then
10:     $j \leftarrow j - 1$ .
11:     $i \leftarrow i - 1$ .
12:    goto loop.
13:    close;
14:   end if
15:    $i \leftarrow i + \max(\delta_1(string(i)), \delta_2(j))$ .
16:   goto top.
17: end procedure
```

---

## 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.