

中药学习与分析平台设计与实现^{*}

何佳怡 茆元慧 谢佳东 胡孔法 胡晨骏 杨 涛

(南京中医药大学人工智能与信息技术学院 南京 210023)

〔摘要〕 结合光学字符识别技术构建基于安卓的中药学习与分析平台, 详细阐述平台设计开发与具体实现, 指出该平台有助于中医药相关领域人员学习与研究中药理论与中药药性配伍规律。

〔关键词〕 Android; 中华本草; 文字识别; 学习平台

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2021.03.014

Design and Realization of Traditional Chinese Medicine Learning and Analysis Platform HE Jiayi, MAO Yuanhui, XIE Jiadong, HU Kongfa, HU Chenjun, YANG Tao, School of Artificial Intelligence and Information Technology, Nanjing University of Chinese Medicine, Nanjing 210023, China

〔Abstract〕 The Traditional Chinese Medicine (TCM) learning and analysis platform based on Android is built by combining Optical Character Recognition (OCR) technology. The paper elaborates on the design, development and implementation of the platform in detail, and points out that the platform is helpful for personnel in the field of TCM to learn and study the TCM theory and compatibility regularity of TCM property.

〔Keywords〕 Android; Chinese materia medica; character recognition; learning platform

1 引言

中药是中医用以防治疾病、养生康复与保健的重要载体^[1], 中药理论学习在中医药事业传承与发展中占据十分重要的地位, 中药药性理论是中医药理论体系核心理论^[2]。中药数量数以万计, 运用传统中药学习方法研究中药及其处方需要耗费大量时

间。互联网、无线移动网络技术与普及推动了智能手机等移动设备使用, 同时文字识别普遍应用, 方便人们快速录入信息, 提高学习与工作效率^[3], 移动学习应运而生, 具有便携性、智能性、良好的交互性等诸多优势^[4]。《中华本草》^[5]是体现 20 世纪中药学科发展水平的综合性本草著作, 本文以其中收录的中药为基础, 构建中药数据库, 结合光学字符识别技术 (Optical Character Recognition,

〔修回日期〕 2020-07-23

〔作者简介〕 何佳怡, 本科生, 发表论文 1 篇; 通讯作者: 谢佳东, 硕士, 助理实验师。

〔基金项目〕 国家自然科学基金委国家自然科学基金青年基金项目“基于中医临床大数据的名老中医诊治慢性肾病经验挖掘方法的应用研究”(项目编号: 81804219); 国家自然科学基金面上项目“面向中医临床大数据的现代名老中医肺癌辨治规律并行挖掘策略及方法学研究”(项目编号: 81503499); 国家重点研发计划项目“中医药数据中心与云平台共性关键技术研究”(项目编号: 2017YFC1703501); 国家重点研发计划项目“病案首页数据采集与中医医疗质量评价系统的创新研究”(项目编号: 2017YFC1703503); 南京中医药大学 2019 年大学生创新训练计划“基于朴素贝叶斯的中医肾病语音电子病历构建与应用研究”(项目编号: 2019NZY101)。

OCR), 识别中药处方图像中的文字信息, 设计开发基于 Android 的中药学习与分析平台, 从四气五味、升降浮沉、归经、毒性等方面分析每个处方以及每味中药的药性, 有利于中医药相关领域人员对中药理论与中药药性配伍规律学习与研究。中医药相关领域人员在从课本、期刊、网页等存储介质上阅读中医药理论、中医病历等文献资料时, 通过手机拍摄图像或者上传本地图像, 利用本平台可以快速查询单个中药的详细信息以及中药与处方分析结果, 有助于学习与研究中药理论与药性配伍规律。

2 平台设计

2.1 功能模块 (图 1)

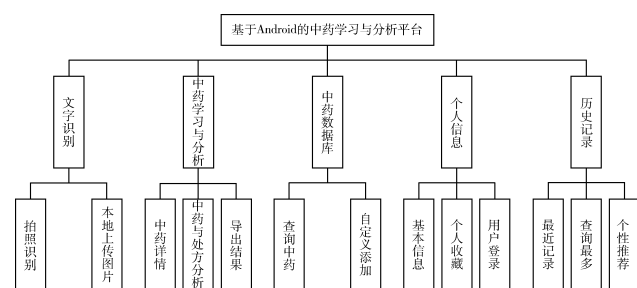


图 1 功能模块

2.1.1 文字识别 用户可选择拍照或本地上传图片进行识别, 平台自动识别出图像中有关中药文字信息, 这些中药均包含在中药数据库中, 不识别与中药无关内容。识别结果为单个或多个中药, 用户可根据个人需要进行中药学习与分析。

2.1.2 中药学习与分析 (1) 中药详情。根据文字识别结果, 用户可选择查看单个中药详细信息, 包括中药名称、拼音、英文名、别名、出处、药材基源、原形态、生态环境、栽培、性状鉴别、化学成分、药理作用、炮制、性味、归经、功能主治、用方用量、注意事项、各家论述等内容。(2) 中药与处方分析。根据文字识别结果, 用户可以选择单个中药或处方进行分析, 主要分析中药性、味、归经 3 个药性, 统计分析结果以图表形式呈现。(3) 导出结果。中药详情信息及中药与处方分析结果均可导出成 Word 文档、Excel 文档及图片,

便于用户学习与分析。

2.1.3 中药数据库 可查询单个中药, 包括中药名称、拼音、英文名、别名、出处、药材基源、原形态、生态环境、栽培、性状鉴别、化学成分、药理作用、炮制、性味、归经、功能主治、用方用量、注意事项、各家论述等内容。某些中药名称在各地用法不一, 无法在中药数据库中查找到, 可通过查询别名或自定义添加这类药属性, 添加的内容将存储至用户个人数据库中。

2.1.4 个人信息 (1) 基本信息。包含用户个人账号、密码、昵称、年龄、性别、手机号、邮箱地址等基本信息, 用户个人与平台管理员可进行管理。(2) 个人收藏。用户在使用平台进行中药学习与分析过程中可将中药详情信息、中药与处方分析结果添加到个人收藏模块, 便于快速查看、学习。(3) 用户登录。用户需使用平台账号与密码进行登录, 提供手机号与邮箱验证, 确保用户信息安全。

2.1.5 历史记录 主要保存用户使用文字识别功能的图片、最近查询的中药、查询次数最多的中药记录, 根据最近查询记录为用户推荐可能感兴趣的中药。

2.2 中药数据库

2.2.1 中药名称规范化 同一个中药往往具有正名、别名、异名、释名等多种不同名称, 在不同处方中可能出现不同名称^[6]。为解决一药多名问题, 本文基于中药数据库对中药名称进行梳理整合、规范统一, 实现正名、别名、异名、释名之间的自动转换, 不断补充、完善与更新中药数据库^[7]。

2.2.2 数据库设计 (1) 中药详情表。本文收集《中华本草》中收录的近万种中药, 内容涉及中药品种、栽培、药材、化学、药理、炮制、制剂、临床应用等中医药学科各个方面^[8], 在此基础上对每种中药属性进行详细划分, 设计中药详情表, 包含拼音、英文名、别名、出处、药材基源、原形态、生态环境、栽培、性状鉴别、化学成分、药理作用、炮制、性味、归经、功能主治、用方用量、注意事项、各家论述等属性, 利用此表可查询单个中药详细信息。(2) 中药药性、药味、归经表。

《中华本草》对中药药性进行更为细致的规范，部分中药性味、归经，见表 1。为准确分析中药与处方，设计中药药性表、中药药味表以及中药归经表。中药药性包括大寒、大热、寒、凉、平、热、寒、微温、温等，反映药物对人体阴阳盛衰、寒热变化的作用倾向，是与所治疾病的寒热性质相对而言的^[9]。中药药味分为淡、甘、苦、平、涩、酸、微甘、微苦、微酸、微辛、微咸、咸、辛等，重在反映中药部分作用的性质和特点，是对众多中药功效的进一步概括。归经主要包括肺经、肝经、胃经、肾经、脾经、心经、膀胱经、大肠经、胆经、三焦经、小肠经、心包经等脏腑经络。中药归经理论是在藏象学说和经络学说的基础上，明确药物作用定位和走向^[10]。

表 1 《中华本草》部分中药

中药	性	味	归经
黄芪	温	甘	心；肺；脾；肾
薏苡仁	微寒	甘；淡	脾；肺；肾
苍术	温	辛；苦	脾；胃；肝
白术	温	苦；甘	脾；胃
茯苓	平	甘；淡	心；脾；肺；肾

3 平台实现

3.1 文字识别技术

3.1.1 概述 文字识别作为模式识别中的一个重要领域，在日常生活中有广泛应用。OCR 技术可以对文字进行较高精度的识别，包括生僻字在内^[11]，是指智能手机等电子设备检查通过摄像头获取到的包含字符的图像，通过检测暗、亮、颜色、对比度、像素类别等指标、模式确定其特征，利用字符识别方法将其特征与模板进行匹配，从而翻译成计算机文字的过程^[12]。本文利用 OCR 技术识别从课本、期刊、网页等存储介质上拍摄或截取的正体字图像，通过字符特征数据库与中药数据库，获取图像中关于中药的文本信息，识别精度高、速度快，为中药学习与分析提供数据来源。文字识别实现过程主要分为图像获取、预处理、字符对比等^[13]，见图 2。

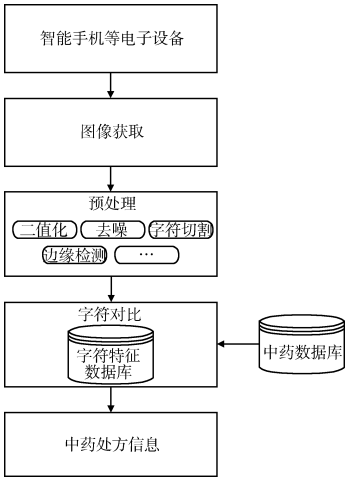


图 2 文字识别

3.1.2 图像获取 通过智能手机等电子设备获取需要识别的图像，包括课本、期刊、网页等存储介质上的正体字图像。

3.1.3 预处理 首先，对图像进行二值化处理，在保证不丢失图像特征的前提下减少图像占用空间，提高处理速度。其次，图像在拍摄、存储、传输等过程中可能会受到噪声干扰而导致质量降低，通过去噪处理获取高质量的图像信息。最后，利用图像分割与边缘检测将图像中的文字拆分成单个字符图像并取得每个字符图像特征。

3.1.4 字符对比 将预处理过程获取的字符图像特征与字符特征数据库中的标准字符进行对比，找出最相近的字符，得到通用字符。将通用字符与中药数据库进行对比，获取关于中药的字符。

3.2 中药与处方分析方法

首先，通过 ORC 文字识别获取图像中关于中药的信息，包括单个中药或处方。其次，基于《中华本草》构建的中药数据库分析每味中药药性，对分析结果进行校验，确保分析结果完整准确。文字识别结果保留图像中中药原始名称，可能存在与中药数据库中的中药名称不一致情况，导致分析结果不完整或不准确，例如“生白术”在中药数据库中名称为“白术”。在这种情况下需进行二次校验，通过正别名替换或用户自定义添加，补充、完善分析结果，同时完善、更新中药数据库，用户自定义添加的内容将

存储至用户个人数据库。最后，以图表形式呈现中药与处方标准分析结果，便于用户理解、分析与使用。中药与处方分析具体流程，见图 3。

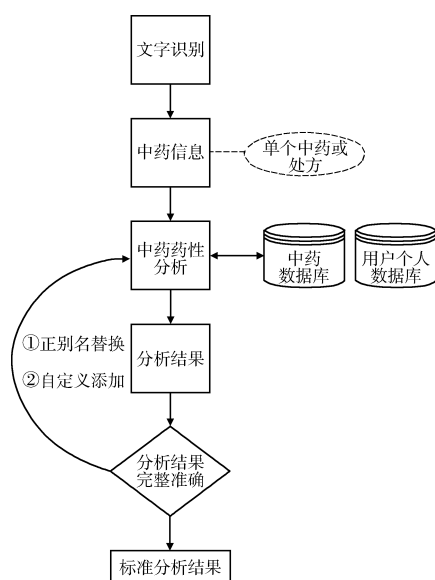


图 3 中药与处方分析流程

3.3 平台界面实现

中药学习与分析平台基于 Android 系统开发，采用模型视图控制器（Model View Controller, MVC）架构。用户进入平台主页面后可以选择文字识别、中药学习与分析、中药数据库、个人信息、历史记录功能。平台界面均建立在 Activity 之上，主界面（Main Activity）采用线性布局（Linear Layout）与相对布局（Relative Layout），添加导航栏（Navigation）、按钮（Button）等，使用 Intent 启动其他 Activity，跳转至文字识别界面（Recognition Activity）、中药学习与分析界面（Study Analyze Activity）、中药数据库界面（Database Activity）、个人信息界面（Personal Activity）、历史记录界面（History Activity）。

3.4 中药学习与分析平台实现

用户登录平台后进入文字识别界面，选择拍摄图像或本地上传图像进行文字识别，识别结果为多个中药，默认勾选全部中药，用户可根据个人需要将不进行分析的中药去除勾选，平台仅选择勾选的中药进行中药与处方分析。点击识别结果中的单个中

药，平台显示该中药详细信息。点击分析按钮，平台完成中药与处方分析后以图表形式呈现分析结果。

4 结语

本文设计开发基于 Android 的中药学习与分析平台，有助于中医药相关领域人员学习与分析中药及其处方。平台线上运行后，性能稳定，便捷性高，文字识别和中药与处方分析效果良好、速度快，提高移动学习者学习兴趣和效果，具有良好应用前景。

参考文献

- 1 王重阳, 刘越, 穆佳玲, 等. 基于移动互联网的中药学习平台研究 [J]. 医学信息学杂志, 2019, 40 (8): 90-94, 76.
- 2 李盼, 陈雨微, 丁丽琴, 等. 辛味中药在治疗消渴证中的应用价值及中药五味理论现代研究的思考 [J]. 中草药, 2019, 50 (22): 5577-5583.
- 3 鲍美英, 申晋祥. 基于 Android 的医疗考试培训系统的设计与实现 [J]. 山西大同大学学报 (自然科学版), 2018, 34 (5): 26-28.
- 4 马玉慧, 赵乐, 李南南, 等. 新型移动学习资源——教育 APP 发展模式探究 [J]. 中国电化教育, 2016 (4): 64-70.
- 5 国家中医药管理局《中华本草》编委会. 中华本草 [M]. 上海: 上海科学技术出版社, 1999.
- 6 谢佳东, 房裴裴, 胡孔法, 等. 现代名老中医临床诊疗方药数据抽取与清理方法研究 [J]. 时珍国医国药, 2017, 28 (11): 2786-2788.
- 7 王连心, 孟庆刚, 王志国, 等. 中药知识库设计浅析 [J]. 世界中医药, 2011, 6 (6): 535-537.
- 8 虞舜, 钱锦栋. “中华本草数据库”的研制 [J]. 江苏中医药, 2006, 27 (12): 62-63.
- 9 李久全. 中药气味药性理论困惑与对策研究 [D]. 哈尔滨: 黑龙江中医药大学, 2016.
- 10 王俊尧. 基于网络药理学研究中药归经与治疗疾病的关系 [D]. 北京: 北京中医药大学, 2018.
- 11 张婷婷, 马明栋, 王得玉. OCR 文字识别技术的研究 [J]. 计算机技术与发展, 2020 (3): 1-8.
- 12 李梵若, 韩莹, 代广斌, 等. 文字识别在虚拟导游系统中的应用研究与实现 [J]. 计算机与现代化, 2019 (12): 83-87.
- 13 程加乐. 基于特征空间的旋转多字体文字识别 [D]. 西安: 长安大学, 2016.