

基于脉搏波的中医体质自动辨识系统研究初探^{*}

张 毅,王宏宇,吕朝磊,张 倩,贾聪敏,王 耘^{**}

(北京中医药大学中药信息工程研究中心 北京 100102)

摘 要:目的:探索光电容积脉搏波频域分析进行中医体质辨识的可行性,建立简单可靠的气虚体质判定模型。方法:对招募到的志愿者应用《中医体质分类和判定标准》体质量表评估体质状态;通过传感器采集脉搏波信号,并进行功率谱分析提取频谱特征信息;利用数据挖掘8种分类算法建立气虚体质判定模型。结果:脉搏波谐波峰值:气虚组>非气虚组;随机森林算法时模型最优,10折交叉验证测试的准确性达到80.4%,受试者工作特征曲线下面积为0.824,模型具有一定的可靠性。结论:应用光电容积脉搏波频域分析进行气虚体质辨识的方法是可行的,可为构建中医体质自动辨识系统提供前期基础,为亚健康人群在家庭保健中进行体质辨识提供简便快速的辅助工具。

关键词:气虚体质 体质辨识 光电容积脉搏波 频域分析 数据挖掘

doi:10.11842/wst.2017.10.017 中图分类号:R24 文献标识码:A

中医体质学说体现了中医因人制宜、辨证论治等核心理念。把握不同人群体质的特征和差异,及由此所决定的发病倾向和对某些病因的易感性,对制定个性化的治疗、预防、养生方法具有重要意义。因此,需要将个体复杂的体质予以辨识。体质辨识是制定个性化预防保健措施的依据及基础,是摆脱亚健康、实现“治未病”的前提^[1,2]。但目前的体质辨识方法还有一定的不足:①中医资源稀缺,不能满足亚健康人群体质辨识的需求;②通过《中医体质分类和判定标准》体质量表进行自测,易受测试者主观因素影响^[3];③基因分类器、经络知热感度测量法^[4]、中医四诊等几种辨识方法,成本及技术要求较高,不利于普及应用等。因此,建立操作简便、能广而用之的体质自动辨识系统辅助亚健康人群进行体质辨识具有重要意义。

光电容积脉搏波描记法(Photoplethysmography, PPG)是借光电手段在活体组织中检测血液容量变化

的一种无创检测方法^[5],可以得到连续的光电容积脉搏波。脉搏波蕴含机体丰富的生理病理信息,对其波形信号变化的研究可以实现对心血管、呼吸、心理等^[6,7]进行无创连续检测,临床上已广泛应用于心率及血氧饱和度的监测^[8]。因其检测方法操作简便和成本低廉等特点在家庭保健中显示出特有的优势。

本研究将光电容积脉搏波频域分析应用于中医体质辨识,建立中医体质辨识模型,并进一步构建中医体质自动辨识系统,以实现在家庭保健中对体质的辅助辨识。气虚体质是最常见、最基本的非生理性体质类型之一^[9],张筠等^[10]研究表明气虚体质是大学新生中最常见的体质类型,陈润东等^[11]讨论了广州地区中医体质类型的分布状况,结果表明气虚质比例最高。因此本研究以气虚体质为例,对不同气虚体质脉搏波频谱之间的差异进行分析,阐明基于脉搏波频域分析建立气虚体质判定模型实现气虚体质自动辨识的可行性,为构建中医体质自动辨识系统提供基础。

收稿日期:2017-09-14

修回日期:2017-10-24

^{*} 国家自然科学基金项目(No. 81373985):药性组合三联密码子的作用规律及其应用研究,负责人:王耘。

^{**} 通讯作者:王耘,教授,博士生导师,主要研究方向:中药信息融合。

1 材料与方法

1.1 研究方法

测试对象主要为北京中医药大学的大学生及研究生。对志愿者分别进行体质辨识及脉搏波的数据采集。志愿者用《中医体质分类和判定》标准下体质量表以自我报告的形式进行体质辨识。根据疾病的临床诊断标准排除各种急慢性疾病的患者,并且排除调查期间服用任何药物、脉搏波信息缺失、调查问卷质量不佳的受试者,最终纳入研究对象112例(非气虚组与气虚组各56例),其中男性40例,女性72例,年龄18~30岁,平均年龄 19.41 ± 2.4 岁。

1.2 体质调查问卷

调查问卷根据中华中医药学会《中医体质分类和判定标准》^[12]体质量表制定,并包括调查对象的性别、年龄、身高、体重、既往病史等内容。王琦教授将中医体质分为平和质、气虚质、阳虚质、阴虚质、痰湿质、湿热质、瘀血质、气郁质、特禀质等9种基本类型。体质量表根据不同体质共有9个亚量表,分别计算原始分及转化分,依标准判定体质类型。

原始分=各个条目分值相加

转化分数=[(原始分-条目数)/(条目数 $\times 4$)] $\times 100$

气虚体质亚量表共有8个条目,每一问题按5级评分(从1(没有发生)至5(总是发生)),总分从8到40(本研究中总分从8到35),转化分从0到100(本研究中转化分从0到85)。转化分 ≥ 40 判定为气虚体质,转化分在30~39分则判定为倾向气虚体质,转化分 < 30 判定为非气虚体质。

1.3 光电容积脉搏波的信息提取

1.3.1 数据采集

利用光电反射式脉搏波传感器与Arduino单片机搭建脉搏波检测设备,检测实验在温度适宜的安静室内进行,检测开始前要求受试者心情平静,脉搏波检测过程中受试者保持静坐。将传感器用绑带固定在受试者右手食指处,绑带的松紧程度保持适中,在保证传感

器不脱离的情况下避免施加压力过大导致脉搏波的波形形变和采集失真。检测过程持续约1 min,并需要获得至少20 s的稳定波形数据,采样率为50 Hz,存储单元为5 bit。图1为采集到的PPG信号。

1.3.2 脉搏波的频域分析

脉搏波频谱由脉搏波所具有的全部频率分量的集合形式表示,保留了脉搏波的全部信息,可以获得时域分析中无法得到的特征信息。

对脉搏波信号进行Welch功率谱分析,然后经过hamming窗平滑,得到信号的功率谱如图2所示。发现主要功率集中在0~11 Hz之间,因此只对这一范围的功率谱进行研究。

A点在功率谱中幅值最大(基波),频率约为1.17 Hz,对应测试者的脉率;B点和C点对应的频率分别是A点频率的整数倍(2倍和3倍),因此可以认为是基波的谐波。研究发现,一般人的脉搏波成分主要都集中在基波(第1次谐波)到5~6次谐波内^[13]。因为受采样率的限制,本研究只取到脉搏波功率谱的第6个谐波。计算得到各脉搏波功率谱的6谐波峰值,利用SAS软件对数据进行统计分析。

1.4 体质判定模型构建方法

提取各脉搏波功率谱0~11 Hz的相位及幅值,利用8种常用数据挖掘分类算法(贝叶斯网络算法、支持向量机算法、K最近邻算法、KStar算法、决策树算法、随机森林算法、AdaBoost算法和Bagging算法)挖掘气虚体质与脉搏波频谱之间可能存在的关系,建立气虚体质判定模型。以上所有分类算法通过数据挖掘平台Weka实现。

本研究所有分类算法均在10折交叉验证的环境下运行。交叉验证是数据建模中常用且有效的一种内部验证方法,常用来测试模型或算法的准确性。

模型预测性能的评价指标主要包括准确性(ACC)、灵敏度(SE)、特异性(SP)、平衡精度(BACC),受试者工作特征曲线曲线下面积(AUC)。灵敏度和特异性通常分别被用来评估模型正确识别真阳性样本和真阴性

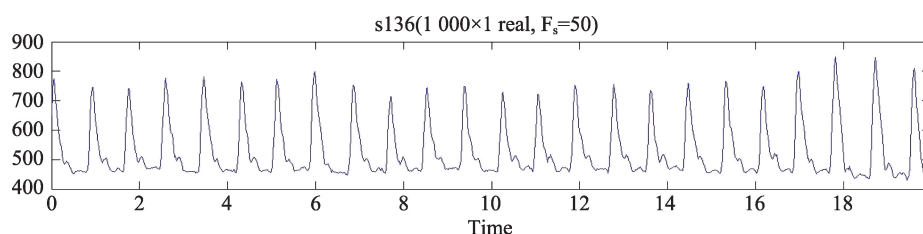


图1 光电容积脉搏波描记法信号

样本的能力。在本研究中,灵敏度指实际为气虚体质,按照本研究所建模型被正确地判断为气虚体质的概率;特异性指实际为非气虚体质,并被模型正确地判断为非气虚体质的概率。平衡精度为灵敏度和特异性的平均值,是模型训练过程中评价模型性能的重要指标^[14, 15]。ROC(receiver operator characteristic)称为受试者工作特征曲线,又称为接收者操作特征曲线^[16, 17]。ROC曲线下面积(AUC)介于0.5~1.0之间。AUC越接近于1,说明模型越可靠。AUC介于0.5~0.7之间时,表明模型具有较低的可靠性。AUC介于0.7~0.9之间时,表明模型具有一定的可靠性。AUC在0.9以上时,表明模型据有较高的可靠性。AUC=0.5时,说明模型完全不起作用,无诊断价值。为了对该模型的预测性能作出合理而全面的评估,本研究将综合考虑以上几项指标。

2 结果与讨论

2.1 体质调查问卷

所有参与人员均完成了体质调查问卷的填写,气虚体质亚量表转化分平均值为 29.24 ± 16.19 。根据体质判定标准筛选出非气虚体质与气虚体质各56例。研究参与人员均为青年群体,平均年龄 19.41 ± 2.4 岁,以避免兼夹体质过多及疾病的干扰,如衰老和慢性炎症。BMI均在正常范围内。

2.2 非气虚与气虚的脉搏波谐波峰值相关性分析

脉搏波频域丰富的谐波幅度与频率对应着不同的时域形态,包含了人体健康的绝大多数信息,中医理论中的脉象实际上在频域上的反应就是谐波幅度的不同^[18]。

经过计算,分别得到非气虚组及气虚组的6个脉搏波谐波的峰值($p_1, p_2, \dots, p_6$),分析结果如下所示。从表2和图3分析可以看出:(1)不同体质的脉搏波谐波峰值均随频率的增加而逐渐下降。(2)气虚体质的谐波峰值均明显大于非气虚体质,其中除 p_6 外均具有显著差异($P < 0.05$)。

从比较结果可以得出脉搏波谐波峰值与不同气虚体质之间存在一定关系,谐波峰值:气虚体质>非气虚体质,即气虚体质的谐波峰值幅度要大于非气虚体质。仅提取谐波峰值,信息较少,因此提取各脉搏波功率谱0~11 Hz的相位及幅值来进一步分析。

2.3 气虚体质判定模型与验证

本研究选取AUC值在0.7以上,ACC值最大的模型为最优模型。综合比较8种分类算法,分类方法为

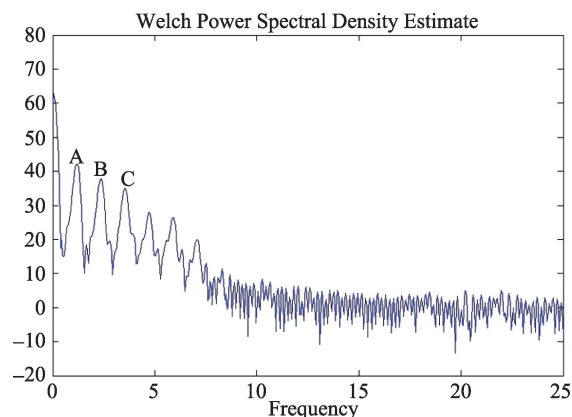


图2 脉搏波的功率谱

表1 调查问卷结果($\bar{x} \pm s$)

	男	女	总计
人数	40	72	112
年龄	18.68 ± 1.23	19.82 ± 2.78	19.41 ± 2.4
BMI	20.89 ± 3.25	20.35 ± 2.14	20.54 ± 2.59
气虚亚量表转化分	27.34 ± 14.89	30.3 ± 16.88	29.24 ± 16.19

表2 非气虚组和气虚组脉搏波谐波峰值比较($\bar{x} \pm s, n=56$)

	非气虚组	气虚组
$p_1(\text{dB})$	41.25 ± 5.88	$43.76 \pm 4.5^*$
$p_2(\text{dB})$	37.43 ± 6.32	$39.98 \pm 4.48^*$
$p_3(\text{dB})$	33.52 ± 6.16	$35.61 \pm 4.54^*$
$p_4(\text{dB})$	25.8 ± 6.18	$28.66 \pm 5.63^*$
$p_5(\text{dB})$	22.35 ± 6.75	$24.8 \pm 5.17^*$
$p_6(\text{dB})$	17.44 ± 7.22	19.92 ± 5.39

注:与非气虚比较,* $P < 0.05$

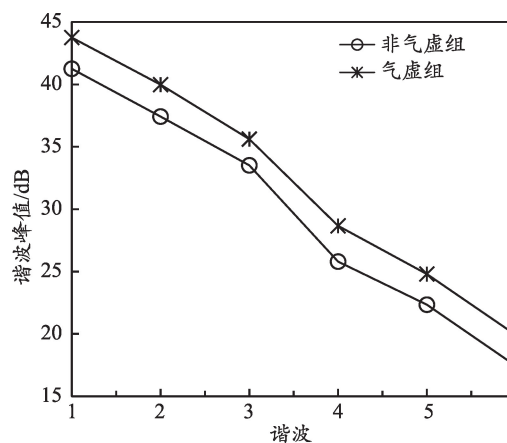


图3 非气虚组和气虚组脉搏波谐波峰值比较

随机森林算法时模型最优,相关指标如表3所示。模型10折交叉验证测试的准确性达到80.4%,识别真阳

表3 最优模型评价相关指标

算法	ACC	SE	SP	BACC	AUC
随机森林	0.804	0.732	0.875	0.804	0.824

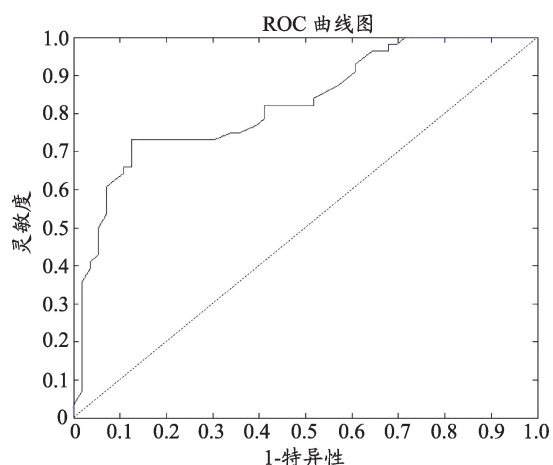


图4 最优模型受试者工作特征曲线

性样本的正确率即灵敏度达到73.2%,特异性、平衡精度及受试者工作特征曲线下面积均在0.8以上,表明本研究所建模型具有一定的可靠性。

对所建立的模型进行了ROC曲线分析。ROC曲线将灵敏度与特异性以图示化的方法结合在一起,从而实现模型的无偏估计,是试验准确性的综合代表。通常情况下,一次随机的预测返回的点将会落在图中虚线的对角线上,而理想的预测返回的点将会落在图中左上角的区域,越靠近左上角说明模型的预测性能越理想。图4中ROC曲线整体均呈现向左上角区域靠拢的趋势,表明本研究建立的气虚体质判定模型具有一定的可靠性。

本研究所建模型目前只适用于年龄18至30岁的青年人群。而大学生中亚健康状态普遍存在^[19],气虚体质更是大学新生中最常见的体质类型^[10]。气虚体质判定模型方便了青年人群如大学生群体对不同气虚体质的辅助辨识,为制定健康干预计划提供一定的依据,对改善亚健康状态具有重要作用。

3 结论

本研究仅是建立基于光电容积脉搏波频域分析的中医体质自动辨识系统的初次尝试。对112名参与者的不同气虚体质脉搏波频谱之间的差异进行了分析,建立了气虚体质判定模型,实现了利用脉搏波的频谱分析对不同气虚体质的自动辨识。表明利用光电容积脉搏波的频域分析实现气虚体质辨识的方法是可行的,为下一步建立中医体质自动辨识系统提供一定的基础。将光电容积脉搏波频域分析引入中医体质辨识是一种新的探索,笔者尚未见相关报道。

在今后的工作中,笔者将在气虚体质判定模型的基础上完善九种体质判定模型并进一步扩大样本量,建立网上平台构建中医体质自动辨识系统,辅助用户对中医体质的辨识。光电容积脉搏波采集设备价格低廉,操作简便,无需复杂的装置,非常适合非医疗技术人员甚至用户自己操作,有利于在家庭中推广使用。有研究表明脉搏波信号的最高频率成分一般不超过20 Hz,在实际分析中采样频率取50–100 Hz即可满足要求^[13]。本研究采样率为50 Hz,保证了脉搏波信号的质量,同时对脉搏波传感器的要求不高,降低成本,方便脉搏波检测设备的开发。

参考文献

- 王琦. 体质辨识是“治未病”的“抓手”. 环球中医药, 2008, 03: 5–6.
- 孙理军, 柏云飞. 体质辨识在“治未病”中的地位及作用. 现代中医药, 2015, 06: 58–61.
- 柳璇, 王琦. 《中医体质分类与判定》标准修改建议及分析. 北京中医药大学学报, 2013, 36(5): 300–304.
- 冼益民. 运用经络知热感度测量法探讨气虚质人群的经络体质特性. 广州: 广州中医药大学硕士学位论文, 2011.
- Allen J. Photoplethysmography and its application in clinical physiological measurement. *Physiol Meas*, 2007, 28(3): R1–39.
- Sugo Y, Ukawa T, Takeda S, *et al.* A Novel Continuous Cardiac Output Monitor Based on Pulse Wave Transit Time. *IEEE engineering in medicine and biology society*, 2010, 1: 2853–2856.
- Mcgrath S P, Ryan K L, Wendelken S M, *et al.* Pulse Oximeter Pethysmographic Waveform Changes in Awake, Spontaneously Breathing. *Journal of the American Medical Informatics Association. Anesth Analg*, 2011, 112(2): 368–374.
- Cennini G, Arguel J, Akşit K, *et al.* Heart rate monitoring via remote photoplethysmography with motion artifacts reduction. *Optics Express*, 2010, 18(5): 4867–75.
- 王琦, 朱燕波. 中国一般人群中中医体质流行病学调查——基于全国9省市21948例流行病学调查数据. 中华中医药杂志, 2009(1): 7–12.
- 张筠, 陈志军, 刘磊, 等. 2296例大学新生气虚体质状况调查. 中国中医基础医学杂志, 2015, 21(3): 310–311.
- 陈润东, 杨志敏, 林熾钊, 等. 中医体质分型6525例调查分析. 南京

- 中医药大学学报, 2009, 25(02): 104–106.
- 12 中华中医药学会. 中医体质分类与判定(ZYYXH/T157–2009). 世界中西医结合杂志, 2009, 4(4): 303
 - 13 罗志昌, 张松, 杨益民. 脉搏波的工程分析与临床应用. 北京: 科学出版社, 2006.
 - 14 Baldi P, Brunak S, Chauvin Y, *et al.* Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics*, 2000, 16(5): 412–424.
 - 15 Peng J, Lu J, Shen Q, *et al.* In silico site of metabolism prediction for human UGT-catalyzed reactions. *Bioinformatics*, 2014, 30(3): 398–405.
 - 16 Liu H, Wu T. Estimating the Area under a Receiver Operating Characteristic Curve For Repeated Measures Design. *J Stat Softw*, 2003, 08(i12): 1–18.
 - 17 Hanley J A, Mcneil B J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 1982, 143(1): 29–36.
 - 18 孟兆辉, 白净, 王苏中, 等. 高血压病人的光电容积脉搏波的频域分析. 北京生物医学工程, 2002, 21(1): 1–4.
 - 19 李明今, 李丹. 基于中医体质理论的女大学生亚健康状态现状调查. 延边大学医学学报, 2012, 35(2): 126–128.

Research on Automatic Identification System of Traditional Chinese Medicine Constitution Based on Pulse Wave

Zhang Yi, Wang Hongyu, Lv Zhaolei, Zhang Qian, Jia Congmin, Wang Yun

(Information Engineering Research Center for Traditional Chinese Medicine, Beijing University
of Chinese Medicine, Beijing 100102, China)

Abstract: This study was aimed to explore the feasibility of identification of traditional Chinese medicine (TCM) constitution based on photoplethysmography (PPG) frequency domain analysis, and to establish an automatic identification system (AIS) of TCM constitution with simple operation and good generalization. The *Classification and Determination of Constitution in TCM* was used to evaluate the subjects' constitutional state. Pulse wave sensor was used to collect the pulse wave signal. Spectral characteristics were analyzed by power spectrum analysis. Eight classification algorithms of data mining were used to establish the model of *qi*-insufficiency constitution. The results showed that the pulse wave harmonic peak was that *qi*-insufficiency constitution was higher than health group; the model was optimal when random forest algorithm was used; and the accuracy of 10–CV test was 80.4%. Sensitivity reached 73.2%. Specificity and receiver operator characteristic were all above 0.8. The model has certain reliability. It was concluded that it was feasible to identify *qi*-insufficiency constitution by means of frequency domain analysis of PPG. It provides preliminary work for the construction of AIS of TCM constitution as well as a simple and quick assistant tool for the constitution identification of subhealth group in family health care.

Keywords: *Qi*-insufficiency constitution, constitution identification, photoplethysmography, frequency domain analysis, data mining

(责任编辑:张 静, 责任译审:王 晶)