

文章编号: 0253-2239(2009)02-0459-05

基于径向基神经网络的测定抗结核药物 主成分质量分数的近红外光谱定量分析模型

逯家辉 王 迪 沈 畏 郭伟良 张益波 滕利荣*

(吉林大学生命科学学院, 吉林 长春 130012)

摘要 建立了同时测定利福平片、异烟肼片、吡嗪酰胺片、异福片和异福酰胺片 5 种抗结核药物中的利福平(RMP)、异烟肼(INH)和吡嗪酰胺(PZA)含量的新方法,应用径向基神经网络(RBFNN)建立 5 种抗结核片剂药物样品的近红外光谱(NIRS)与其中 RMP、INH 和 PZA 含量间相关模型。模型以交互验证均方根误差(RMSECV)为评价标准,选择最有效光谱区域、对网络结构参数和扩展常数进行优化,得到最优定量分析模型。最优模型的 RMSECV 分别为 0.0127、0.0104、0.0078,应用最优模型对预测集样品中 RMP、INH 和 PZA 含量进行预测,预测均方根误差(RMSEP)分别为 0.0125、0.0109、0.0103。内部交互验证和外部验证均表明,该方法具有较高的准确度,能够满足 5 种抗结核药物生产中 RMP、INH 和 PZA 的同时检测精度要求。

关键词 医用光学与生物技术;抗结核药物;径向基神经网络;近红外光谱;定量分析

中图分类号 O657.33; R927.2

文献标识码 A

doi: 10.3788/AOS20092902.0459

The Radial Basis Function Neural Network Quantitative Analysis Model for Determination of Anti-tuberculosis Tablets Using Near Infrared Spectroscopy

Lu Jiahui Wang Di Shen Wei Guo Weiliang Zhang Yibo Teng Lirong

(College of Life Science, Jilin University, Changchun, Jilin 130012, China)

Abstract A new approach is established to determine the contents of rifampicin (RMP), isoniazide (INH) and pyrazinamide (PZA) in five varieties of anti-tuberculosis tablets, i. e. rifampicin tablets, isoniazide tablets, pyrazinamide tablets, rifampicin and isoniazide tablets and rifampicin isoniazide and pyrazinamide tablets. Near infrared spectroscopy (NIRS) with radial basis function neural network (RBFNN) is applied to establishing the analytical protocol for the determination of the contents of RMP, INH and PZA in the five anti-tuberculosis tablets. The root mean square error of calibration set obtained by cross-validation (RMSECV) is used as the evaluation of the model. The optimal quantitative analysis models are resulted from selecting the most effective spectral region, suitable topological parameter and spread constants. RMSECV of the optimum models are 0.0127, 0.0104 and 0.0078. Using these optimum models for determination of the RMP, INH and PZA contents in prediction set, the root mean square error of prediction set (RMSEP) are 0.0125, 0.0109 and 0.0103. Internal cross-validation and external validation are verified and the method has a high accuracy and can meet the accuracy in simultaneous determination of RMP, INH and PZA contents in 5 varieties of anti-tuberculosis tablets.

Key words medical optics and biochemistry; anti-tuberculosis tablets; radial basis function neural network; near infrared spectroscopy; quantitative analysis

收稿日期: 2008-06-20; 收到修改稿日期: 2008-08-15

基金项目: 中国医学基金会新药发展基金(20061108)资助课题。

作者简介: 逯家辉(1965—),男,高级工程师,主要从事生物药物的非破坏在线质量控制方面的研究与应用。

E-mail: lujh@jlu.edu.cn

* 通信联系人。E-mail: tenglr543@gmail.com

1 引 言

结核病是由结核分支杆菌引起的慢性传染性疾病。近年来,结核病疫情在全球呈回升趋势,已经成为世界性的公共卫生问题。利福平(RMP)、异烟肼(INH)和吡嗪酰胺(PZA)是目前抗结核比较有效的药物。因单一用药,耐药性产生较快,异福片(RMP和INH复方制剂)和异福酰胺(RMP、INH和PZA复方制剂)可显著提高药物的抗菌活性,延缓耐药性的产生,更有利于患者的用药依从性。RMP、INH和PZA的常规测定方法主要有反相高效色谱法^[1],二极管阵列光度法^[2],薄层层析法^[3]和紫外分光光度法^[4],这些方法均需要对样品进行预处理,操作过程繁琐费时,且需要有机试剂。

现代近红外光谱(NIR)分析技术,可充分利用多波长下的光谱数据进行定性或定量分析。由于近红外光谱分析技术具有速度快、效率高、成本低、测试重现性好等优点,已经被越来越多地应用于食品工业、石油化工、制药工业等领域^[5~7]。此外,近红外光谱分析技术还具有如下特点:样品测量一般无需预处理,光谱测量方便。通过相应的测样器件可以直接测量液体、固体、半固体和胶状类等不同物态的样品^[8]。便于实现在线分析^[8]。倪力军等通过近红外光谱对丹参水提过程质量控制实现了在线检测^[9],是典型的无损分析技术。鉴于这一特点,该技术在活体分析和医药临床领域正得到越来越多的应用^[8]。

前向型神经网络具有很强的函数逼近功能,已被广泛应用于许多领域^[10]。径向基(RBF)神经网络

主要特点,一是可以在训练的参数优化过程中自动确定隐层神经元个数,二是训练速度快(学习比通常的(BP)算法快1000~10000倍),三是有良好的逼近和分类能力。理论上讲,RBF神经网络(RBFNN)具有任意函数的逼近能力^[11]。

本文应用主成分分析(PCA)和RBFNN相结合的方法,实现了5种抗结核药物中RMP、INH和PZA含量的近红外光谱法同时测定,结果令人满意。

2 材料与方 法

2.1 实验材料与仪器

利福平(99.59%)、异烟肼(99.87%)和吡嗪酰胺(99.66%)对照品购于中国药品生物制品检定所;辅料购买于成都市泰山薄膜包衣有限公司。

UV-3150型紫外可见近红外分光光度计(日本岛津公司),附件ISR-3100积分球;Waters486高效液相色谱仪(美国Waters公司),Nucleosil ODS(4.6 mm×150 mm, 10 μm)色谱柱(大连江申分离科技公司)。

2.2 样品的制备

分别按照利福平片、异烟肼片、吡嗪酰胺片、异福片和异福酰胺片处方量,精密称取各主药适量,按处方比例加入辅料,混合均匀,制成片剂样品,样品的RMP、INH和PZA含量采用高效液相色谱法(HPLC)^[3]测定。样品的各组分HPLC测定含量统计结果如表1。

表1 RMP, INH and PZA 质量分数统计

Table 1 Statistical mass fraction of RMP, INH and PZA

Samples	Components	Samples numbers	Average	Ranges
Rifampicin tablets	RMP	51	0.3749	0.3374~0.4119
Isoniazide tablets	INH	44	0.9004	0.8052~0.9947
Pyrazinamide tablets	PZA	35	0.8950	0.8043~0.9840
Rifampicin and isoniazide tablets	RMP	48	0.5233	0.4246~0.6193
	INH		0.2522	0.1999~0.2975
Rifampicin Isoniazide and Pyrazinamide tablets	RMP	40	0.1987	0.1501~0.2474
	INH		0.1390	0.1001~0.1779
	PZA		0.4085	0.3599~0.4572

2.3 光谱扫描

光谱带宽为12 nm,光谱扫描波长范围为800~2500 nm,每个样品扫描3次,取其平均值作为该样品的NIR光谱。随机选取的利福平片、异烟肼片、吡嗪酰胺片、异福片和异福酰胺片样品的NIR光谱

如图1所示。由图1可见,5种抗结核药物样品的NIR光谱严重重叠,用常规光谱分析方法难以进行定量分析,必须应用化学计量学方法进行光谱解析,实现样品浓度预测。实验中RBFNN程序利用Matlab7.0软件编写。

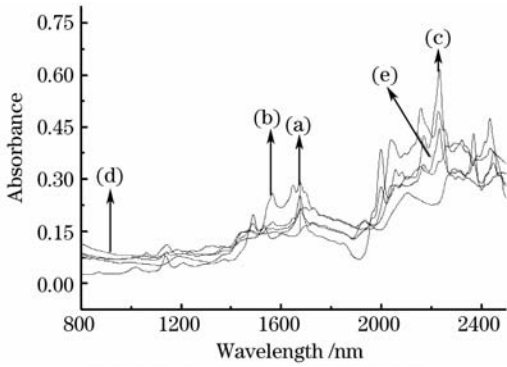


图1 样品的近红外光谱图

Fig. 1 NIR spectra of samples. (a): rifampicin tablet; (b): isoniazide tablet; (c): pyrazinamide tablet; (d): rifampicin and isoniazide tablet; (e): rifampicin isoniazide and pyrazinamide tablets

2.4 建模标准

建立 RBFNN 定量分析模型时, 218 个样品随机分成 2 组: 校正集包含 163 个样品, 预测集包含 55 个样品。采用 TQ Analyzer V6.0 分析软件自动选取最佳的光谱波长范围, 选择最合适的输入层节点数, 隐层节点数和最佳扩展常数, 建立最优的校正模型。以交互验证均方根误差 (RMSECV) 和预测均方根误差 (RMSEP) 作为指标, 来考察所建立校正模型的性能。RMSECV 和 RMSEP 计算公式如下:

$$E_{\text{RMS}} = \sqrt{\frac{\sum_{i=1}^n (C_{\text{NIR}_i} - C_{\text{REF}_i})^2}{n}}$$

对于 RMSECV, C_{NIR_i} 为校正集中各组分质量分数的交互验证预测值, C_{REF_i} 为校正集中各组分质量分数的 HPLC 测定值, 对于 RMSEP, C_{NIR_i} 为模型对预测集样品中各组分含量预测值, C_{REF_i} 为预测集中各组分含量的 HPLC 测定值。

3 结果与讨论

3.1 光谱波长范围的选择

光谱波长范围对建立模型的性能和预测准确性有很大的影响, 光谱波长范围的选择过程实际就是样品各组分有效光谱信息的筛选过程, 它直接影响 NIR 模型的预测能力。采用 TQ Analyzer V6.0 自动优选最佳光谱波长范围, 得到测定各组分的最佳光谱波长范围为: RMP 1981~2195 nm; INH 1540~1717 nm 和 2086~2197 nm; PZA 1460~1537 nm, 1956~2022 nm 和 2268~2393 nm。

3.2 输入层节点数的选择

由于每个样品的 NIR 光谱吸光度值数据较多,

输入层节点数较多, 用 RBFNN 法对 NIR 光谱进行解析时的计算量就较大, 网络运行的时间也就长。因此, 在本实验中采用主成分分析 (PCA) 法提取 NIR 的有效信息, 完成数据压缩, 减少计算量。并以前 4~15 个主成分得分作为神经网络的输入层节点, 在隐层节点数为 13, 扩展常数为 2 时, 不同输入层节点数对 RMSECV 值的影响见图 2。RMP, INH 和 PZA 3 个组分在输入层节点数分别为 7, 10 和 12 时, RMSECV 为最小值, 分别为 0.0144, 0.0108 和 0.0080。

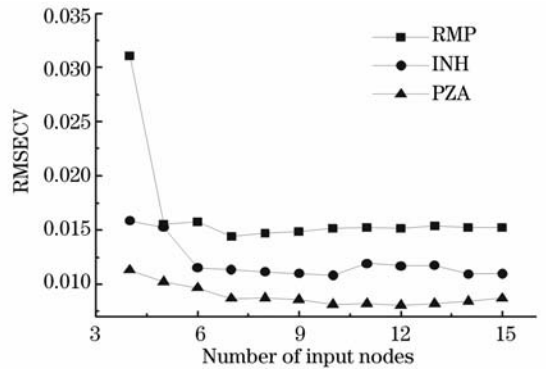


图2 输入层节点数的影响

Fig. 2 Effect of the number of input nodes

3.3 隐层节点数的选择^[12]

隐层节点数的选择对定量分析模型的建立影响较大, 其数目太少可能使网络不能收敛, 造成网络不能准确预测未知样本, 容错性差, 但太多又易造成网络学习时间过长及过拟合现象的出现。

分别以 7, 10 和 12 作为测定 RMP, INH 和 PZA 3 个组分含量的输入层节点数, 固定扩展常数为 2, 不同隐层节点数对模型的 RMSECV 值的影响见图 3。在测定 RMP, INH 和 PZA 含量时, 模型的隐层节点数分别达到 18, 15, 14 后, RMSECV 值已趋于平稳, 说明此时模型已得到充分拟合, 无需再增

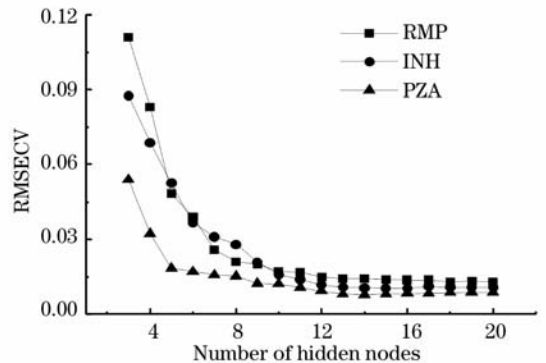


图3 隐层节点数的影响

Fig. 3 Effect of the number of hidden nodes

加隐层神经元而导致计算时间的增加。故测定 RMP, INH 和 PZA 含量最佳模型的隐层节点数分别为 18, 15 和 14。

3.4 扩展常数的选择

采用 2.2 和 2.3 优化出的最佳输入层节点数和

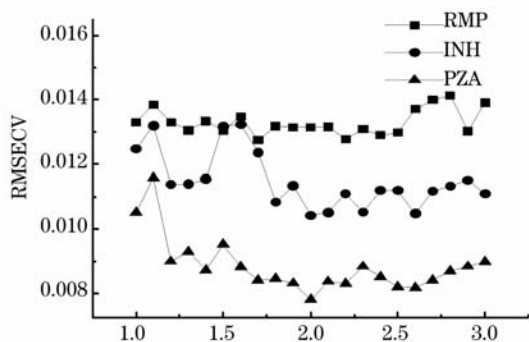


图 4 扩展常数的影响

Fig. 4 Effect of spread constant

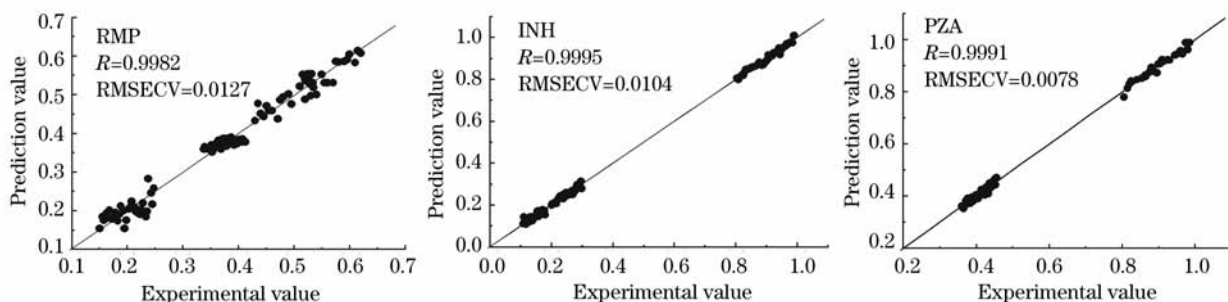


图 5 模型校正集交互验证预测值与测量值相关图

Fig. 5 Correlation between prediction values and experimental values obtained by cross-validation

3.6 最优模型的外部验证

外部验证一般采用未参与定标,但样品性质与参与定标的样品性质相似的样品,通过比较这些未参与定标样品的预测值与化学值的差异来判断模型的预测准确性。实验采用 55 个样品作为预测集,用

隐层节点数考察不同扩展常数对模型 RMSECV 的影响见图 4。测定 RMP, INH 和 PZA 含量扩展常数分别为 1.7, 2 和 2 时, RMSECV 为最小值。

3.5 最优模型的内部交互验证

经过优化得到最佳神经网络参数,测定 RMP、INH 和 PZA 质量分数的定量分析模型的最佳输入层节点数分别为 7, 10 和 12, 最佳隐含节点数为 18, 15 和 14, 最适扩展常数分别为 1.7, 2 和 2。以 HPLC 测定抗结核药物校正集样品中 RMP, INH 和 PZA 的含量为实验测量值,与 RBFNN 分析模型校正集交互验证预测值的相关图见图 5。其相关系数 R 分别为 0.9982, 0.9995 和 0.9991, RMSECV 值分别为 0.0127, 0.0104 和 0.0078, 可知交互验证预测值较好地逼近了 HPLC 测定值,模型的稳定性及拟合效果较好。

上述各优化参数所建校正模型对其进行外部验证。预测集样品各组分的预测值与 HPLC 法测定值进行成对数据双边 t 检验(见表 2), 各组分 HPLC 测定值和模型预测值间的 t 检验值均未达到显著水平,表明预测值与 HPLC 测定值间无显著差异。

表 2 模型预测集结果评价

Table 2 Evaluation of the prediction models with validation sets

Components	Samples	Actual average/%	Predicted average/%	t_{values}	$t_{0.05/n=60}$
RMP	55	22.23	22.08	0.969	2.000
INH	55	26.85	26.90	0.993	2.000
PZA	55	23.59	23.49	0.987	2.000

预测集 RMP, INH 和 PZA 的预测值,与 HPLC 法测定值的相关关系图见图 6。由图 6 可见,测定 RMP、INH 和 PZA 含量的定量分析模型

的 RMSEP 分别为 0.0125、0.0109 和 0.0103, 实验结果表明,所建的校正模型具有良好的预测能力,能够满足常规分析方法的精度要求。

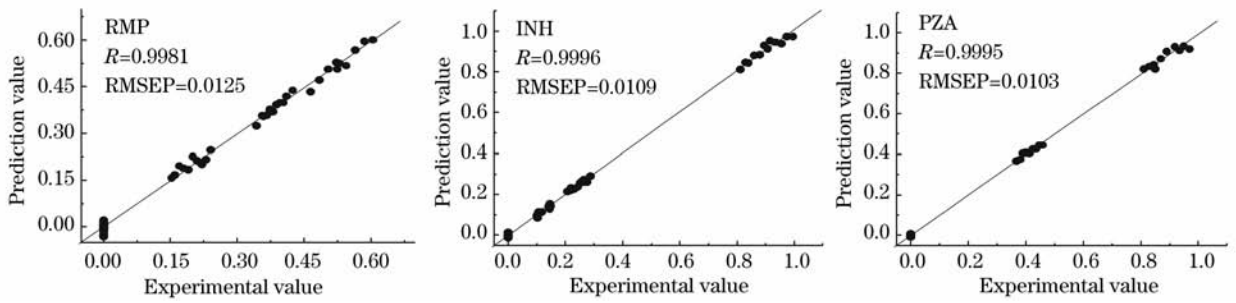


图 6 抗结核药物主成分质量分数预测值与测量值关系图

Fig. 6 Relationship between prediction and experimental values of anti-tuberculosis principal component mass fraction

4 结 论

应用近红外漫反射光谱与径向神经网络相结合,无损非破坏同时测定 5 种抗结核药物中 RMP, INH 和 PZA 含量是可行的。同传统方法相比,无需对样品进行预处理,因此可达到快速分析的目的,预测精度能够满足利福平片、异烟肼片、吡嗪酰胺片、异福片和异福酰胺片 5 种抗结核药物生产中定量分析的要求,可以分别应用于 5 种抗结核药物实际生产中的质量控制。

参 考 文 献

- 1 E. Calleri, E. D. Lorenzi, S. Furlanetto *et al.*. Validation of a RP-LC method for the simultaneous determination of isoniazid, pyrazinamide and rifampicin in a pharmaceutical formulation [J]. *J. Pharm. Biomed. Anal.*, 2002, **19**: 1089~1096
- 2 Espinosa-Mansilla A, Valenzuela MIA, de la Pena AM *et al.*. Determination of antitubercular drugs in urine and pharmaceuticals by LC using a gradient flow combined with programmed diode array photometric detection [J]. *Talanta*, 2002, **58**: 273~280
- 3 A. P. Argekar, S. S. Kunjir, K. S. Purandare. Simultaneous determination of rifampicin, isoniazid and pyrazinamid by high performance thin layer chromatography [J]. *J. Pharm. Biomed. Anal.*, 1996, **4**: 1645~1650
- 4 Espinosa-Mansilla A, Valenzuela MIA, de la Pena AM *et al.*. Comparative study of partial least squares and a modification of hybrid linear analysis calibration in the simultaneous spectrophotometric determination of rifampicin, pyrazinamide and isoniazid [J]. *Anal. Chim. Acta*, 2001, **427**: 129~136
- 5 Feng Fengqin, Zhang Hui, Wang Li *et al.*. Rapid detection of the quality indicators in glycerol monolaurate products using Fourier transform infrared spectroscopy [J]. *Acta Optica Sinica*, 2008, **28** (2): 326~330
冯凤琴, 张 辉, 王 莉 等. 应用红外光谱技术快速检测月桂酸单甘油脂的品质指标[J]. *光学学报*, 2008, **28** (2): 326~330
- 6 Zou Xiaobo, Zhao Jiewen. Methods of characteristic wavelength region and wavelength selection based on genetic algorithm[J]. *Acta Optica Sinica*, 2007, **27**(7): 1316~1321
邹小波, 赵杰文. 用遗传算法快速提取近红外光谱特征区域和特征波长[J]. *光学学报*, 2007, **27**(7): 1316~1321
- 7 Shi Youming, Liu Gang, Liu Jianhong *et al.*. Identification of auricularia auricula from different regions by Fourier transform infrared spectroscopy[J]. *Acta Optica Sinica*, 2007, **27**(1): 129~132
时有明, 刘 刚, 刘剑虹 等. 不同产地黑木耳的傅里叶变换红外光谱鉴别[J]. *光学学报*, 2007, **27**(1): 129~132
- 8 Xu Guangtong, Yuan Hongfu, Lu Wanzhen. Development of modern near infrared spectroscopic techniques and its applications [J]. *Spectroscopy and Spectral Analysis*, 2000, **20**(2): 134~142
徐广通, 袁洪福, 陆婉珍. 现代近红外光谱技术及应用进展[J]. *光谱学与光谱分析*, 2000, **20**(2): 134~142
- 9 Ni Lijun, Shi Xiaohao, Gao Xiujiao *et al.*. Research on the application of on-line detection and analytical technique by NIR in quality control of water extracting process of salvia miltiorrhiza[J]. *Chin Pharm J.*, 2004, **39**(8): 628~630
倪力军, 史晓浩, 高秀蛟 等. NIR 在线检测、分析在丹参水提过程质量控制中的应用[J]. *中国药学杂志*, 2004, **39**(8): 628~630
- 10 Qi Feng, Liu Wenqing, Zhou Bin *et al.*. Improving DOAS system measurement precision with artificial neural network method [J]. *Acta Optica Sinica*, 2002, **22**(11): 1345~1349
齐 锋, 刘文清, 周 斌 等. 利用人工神经网络方法提高差分光学吸收光谱系统测量精度研究[J]. *光学学报*, 2002, **22**(11): 1345~1349
- 11 You Jiaxun, Chen Chao, Guo Chuangxin. Application of RBF neural networks in the fruit color sorting machine[J]. *Light Industry Machinery*, 2008, **26**(3): 24~26
游家训, 成 超, 郭创新. 径向神经网络在水果颜色分级机中的应用[J]. *轻工机械*, 2008, **26**(3): 24~26
- 12 J. F. R. Gowe. Observations of in situ fluorescence of chlorophyll-a in Saavich inlet [J]. *Boundary-Layer Meteorology*, 1980, **18**: 235~248