

基于近红外光谱技术和分类与回归树算法建立天舒片崩解时间预测模型

刘秋安¹, 徐芳芳^{2,3*}, 张欣^{2,3}, 姜欣汝², 徐冰⁴, 吴云^{2,3}, 肖伟^{2,3}, 王振中^{1,2,3*}

1. 南京中医药大学, 江苏 南京 210023

2. 江苏康缘药业股份有限公司, 江苏 连云港 222001

3. 中药制药过程新技术国家重点实验室, 江苏 连云港 222001

4. 北京中医药大学, 北京 100029

摘要:目的 基于近红外光谱(near infrared spectrum, NIRS)技术, 建立一种快速预测天舒片崩解时间的方法。方法 采集39个批次共468个样品的NIRS, 对比分类和回归树(classification and regression trees, CART)算法与偏最小二乘(partial least-square, PLS)算法2种模型的预测效果, 建立天舒片崩解时间预测模型。结果 经基线校正处理后建立的CART模型性能最优。与PLS模型相比该模型将相对校正均方根偏差(relative root mean square error of correction, RRMSEC)由7.43%降低至4.94%, 相对预测均方根偏差(relative root mean square error of prediction, RRMSEP)由7.84%降低至7.66%。结论 NIRS技术结合CART算法预测天舒片崩解时间是可行的, 为天舒片崩解时间快速无损检测提供了一种新方法。

关键词: 近红外光谱技术; 分类和回归树算法; 崩解时间; 天舒片; 偏最小二乘算法; 相对校正均方根偏差; 相对预测均方根偏差

中图分类号: R283.6 文献标志码: A 文章编号: 0253-2670(2021)16-4837-07

DOI: 10.7501/j.issn.0253-2670.2021.16.008

Based on near-infrared spectroscopy technology and classification and regression trees algorithm to establish a prediction model of Tianshu Tablets disintegration

LIU Qiu-an¹, XU Fang-fang^{2,3}, ZHANG Xin^{2,3}, JIANG Xin-ru², XU Bing⁴, WU Yun^{2,3}, XIAO Wei^{2,3}, WANG Zhen-zhong^{1,2,3}

1. Nanjing University of Chinese Medicine, Nanjing 210023, China

2. Jiangsu Kanion Pharmaceutical Co., Ltd., Lianyungang 222001, China

3. State Key Laboratory of New-tech for Chinese Medicine Pharmaceutical Process, Lianyungang 222001, China

4. Beijing University of Chinese Medicine, Beijing 100029, China

Abstract: Objective A rapid method was established to predict the disintegration time of Tianshu Tablets (天舒片) based on near infrared spectroscopy (NIRS). **Methods** The near-infrared spectra of 468 samples from 39 batches were collected, and the disintegration time prediction model of Tianshu tablets was established by comparing the prediction effects of the partial least squares (PLS) and classification and regression tree (CART) models. **Results** The performance of the CART model was the best after the spectrum was preprocessed by the baseline correction, relative root mean square error of correction (RRMSEC) value of this model was decreased from 7.43% to 4.94%, relative root mean square error of prediction (RRMSEP) value was decreased from 7.84% to 7.66%. **Conclusion** It is feasible to predict the disintegration time of Tianshu Tablets with NIR spectroscopy technology and CART algorithm, which provides a new method for rapid and non-destructive testing of the disintegration time of Tianshu tablets.

Key words: near infrared spectroscopy; classification and regression tree algorithm; disintegration time; Tianshu Tablets; partial least-square; relative root mean square error of correction; relative root mean square error of prediction

收稿日期: 2021-03-12

基金项目: 国家“重大新药创制”科技重大专项: 基于功效成分群的中药口服固体制剂先进制药与信息化技术融合示范应用(2018ZX09201010-004)

作者简介: 刘秋安, 硕士研究生。E-mail: lqa82119245@163.com

*通信作者: 王振中, 研究员, 研究方向为中药新药研发。E-mail: kywyz@163.com

徐芳芳(1990—), 女, 博士, 研究方向为过程分析技术。E-mail: 879164331@qq.com

天舒片由天麻和川芎2味药组成,具有活血平肝、通络止痛的功效^[1]。崩解时间是天舒片质量控制中最重要的指标之一。目前,天舒片崩解时间按照《中国药典》2020年版中方法^[2]使用崩解仪进行测定,这种方法具有破坏性、劳动强度大且耗时。因此,亟需建立一种中药片剂崩解时间的快速检测方法,以应用于天舒片生产过程中的质量控制。

近几年,近红外光谱(near infrared spectrum, NIRS)技术作为一种无损的快速分析手段已经被研究应用于中药制药行业中,例如测定活性成分含量^[2-4]、水分^[5-7]以及混合均匀性^[8-10]等。Donoso等^[11]利用偏最小二乘回归(partial least-square, PLS)法分析茶碱片剂崩解时间与近红外光谱的关系,结果表明,崩解时间的增加会导致近红外吸收率的增加。王昀等^[12]采用PLS法建立基于NIRS技术的安乃近崩解时间快速分析模型,相关系数高达0.990 1。Zannikos等^[13]研究发现片剂的NIRS与溶解度都与水分变化有关,从而建立了NIRS预测卡马西平片溶出度的模型。上述研究表明,NIRS技术可以用来预测片剂崩解时间。

目前,国内对以NIRS技术预测中药片剂崩解时间的研究较少,且大多使用PLS法建立模型,NIRS技术结合分类和回归树(classification and regression trees, CART)算法的应用未见报道。CART算法是决策树的一种实现,本质是一种二分递归分割算法,可以直观的展示相关预测变量的作用和相对重要性,并且可以通过分段建模减弱线性回归中共线性问题对模型效果的影响^[14]。本研究尝试应用CART算法建立模型,旨在开发一种基于NIRS技术预测天舒片崩解时间的方法,以便后续为天舒片的先进制药提供技术支撑。

1 仪器与材料

1.1 仪器

ZB-1E型智能崩解仪,天津市天大天发科技有限公司;Antaris II傅立叶变换近红外分析仪,配积分球漫反射检测器,美国Thermo Fisher Scientific公司。

1.2 材料

天舒片素片由江苏康缘药业股份有限公司提供。39批次,每批12个共468个样本,批号分别为200701、200702、200703、200704、200705、200706、200707、200708、200801、200802、200901、200902、200903、200904、200905、200906、200907、

200908、200909、200910、201001、201002、201101、201201、201202、201203、201204、201205、201206、201210、201211、201212、201213、201214、201215、201216、201217、201218、201101。

2 方法与结果

2.1 NIRS 采集

将天舒片素片置于近红外光谱仪圆形药片固定装置卡槽内。采用近红外光谱仪采集光谱,波数范围为10 000~4000 cm^{-1} ,分辨率为8 cm^{-1} ,2倍增益,参比背景为空气,每条数据为4次扫描的平均值。按上述方法采集样品NIRS,结果如图1所示。可见,468个样品的NIRS图,吸光度略有差异,但整体趋势相似。

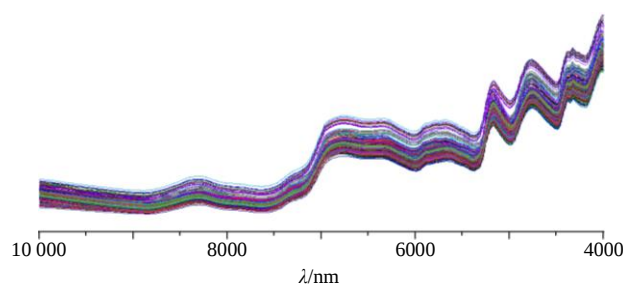


图1 468个样品的NIRS图($n=4$)

Fig. 1 NIRS of 468 samples ($n=4$)

2.2 崩解时间测量

采用《中国药典》2020年版四部通则崩解时限检查法,分别测定“2.1”项中采集NIRS后的天舒片素片的崩解时间。取药片1片置于崩解仪玻璃管中,因天舒片素片黏附挡板,故不加挡板,烧杯内盛有 $(37\pm1)^{\circ}\text{C}$ 的纯化水,往返频率为每分钟30~32次。39批共468个天舒片样品崩解时间如表1所示,可见,天舒片样品的崩解时间最小为18 min,最大为45 min。

2.3 数据处理

采用UnscrambleX 10.4(Camo software AS,挪威)软件对光谱进行预处理,采用Matlab 2020(美国MathWorks公司)软件进行PLS模型的样本划分、变量筛选及模型构建,采用SPM 8.3(美国Salford Systems公司)进行CART模型的样本划分及模型构建,采用Origin 8.0(美国OriginLab公司)软件绘图。

2.4 定量模型的建立

2.4.1 评价指标的确定及建模方法 本研究分别使用PLS算法与CART算法来建立预测模型。并通过训练集相关系数(correlation coefficient of training

表 1 39 批样品崩解时间 (n = 12)

Table 1 Disintegration time of 39 batches of samples (n = 12)

批号	崩解时间/min	RSD/%	批号	崩解时间/min	RSD/%	批号	崩解时间/min	RSD/%	批号	崩解时间/min	RSD/%
200701	37.17	3.78	200901	30.42	2.20	201001	33.67	3.19	201211	30.33	2.57
200702	25.92	2.58	200902	27.17	1.43	201002	34.67	3.95	201212	39.50	6.52
200703	23.33	3.34	200903	25.25	3.43	201101	19.17	4.89	201213	36.33	5.16
200704	27.75	4.89	200904	27.67	2.35	201201	20.67	7.25	201214	33.33	2.66
200705	23.08	2.23	200905	27.42	3.63	201202	22.33	3.97	201215	30.00	2.84
200706	24.17	2.97	200906	33.17	3.60	201203	20.75	5.09	201216	32.92	5.26
200707	23.75	3.17	200907	38.50	2.07	201204	30.75	2.82	201217	34.75	3.04
200708	33.58	3.90	200908	20.50	3.29	201205	31.83	1.81	201218	37.75	5.43
200801	37.50	3.86	200909	32.17	2.91	201206	29.42	3.39	210101	22.08	5.94
200802	37.92	4.56	200910	33.25	2.60	201210	33.25	1.87			

set, R_{cal})、验证集相关系数 (correlation coefficient of verification set, R_{pre})、校正均方根偏差 (root mean square error of correction, RMSEC)、预测均方根偏差 (root mean square error of prediction, RMSEP)、相对校正均方根偏差 (relative root mean square error of correction, RRMSEC)、相对预测均方根偏差 (relative root mean square error of prediction, RRMSEP)、性能偏差比 (ratio of performance deviation, RPD) 为指标评价模型优劣, 优选潜变量数、预处理方法以及变量筛选方法。 R 为 Pearson 相关系数, 其他相关指标计算公式如下。

$$RMSEC = [\sum_{i=1}^n (y_i - y_{ic})^2 / (n-1)]^{1/2} \quad (1)$$

$$RMSEP = [\sum_{i=1}^m (y_i - y_{ip})^2 / (m-1)]^{1/2} \quad (2)$$

$$RRMSEC = RMSEC / Y_{mc} \quad (3)$$

$$RRMSEP = RMSEP / Y_{mp} \quad (4)$$

$$RPD = [\sum_{i=1}^n (y_i - y_{mp})^2 / (n-1)]^{1/2} / RMSEP \quad (5)$$

y_i 为第 i 个样品的实测值, y_{ic} 是训练集中第 i 个样品的预测值, y_{ip} 是验证集中第 i 个样品的预测值, y_{mc} 为训练集样品的平均值, y_{mp} 为验证集样品的平均值, RMSEC、RMSEP、RRMSEC、RRMSEP 越小, R_{cal} 、 R_{pre} 、RPD 越大表示模型性能越好

CART 算法^[15]由以下 2 步组成: (1) 决策树生成: 基于训练数据生成决策树, 生成的决策树要尽量大; (2) 决策树剪枝: 用验证集对已生成的树进行剪枝并选择最优子树, 这时用损失函数最小作为剪枝的标准。

决策树的生成就是递归地构建二叉决策树的过

程。在训练数据集所在的输入空间中, 递归地将每个区域划分为 2 个子区域并决定每个子区域上的输出值, 构建二叉决策树。

(1) 选择最优切分变量 j 与切分点 s , 求解公式 (6)。

$$\min_{j,s} [\min_{c_1} \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 - \min_{c_2} \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2] \quad (6)$$

遍历变量 j 对固定的切分变量 j 扫描切分点 s , 选择使公式 (6) 达到最小值的对 (j, s) 。

(2) 用选定的对 (j, s) 划分区域并决定相应的输出值。

$$R_1(j, s) = \{x | x^{(j)} \leq s\}, R_2(j, s) = \{x | x^{(j)} > s\} \quad (7)$$

$$\hat{c}_m = \frac{1}{N_m} \sum_{x_i \in R_m(j,s)} y_i, x \in R_m, m=1, 2 \quad (8)$$

(3) 继续对 2 个子区域调用步骤 (1) (2), 直至满足停止条件。

(4) 将输入空间划分为 M 个区域 $R_1, R_2, \dots, R_M$, 生成决策树。

$$f(x) = \sum_{m=1}^M c_m I(x \in R_m) \quad (9)$$

CART 剪枝算法由 2 步组成: 首先从生成算法产生的决策树 T_0 底端开始不断剪枝, 直到 T_0 的根节点, 形成 1 个子树序列 $\{T_0, T_1, \dots, T_n\}$; 然后通过交叉验证法在独立的验证数据集上对子树序列进行测试, 从中选择最优子树。

2.4.2 PLS 算法建模

(1) 样本集划分: 采用 Kennard-Stone (K-S) 法将 468 个样品划分为训练集 (375 个) 和验证集 (93 个)。K-S 算法旨在通过最大化所选样本的仪器

响应向量 ($\vec{x}$) 之间的欧式距离, 以统一的方式覆盖多维空间。首先, 挑选样本集中欧式距离最大的 2 个样本作为校正集, 随后在每一次迭代中, 算法会选择距离 2 个样本的最小距离最大的样本加入校正集, 直至训练集样本数达到设定的数目^[16]。

(2) 光谱预处理: 采用光谱作为化学信息来源时, 在进行信息的提取和解析之前, 常常需要对光谱数据进行预处理, 以消除基线漂移和噪声对光谱带来的干扰。一阶导数 (1st) 可以消除背景的常数平移; 二阶导数 (2nd) 可以消除线性背景平移; 标准正则变换 (standard normal variate, SNV) 可以校正因样品不均匀造成的散射而引起的误差; 多元散射校正 (multiplicative scatter correction, MSC) 作用与标准正则变换相似; 基线校正 (baseline) 可扣除仪器背景和漂移对光谱信号的影响; S-G 平滑 (Svaitzky-Golay smoothing) 是通过多项式来对窗口内的数据进行多项式最小二乘拟合以消除噪音, 提高信噪比; 矢量归一化法 (normalize) 可消除光程变化对样品产生的影响^[17]。

本研究对比了以下几种预处理方法: S-G 平滑 (S-G)、矢量归一化法、SNV、MSC、基线校正、S-G 平滑+一阶导数 (S-G+1st)、S-G 平滑+二阶导数 (S-G+2nd)、标准正则变换+一阶导数 (SNV+1st)、多元散射校正+一阶导数 (MSC+1st)、基线校正+一阶导数 (基线校正+1st)、矢量归一化法+一阶导数 (归一化+1st)。

(3) 特征变量筛选: 组合间隔偏最小二乘法

(synergy interval PLS, siPLS) 是将光谱等分为多个区间, 并通过多个区间的随机组合建立 PLS 并选出最优组合。本研究考察了将全光谱等分为 20、30、40 个子区间, 并选择 2~4 个区间的随机组合建立 PLS 模型, 最终筛选最优组合区间为将全光谱分为 20 个子区间并选择 4 个随机组合。

移动窗口偏最小二乘法 (moving window PLS, mwPLS) 是基于移动窗口理念的间隔偏最小二乘模型, 对于每个变量, 将使用给定大小的窗口来进行 PLS 建模并选择最佳因子数。本研究考察了窗口宽度 (分别考察了全光谱总波数的 2%、5%、10%、15%, 即 31、77、155、233) 对模型的影响, 最终优选窗口宽度为 233 时的最佳建模波段。

(4) 不同预处理方法对 PLS 模型的影响: 不同预处理方法对 PLS 模型性能的影响如表 2 所示。综合比较各预处理方法所得模型的 RRMSEC、RRMSEP、RPD 可知, 采用 SNV 对光谱进行预处理后建立的 PLS 模型 RRMSEC 为 7.43%, RRMSEP 为 7.84%, RPD 为 2.40, 模型性能最好。

(5) 不同变量筛选方法对 PLS 模型的影响: 不同变量筛选方法对模型性能的影响如表 3 所示。综合比较采用各变量筛选方法所得模型的 RRMSEC、RRMSEP 以及 RPD 值可知, 采用全波长建模的模型 RRMSEP 最小, RPD 最大, 模型性能优于其他筛选变量后的模型, 故不进行变量筛选, 选用全光谱模型。

2.4.3 CART 算法建模

表 2 不同预处理方法对 PLS 模型性能的影响

Table 2 Influence of different pretreatment methods on performance of PLS model

预处理方法	训练集			验证集			RPD	潜变量数
	R_{cal}	RMSEC/min	RRMSEC/%	R_{pre}	RMSEP/min	RRMSEP/%		
无预处理	0.92	2.27	7.67	0.85	3.07	10.02	1.93	12
S-G	0.90	2.60	8.76	0.84	3.16	10.32	1.87	10
S-G+1st	0.93	2.22	7.46	0.88	2.72	9.02	2.08	7
S-G+2nd	0.93	2.24	7.55	0.81	3.22	10.52	1.72	15
SNV	0.93	2.22	7.43	0.91	2.36	7.84	2.40	7
SNV+1st	0.93	2.08	6.98	0.87	3.17	10.55	2.03	11
基线校正	0.93	2.25	7.64	0.81	3.23	10.34	1.66	9
基线校正+1st	0.94	2.05	6.96	0.76	3.59	11.52	1.55	5
归一化	0.92	2.32	7.86	0.87	2.81	9.06	2.07	6
归一化+1st	0.93	2.15	7.27	0.80	3.66	11.84	1.68	11
MSC	0.93	2.22	7.44	0.91	2.36	7.99	2.40	8
MSC+1st	0.93	2.08	6.99	0.87	3.18	10.57	2.03	11

表 3 不同变量筛选方法对 PLS 模型的影响

Table 3 Performance of PLS model established by different variable screening methods

变量筛选方法	变量区间/个数	R_{cal}	R_{pre}	RRMSEC/%	RRMSEP/%	RPD	潜变量数
无	3 999.640~10 001.030	0.93	0.91	7.43	7.84	2.40	7
siPLS	5 203.003~5 499.987	0.94	0.90	7.02	8.32	2.31	13
	5 804.685~6 101.669						
	7 910.570~8 207.555						
	8 211.411~8 508.396						
mwPLS	5 631.123~6 525.932	0.92	0.84	7.87	10.04	1.81	12

(1) 样本集划分: 样本集划分采用软件自带样本划分功能, 按照随机种子数随机抽取样本, 将 468 个样本划分为训练集 (382 个) 和验证集 (86 个), 统计结果见表 4。

(2) 不同预处理方法对 CART 模型的影响: 光谱预处理方法与“2.4.2”项中所用光谱预处理方法相同。结果如表 5 所示, 采用基线校正对光谱进行预处理后建立的 CART 模型, RRMSEP 值最小为 7.66%, RRMSEC 为 4.94%, 模型性能最优, 验证集预测结果如图 2 所示。

2.5 PLS 算法模型与 CART 算法模型比较

2 种算法的最优模型如表 6 所示。CART 算法所得模型预测误差较小, RRMSEC 为 4.94%, RRMSEP 为 7.66%, 模型预测精度更准确。

3 讨论

本研究建立了一种基于 NIRS 技术预测天舒片崩解时间的分析方法, 可以高效、快速、无损的预测天舒片崩解时间, 将崩解时间测量时间由 40 min 缩短至 5 s 左右, 并由破坏性检测转变为无损检测, 提高了天舒片崩解时间的检测效率。光谱采用基线

表 4 CART 模型训练集和验证集数据统计结果

Table 4 Statistical results of calibration set and validation set of CART model

样本	数量	占比/%	崩解时间/min					
			平均值	最小	下四分位	中位数	上四分位	最大
训练集	382	81.62	29.80	18.00	25.00	31.00	34.00	45.00
验证集	86	18.38	29.94	18.00	24.00	30.00	35.00	40.00

表 5 不同预处理方法对 CART 模型性能的影响

Table 5 Influence of different pretreatment methods on performance of CART model

预处理方法	训练集			验证集		
	R_{cal}	RMSEC/min	RRMSEC/%	R_{pre}	RMSEP/min	RRMSEP/%
无预处理	0.75	2.92	9.78	0.58	4.08	13.63
MSC	0.94	1.36	4.57	0.71	3.39	11.32
MSC+1st	0.93	1.54	5.17	0.62	3.95	13.19
SG	0.76	2.86	9.58	0.63	3.80	12.68
SG+1st	0.97	1.02	3.42	0.72	3.36	11.22
SG+2	0.89	1.93	6.47	0.68	3.54	11.83
SNV	0.94	1.36	4.57	0.81	2.73	9.11
SNV+1st	0.97	0.91	3.07	0.62	3.98	13.29
基线校正	0.94	1.47	4.94	0.87	2.29	7.66
基线校正+1st	0.91	1.77	5.94	0.77	3.05	10.19
均一化	0.98	0.87	2.91	0.70	3.56	11.89
均一化+1st	0.81	2.52	8.47	0.68	3.56	11.88

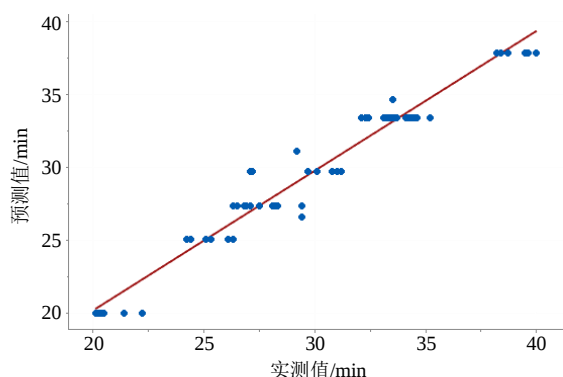


图 2 天舒片崩解时间的 NIRS 预测值与实测值相关性分析
Fig. 2 Correlation analysis between predicted value of NIRS and measured value of disintegration time of Tianshu Tablets

表 6 2 种算法所得模型效果比较

Table 6 Comparison of model effects obtained by two algorithms

算法	训练集			验证集		
	R_{cal}	RMSEC/min	RRMSEC/%	R_{pre}	RMSEP/min	RRMSEP/%
PLS	0.93	2.22	7.43	0.91	2.36	7.84
CART	0.94	1.47	4.94	0.87	2.29	7.66

江苏康缘药业股份有限公司研发了中药生产过程知识信息管理系统^[20] (process knowledge system, PKS), 后续可以将此模型植入到该 PKS 系统中, 实现压片过程中天舒片崩解时间的快速准确预测, 实时反馈产品质量信息, 及时调整相关工艺参数, 尝试进行天舒片压片工段的智能放行, 以保证产品质量。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] 中国药典 [S]. 二部. 2020: 129, 628.
- [2] 何月云, 梁华伦, 苏曹豪, 等. 近红外技术在小柴胡颗粒中黄芩苷快速检测的应用研究 [J]. 今日药学, 2019, 29(7): 461-463.
- [3] 高瑞琳, 杨鹏硕, 许刚, 等. 基于系统建模思想的脑心通胶囊中丹酚酸 B 近红外定量建模 [J]. 光谱学与光谱分析, 2020, 40(11): 3573-3578.
- [4] 吕尚, 周海滨, 汪俊, 等. 基于近红外光谱的银杏叶提取液总黄酮醇苷快速检测研究 [J]. 药物分析杂志, 2017, 37(5): 927-933.
- [5] 张恩华, 邱选兵, 魏永卜, 等. 基于方波激励的近红外 LED 中药水分传感器 [J]. 光谱学与光谱分析, 2020, 40(5): 1656-1660.
- [6] 时博, 谢惠英, 雷敬卫, 等. 近红外光谱法在知母和黄柏饮片水分分析中的应用 [J]. 中华中医药学刊, 2017,

校正方法进行预处理, CART 算法建立模型。结果显示, 模型预测误差为 2.29 min, 实际生产过程中产品崩解时间分布于 (29.82 ± 5.87) min, 《中国药典》2020 年版要求为 ≤ 60 min, 故本模型预测性能可满足生产过程中产品崩解时间的检测。

随着过程分析技术 (process analysis technology, PAT) 的发展, 各个制药企业都在尝试应用 PAT 工具来提升产品质量控制水平, 提高生产效率。礼来公司开发了一种用于压片机供料靴内的 PAT 工具, 用于实时监测最终混合物的活性药物成分 (active pharmaceutical ingredient, API) 浓度, 该工具可以通过自动触发压片机卸料斜槽而清除不合格的药片, 从而保证产品含量^[18-19]。

35(3): 669-672.

- [7] 王晴, 徐芳芳, 张欣, 等. 在线近红外光谱监测桂枝茯苓胶囊流化床干燥过程水分的方法研究 [J]. 中草药, 2019, 50(22): 5429-5438.
- [8] 刘燎原, 梁志毅, 刘丽萍, 等. 基于近红外光谱技术的砂仁提取物混合工艺研究 [J]. 中国现代中药, 2020, 22(1): 85-88, 93.
- [9] 万娜, 林环玉, 伍振峰, 等. 基于中药粒子设计结合近红外光谱技术研究清润丸原料粉末的混合均匀性 [J]. 中草药, 2020, 51(17): 4425-4432.
- [10] 杨婵, 徐冰, 张志强, 等. 基于移动窗 F 检验法的中药配方颗粒混合均匀度近红外分析研究 [J]. 中国中药杂志, 2016, 41(19): 3557-3562.
- [11] Donoso M, Ghaly E S. Prediction of tablets disintegration times using near-infrared diffuse reflectance spectroscopy as a nondestructive method [J]. *Pharm Dev Technol*, 2005, 10(2): 211-217.
- [12] 王昀, 孟庆华. 近红外光谱法测定药物崩解时限应用研究 [J]. 海峡药学, 2011, 23(12): 69-71.
- [13] Zannikos P N, Li W I, Drennen J K, et al. Spectrophotometric prediction of the dissolution rate of carbamazepine tablets [J]. *Pharm Res*, 1991, 8(8): 974-978.
- [14] Loh W Y. Classification and regression trees [J]. *WIREs Data Min Knowl Disc*, 2011, 1(1): 14-23.

- [15] 李航. 统计学习方法 [M]. 北京: 清华大学出版社, 2012: 81.
- [16] Galvao R, Araujo M, Jose G, *et al.* A method for calibration and validation subset partitioning [J]. *Talanta*, 2005, 67(4): 736-740.
- [17] 倪力军, 张立国. 基础化学计量学及其应用 [M]. 上海: 华东理工大学出版社, 2011: 56-70.
- [18] Harms Z D, Shi Z Q, Kulkarni R A, *et al.* Characterization of near-infrared and Raman spectroscopy for in-line monitoring of a low-drug load formulation in a continuous manufacturing process [J]. *Anal Chem*, 2019, 91(13): 8045-8053.
- [19] Manley L, Hilden J, Valero P, *et al.* Tablet compression force as a process analytical technology (PAT): 100% Inspection and control of tablet weight uniformity [J]. *J Pharm Sci*, 2019, 108(1): 485-493.
- [20] 王磊. 热毒宁注射液关键生产工段智能放行技术的研究 [D]. 杭州: 浙江大学, 2018.

[责任编辑 郑礼胜]