

茶叶中低含量氨基酸近红外光谱定量分析模型研究*

郝 勇^{1,2} 陈 斌²

(1. 华东交通大学机电工程学院, 南昌 330013; 2. 江苏大学食品与生物工程学院, 镇江 212013)

摘要:应用近红外光谱分析方法对茶叶中游离氨基酸进行定量分析。连续小波导数(CWD)和标准正态变量变换(SNV)用于光谱预处理;偏最小二乘回归(PLSR)方法用于校正模型构建;采用蒙特卡洛无信息变量消除(MCUVE)方法和连续投影算法(SPA)对建模变量进行优化。结果表明,CWD-SNV方法可以有效地提高茶叶光谱质量,消除光谱的平移误差;基于MCUVE-SPA的变量筛选方法极大地改善了模型的精度,实现了建模变量的有效压缩,模型的预测相关系数(R_p)和预测均方根误差(RMSEP)分别由0.851和0.117改善为0.895和0.107,建模变量由4 148减小为18;当氨基酸百分含量大于0.1%时,近红外光谱结合化学计量学方法可以得到较优的定量分析模型。为茶叶中低含量氨基酸的分析提供了一种快速简便的分析方法。

关键词:茶叶 氨基酸 连续小波导数 变量筛选 近红外光谱

中图分类号: O657.33; O656.3 **文献标识码:** A **文章编号:** 1000-1298(2014)06-0216-05

引言

茶叶中含有20多种游离氨基酸,主要包括茶氨酸、谷氨酸、谷氨酰胺、丙氨酸、丝氨酸、天冬氨酸等,茶叶中游离氨基酸总量是茶叶中所有游离氨基酸含量之和,其含量直接影响茶叶的品质,是评价茶叶鲜爽度的重要指标^[1-4]。氨基酸是生物体代谢过程中起重要作用的物质,在许多生理机能方面都起着不可替代的作用。

茶叶中游离氨基酸的检测方法主要包括国家标准检测方法GB/T 8314—2002和高效液相色谱^[5]、液相色谱-质谱联用^[6]、气相色谱-质谱联用^[7]和氨基酸分析仪^[8]等仪器分析方法。其中,国家标准方法需要配制缓冲液、茚三酮溶液以及氨基酸标准溶液等,操作过程复杂耗时;仪器分析方法需要对样品进行衍生化和显色反应,样品预处理程序同样比较繁琐。

近红外光谱分析技术是一种快速、无损、低成本、无污染的分析技术,它可以充分利用全谱段或多波长下的光谱信息对物质的品质、种类、化学成分进行定量和定性分析,在农产品品质检测中得到广泛应用^[9-12]。本研究采用近红外光谱分析技术对茶叶中的游离氨基酸组分进行定量分析,探讨不同的光谱预处理方法和特征变量筛选方法对氨基酸分析精度的影响,为近红外光谱检测茶叶中低含量的游

离氨基酸提供一种快速、简便无损的分析方法。

1 实验材料与方法

1.1 实验样品及仪器

收集江苏、浙江、云南等不同地区的茶叶样品135个,品种包括白茶、高山茶和乌龙茶,样品经研磨粉碎过500目筛制得茶粉。试验采用德国布鲁克公司生产的MPA(Multi Purpose Analyzer)型多功能傅里叶变换近红外光谱仪。光谱扫描范围10 000~4 000 cm^{-1} ,分辨率为4 cm^{-1} ,扫描次数为32次。试验温度为20℃。以仪器内置金体作为背景。为了确保光谱数据的准确度,每个样品采集3次,平均光谱作为茶叶样品的最终光谱。图1所示为茶叶样品的近红外光谱图。

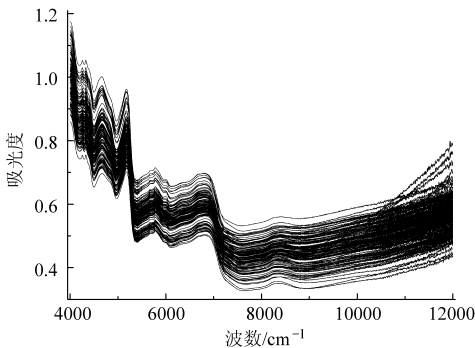


图1 茶叶样品的近红外光谱图

Fig.1 NIR spectra of tea samples

采集光谱后,按照 GB/T 8314—2002 测定茶叶样本中游离氨基酸的百分含量。按照氨基酸百分含量平均分布原则,将 135 个样品按照 2:1 的比例进行校正集(90 个)和验证集(45 个)的划分,从而保证了校正集的样品组分百分含量信息包含验证集样品的信息。茶叶氨基酸百分含量的统计信息如表 1 所示。

表 1 茶叶氨基酸百分含量的统计信息

Tab.1 Statistic information of amino acid contents in tea samples				%
样品集	样本数	范围	平均值	标准偏差
全部	135	0.020 ~ 1.230	0.547	0.225
校正	90	0.020 ~ 1.230	0.546	0.225
测试	45	0.040 ~ 1.150	0.547	0.226

1.2 光谱数据预处理及变量筛选

在光谱分析中,原始光谱含有与样品组分无关的信息,如环境温度、检测器电噪声以及样品背景干扰等因素,需采用预处理方法进行消除。分别采用标准正态变量变换(SNV)、连续小波导数(CWD)^[13]以及 SNV-CWD 方法对光谱进行有效信息提取,探讨预处理方法对模型精度的影响。

为了从大量的光谱变量中提取与待分析目标物间存在相关性的特征变量,分别采用蒙特卡罗无信息变量消除(Monte Carlo uninformation variable elimination, MCUVE)和连续投影算法(Successive projections algorithm, SPA)对建模变量进行优选。MCUVE 方法是基于偏最小二乘回归(Partial least squares regression, PLSR)系数提出的一种波长选择方法^[14-16]。该方法通过建立一系列的 PLSR 模型,并引入稳定性(Stability)来对 PLSR 模型的回归系数进行统计,评价模型中每个变量的可靠性,以此决定每个变量的取舍。该方法已被广泛地应用于光谱变量的选择,是目前最为流行的一种波长选择方法。SPA 方法能够有效消除众多波长变量间的共线性影响,降低模型的复杂度,以其简便、快速的特点得到广泛应用,在多种样品组分的特征波长选取中取得了很好的效果^[17-18]。连续投影算法是一种前向循环选择方法,它从一个波长开始,每次循环,计算它在未选入波长上的投影,将投影向量最大的波长引入到波长组合。每一个新选入的波长都与前一个线性关系最小。

1.3 模型的建立和评价

采用 PLSR 方法建立光谱校正模型,模型的评价指标包括交叉验证均方根误差(Root mean square error of cross-validation, RMSECV)、交互验证相关系数(Correlation coefficient of cross-validation, R_{cv})、预

测均方根误差(Root mean square error of prediction, RMSEP)和预测相关系数(Correlation coefficient of prediction, R_p)。Matlab R2010a 用于光谱数据预处理、变量筛选程序计算和 PLSR 模型的建立。其中 R_{cv}/R_p 值越大, RMSECV/RMSEP 值越小,模型的预测能力越好。

2 结果与分析

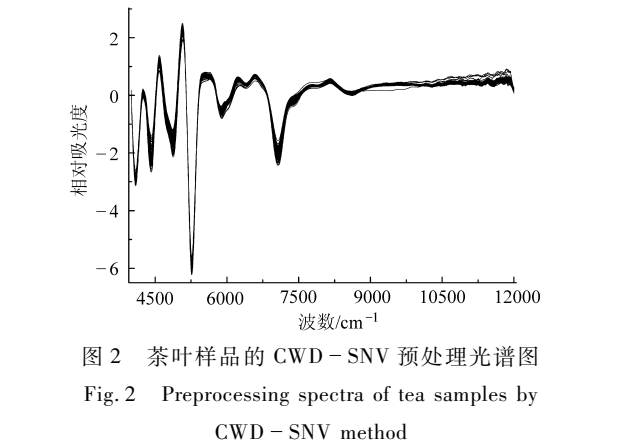
2.1 光谱预处理方法的选择

茶叶样品的近红外光谱包含了丰富的物质信息,样品光谱间的差异,除包含样品自身的组分信息外,还包含检测器噪声、样品背景和杂散光等引起的光谱噪声。为了降低光谱的平移噪声,提高光谱的信噪比,减少各种干扰的影响。分别采用 CWD、SNV 和 CWD-SNV3 种预处理方法对原始光谱进行有效信息提取,并对原始光谱和预处理后的光谱进行 PLSR 建模分析,模型结果如表 2 所示。

表 2 不同光谱预处理方法对茶叶氨基酸分析模型精度的影响结果

Tab.2 Results of PLSR models for amino acid with different spectral preprocessing methods						%
预处理方法	因子数	R_{cv}	RMSECV	R_p	RMSEP	
无	15	0.817	0.133	0.787	0.140	
SNV	9	0.811	0.136	0.843	0.122	
CWD	15	0.850	0.123	0.851	0.118	
CWD-SNV	16	0.853	0.121	0.851	0.117	

由表 2 可知,光谱经 CWD-SNV 预处理方法得到了较好的分析结果。图 2 所示为茶叶样品的 CWD-SNV 光谱,从图中看出,预处理后的光谱有效地消除了平移误差,光谱的平滑度得到显著提高。



2.2 组合变量优选方法提高模型分析精度

在建立多元校正模型时,通过选择包含样品性质特征或其所含成份性质特征的波长(变量),代替全谱去建立模型,会得到更准确的校正模型。分别采用 MCUVE 和 MCUVE-SPA 两种变量筛选方法

对茶叶的 CWD - SNV 光谱进行建模变量优选。MCUVE 方法选择变量过程中,所选变量构建模型的 RMSECV 随选择的变量数的变化趋势如图 3a 所示。从图中可知,当入选变量数为 580 时,模型的 RMSECV 最小,模型精度最高。采用 MCUVE - SPA 方法进行变量优化时,不同变量数构建的模型的 RMSECV 随变量数目的变化曲线如图 3b 所示。从图中可知,当选择变量数为 18 时,模型的 RMSECV 最小。18 个变量在光谱中的分布如图 4 所示。

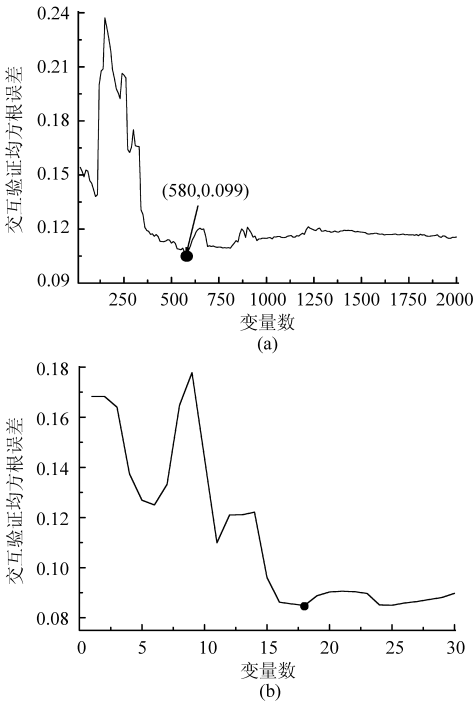


图 3 不同变量构建的氨基酸模型的 RMSECV 随保留变量数目的变化曲线

Fig.3 Variations of RMSECV for amino acid models with different reserved variables
(a) MCUVE (b) MCUVE - SPA

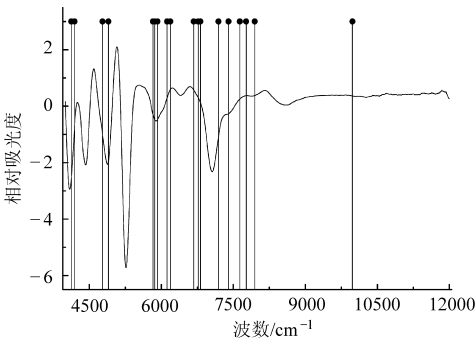


图 4 MCUVE - SPA 方法选择变量的分布图
Fig.4 Distribution of variables selected by MCUVE - SPA methods

比较所选变量在 CWD - SNV 预处理光谱和原始光谱中的分布(原始光谱进行了幅值放大处理,峰位没有改变),由图 4 可知,选择的 18 个变量对应的波数分别为:4 126.9、4 194.3、4 774.8、4 898.2、

5 825.8、5 864.4、5 922.2、6 120.9、6 188.3、6 676.2、6 766.9、6 815.1、7 191.1、7 399.4、7 634.7、7 769.7、7 947.1 和 9 981.6 cm^{-1} ,其中,4 100 ~ 4 200 cm^{-1} 间包含芳香族化合物的 C—H 键合频吸收;4 900 cm^{-1} 附近为 O—H 键的合频吸收;5 800 cm^{-1} 附近为 C—H 键的倍频吸收和 CH_3 基团吸收;6 700 cm^{-1} 附近为 N—H 键的倍频吸收;7 000 cm^{-1} 附近为 O—H 的倍频吸收;7 600 cm^{-1} 为 CH_2 基团的吸收;10 000 cm^{-1} 附近为 C—H 的二倍频吸收,峰位较好地反映了物质的特征吸收^[19-20]。

2.3 模型的验证

分别对 MCUVE 和 MCUVE - SPA 所选变量进行 PLSR 模型的构建,模型的校正和验证结果如表 3 所示。从表中可知,与全谱和 MCUVE 方法相比,MCUVE - SPA 方法不仅提高了模型的校正和预测精度,而且极大地减小了建模变量数目,模型 R_{cv} 和 RMSECV 分别改善为 0.911 和 0.093,建模变量数由原始光谱的 4 148 减少为 18。采用优选出的 18 个变量进行建模并对验证集样品进行预测,预测结果如表 3 所示。模型的 R_p 和 RMSEP 分别由原始光谱的 0.851 和 0.117 改善为 0.895 和 0.107。

表 3 茶叶氨基酸精简模型的构建及验证
Tab.3 Construction and validation of concise models for amino acid contents in tea %

变量筛选方法	变量数	因子数	R_{cv}	RMSECV	R_p	RMSEP
全部光谱	4 148	16	0.853	0.121	0.851	0.117
MCUVE	580	11	0.898	0.099	0.868	0.112
MCUVE - SPA	18	14	0.911	0.093	0.895	0.107

模型的校正集样品和验证集样品的氨基酸百分含量的标准方法(GB/T 8314—2002)测定值和经 CWD - SNV - MCUVE - SPA 预处理后光谱的 PLSR 模型的交互验证/预测值的拟合结果如图 5a 所示。由图可知,氨基酸百分含量的参考值和预测值存在较好的相关关系。然而对于参考百分含量为 0.04% 的样品校正和预测结果均较差(出现负值)。虽然近红外光谱分析方法文献报道的对农产品组分的检测限为 0.01%,然而对于茶叶中氨基酸组分,当含量较低时是无法正确检测的。

将氨基酸百分含量低于 0.1% 的样品从校正和验证集中去掉(剩余 131 个样品),采用所选的 18 个建模变量进行模型的构建和验证(即 87 个样品建模,预测 44 个样品)。模型的 RMSECV 和 RMSEP 分别为 0.095 和 0.102; R_{cv} 和 R_p 分别为 0.888 和 0.888。模型校正集/验证集中氨基酸百分含量的校正/预测值和参考值的相关性如图 5b 所示,虽然氨

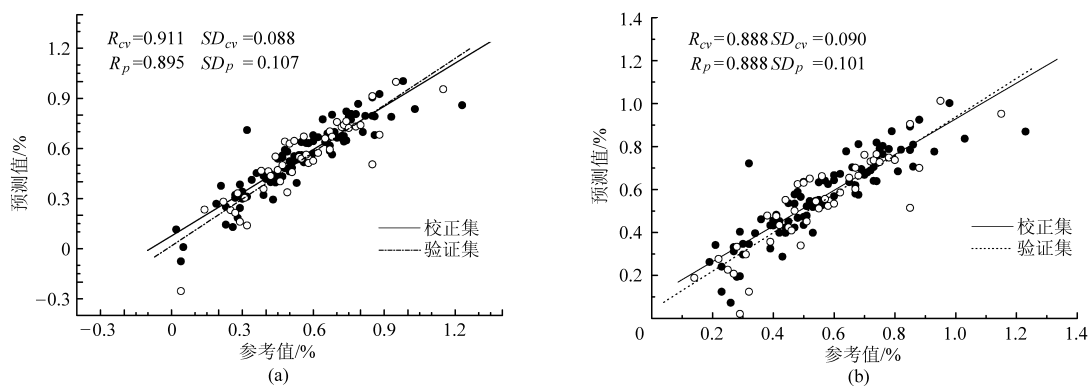


图 5 氨基酸样品集模型的参考值和预测值的相关关系

Fig. 5 Correlation relationship of reference and predicted percentage contents for amino acid with different concentration range
(a) 百分含量大于 0.01% (b) 百分含量大于 0.1%

氨基酸含量接近 0.1% 的样品的预测结果较差,然而模型的交互验证和预测结果均为正值;随着氨基酸百分含量的增加,模型的预测结果呈现接近参考结果的趋势,预测结果趋于准确。

3 结束语

近红外光谱分析方法应用于茶叶中低含量氨基酸的定量分析,不同的光谱预处理方法和特征波长

选择方法用于光谱模型的优化。结果表明,茶叶的近红外光谱经连续小波导数-标准正态变量变换(CWD-SNV)预处理后,结合蒙特卡罗无信息变量消除-连续投影算法(MCUVE-SPA)的变量筛选方法,可以对茶叶中百分含量大于 0.1% 的氨基酸进行定量分析。近红外光谱结合化学计量学方法为检测茶叶中低含量的游离氨基酸提供一种快速、高效和简便无损的分析方法。

参 考 文 献

1 Huo Danqun, Wu Yu, Yang Mei, et al. Discrimination of Chinese green tea according to varieties and grade levels using artificial nose and tongue based on colorimetric sensor arrays [J]. Food Chemistry, 2014, 145: 639 – 645.

2 陈美丽,唐德松,张颖彬,等. GC-MS 结合化学计量学对茶叶品质的判别研究[J]. 浙江大学学报:农业与生命科学版, 2013, 39(1): 84 – 91.

Chen Meili, Tang Desong, Zhang Yingbin, et al. Study on the tea quality evaluation using GC-MS coupling chemometrics[J]. Journal of Zhejiang University: Agriculture and Life Science, 2013, 39(1): 84 – 91. (in Chinese)

3 Kocadagli T, Ozdemir K S, Gokmen V. Effects of infusion conditions and decaffeination on free amino acid profiles of green and black tea[J]. Food Research International, 2013, 53(1): 720 – 725.

4 Horanni R, Engelhardt U H. Determination of amino acids in white, green, black, oolong, puerh teas and tea products[J]. Journal of Food Composition and Analysis, 2013, 31(1): 94 – 100.

5 Zhang Xiaoli, Zhao Tong, Cheng Ting, et al. Rapid resolution liquid chromatography (RRLC) analysis of amino acids using pre-column derivatization[J]. Journal of Chromatography B, 2012, 906: 91 – 95.

6 Le A, Ng A, Kwan T, et al. A rapid, sensitive method for quantitative analysis of underivatized amino acids by liquid chromatography-tandem mass spectrometry (LC-MS/MS) [J]. Journal of Chromatography B, 2014, 944: 166 – 174.

7 Pinto M C X, de Paiva M J N, Oliveira-Lima O C, et al. Neurochemical study of amino acids in rodent brain structures using an improved gas chromatography-mass spectrometry method[J]. Journal of Chemical Neuroanatomy, 2014, 55: 24 – 37.

8 Zhang Hua, Wang Zhenyu, Yang Xin, et al. Determination of free amino acids and 18 elements in freeze-dried strawberry and blueberry fruit using an amino acid analyzer and ICP-MS with micro-wave digestion[J]. Food Chemistry, 2014, 147: 189 – 194.

9 Chen Quansheng, Guo Zhiming, Zhao Jiewen, et al. Comparisons of different regressions tools in measurement of antioxidant activity in green tea using near infrared spectroscopy[J]. Journal of Pharmaceutical and Biomedical Analysis, 2012, 60: 92 – 97.

10 Coppa M, Revello-Chion A, Giaccone D, et al. Comparison of near and medium infrared spectroscopy to predict fatty acid composition on fresh and thawed milk [J]. Food Chemistry, 2014, 150: 49 – 57.

11 吴伟斌,冯灼峰,洪添胜,等. 基于近红外反射的柑橘重叠叶片叶面积指数测量系统[J]. 农业机械学报, 2013, 44(1): 179 – 184.

Wu Weibin, Feng Zhuofeng, Hong Tiansheng, et al. Measurement system for citrus LAI based on near-infrared reflection[J]. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(1): 179 – 184. (in Chinese)

12 李江波,赵春江,陈立平,等. 基于可见/近红外光谱谱区有效波长的梨品种鉴别[J]. 农业机械学报, 2013, 44(3): 153 – 158.

Li Jiangbo, Zhao Chunjiang, Chen Liping, et al. Variety identification of pears based on effective wavelengths in visible/near infrared region [J]. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(3): 153 – 158. (in Chinese)

13 Mohammadpour K, Sohrabi M R, Jourabchi A. Continuous wavelet and derivative transform applied to the overlapping spectra for the quantitative spectrophotometric multi-resolution of triamterene and hydrochlorothiazide in triamterene-H tablets[J]. Talanta, 2010, 81(4): 1821 – 1825.

14 郝勇, 孙旭东, 潘圆媛, 等. 蒙特卡罗无信息变量消除方法用于近红外光谱预测果品硬度和表面色泽的研究[J]. 光谱学与光谱分析, 2011, 31(5): 1225 – 1229.

Hao Yong, Sun Xudong, Pan Yuanyuan, et al. Detection of firmness and surface color of pear by near infrared spectroscopy based on monte carlo uninformative variables elimination method [J]. Spectroscopy and Spectral Analysis, 2011, 31(5): 1225 – 1229. (in Chinese)

15 Huang Zhuangrong, Sha Sha, Rong Zhengqin, et al. Feasibility study of near infrared spectroscopy with variable selection for non-destructive determination of quality parameters in shell-intact cottonseed[J]. Industrial Crops and Products, 2013, 43: 654 – 660.

16 Cai Wensheng, Li Yankun, Shao Xueguang. A variable selection method based on uninformative variable elimination for multivariate calibration of near-infrared spectra[J]. Chemometrics and Intelligent Laboratory Systems, 2008, 90(2): 188 – 194.

17 Marques A D S, de Melo M C N, Cidral T A, et al. Feature selection strategies for identification of *Staphylococcus aureus* recovered in blood cultures using FT – IR spectroscopy successive projections algorithm for variable selection; a case study[J]. Journal of Microbiological Methods, 2014, 98(1): 26 – 30.

18 Insausti M, Gomes A A, Cruz F V, et al. Screening analysis of biodiesel feedstock using UV-vis, NIR and synchronous fluorescence spectrometries and the successive projections algorithm[J]. Talanta, 2012, 97: 579 – 583.

19 Rodriguez-Saona L E, McLaughlin M A, Fry F S, et al. Rapid analysis of sugars in fruit juices by FT – NIR [J]. Carbohydrate Research, 2001, 336(1): 63 – 74.

20 Burns D A, Ciurczak E W. Handbook of near-infrared analysis, 3rd edition: practical spectroscopy[M]. Boca Raton: CRC Press Inc., 2007.

Quantitative Determination of Low Amino Acid Contents in Tea by Using Near-infrared Spectroscopy

Hao Yong^{1,2} Chen Bin²

(1. College of Mechanical and Electronic Engineering, East China Jiaotong University, Nanchang 330013, China
2. School of Food and Biological Engineering, Jiangsu University, Zhenjiang 212013, China)

Abstract: Near-infrared spectroscopy (NIRS) was used for quantitative determination of free amino acid contents in tea samples. Two spectral preprocessing methods including continuous wavelet derivative (CWD) and standard normal variate (SNV) were used for spectral transform. Partial least squares regression (PLSR) was used for modeling. Monte Carlo uninformaton variable elimination (MCUVE) and successive projections algorithm (SPA) were used for optimizing the modeling variables. It was shown that CWD – SNV method could effectively improve spectral quality, and eliminate translation error. MCUVE – SPA method could greatly improve the precision of model, and compress the modeling variables. The correlation coefficient of prediction (R_p) and root mean square error of prediction (RMSEP) of analytical models were optimized from 0.851 and 0.117 to 0.895 and 0.107, and modeling variables were reduced from 4 148 to 18. NIRS combined with chemometrics could get a better analytical model when amino acid contents exceeded 0.1%. It can provide a fast and simple analytical procedure for the determination of low contents amino acid.

Key words: Tea Amino acid Continuous wavelet derivative Variable selection Near-infrared spectroscopy