

不同预处理方法对多位置种蛋性别鉴定的影响研究

张 伏^{1,2,3}, 崔夏华¹, 丁 轲^{4*}, 张亚坤¹, 王永县¹, 潘孝青⁵

1. 河南科技大学农业装备工程学院, 河南 洛阳 471003
2. 江苏大学现代农业装备与技术教育部重点实验室, 江苏 镇江 212013
3. 机械装备先进制造河南省协同创新中心, 河南 洛阳 471003
4. 河南科技大学动物科技学院, 河南 洛阳 471023
5. 江苏省农业科学院畜牧研究所, 江苏 南京 210014

摘 要 由于猪瘟影响, 鸡蛋作为动物性蛋白的需求大幅增加, 蛋鸡养殖行业也在逐步发展与壮大以满足人们的需求。如何在幼稚阶段甚至胚胎发育阶段能够简单、准确与方便地判断出雌雄, 对蛋鸡养殖行业的发展具有重要意义。对此, 选用 96 个壳色相近、表面无裂纹的新鲜种鸡蛋, 以其可见/近红外漫反射光谱为研究对象, 探讨数据采集位置、光谱预处理方法对种蛋胚胎性别鉴定定性模型的影响。在样本表面钝端、锐端和赤道共三个不同位置分别采集漫反射光谱强度, 经校正后选择(440.27~874.6) nm 作为有效光谱波段进行分析, 并按照 2:1 的比例划分训练集和测试集, 分析了采用归一化(Normalize), 二阶导数(2nd Derivative), 标准正态变量变换(SNV), 多元散射校正(MSC), 去趋势法(Detrend), 光谱转换法(Spectroscopic)共六种方法预处理后的偏最小二乘法判别模型和线性判别分析模型的预测集准确率, 与原始数据(Raw)的预测集准确率进行对比, 得出准确率差异。通过对入孵后 216, 240, 264, 288 和 312 h 不同位置所采光谱数据, 与种蛋性别信息综合分析, 发现胚胎发育 288 h 时的预处理效果最好, 有效提高了 35 个模型的准确率; 264 h 时的预处理效果在分析时间内最差, 其处理降低了 19 个模型的判别准确率; 312 h 的预处理降低了 12 个模型的判别准确率; Detrend 和 Spectroscopic 两种预处理方法可显著提高判别效果, 其中经 Spectroscopic 处理后的模型可能无法进行预测; SNV 和 MSC 对模型具有相同作用效果, Normalize 对模型的作用效果不能确定, 经 2nd Derivative 处理后准确率不确定, 有时与原始数据建模效果一致。综合实验结果发现, 对种鸡蛋胚胎发育 288 h 的数据进行预处理可有效提高模型的判别准确率, 其中对种蛋钝端处数据进行 Detrend 去趋势预处理并利用 LDA 模型分析较好。研究结果为建立种鸡蛋基于可见/近红外的性别信息早期、快速检测模型提供了参考。

关键词 可见/近红外光谱; 漫反射; 种鸡蛋; 预处理方法; 判别位置; 性别鉴定

中图分类号: O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)02-0434-06

引 言

鸡蛋是营养丰富、优质的动物蛋白质来源, 受到猪瘟的影响, 鸡蛋作为动物性蛋白的需求大幅增加, 蛋鸡养殖行业也在逐步发展与壮大以满足人们的需求。如何在幼稚阶段甚至胚胎发育阶段能够简单、准确与方便地判断出雌雄, 对蛋

鸡养殖行业的发展具有重要意义^[1]。目前, 常用的幼稚雌雄判别方法多采用通过相关仪器肉眼识别或性状体现进行, 机械法判别雏鸡形状已经成熟。上述研究方法为雏鸡和鸡胚胎雌雄判别提供了理论基础与方法依据。但在孵化过程中, 种蛋无损、快速鉴定性别的方法依然有待发展。

20 世纪 80 年代后期迅速发展起来的近红外光谱分析技术以其分析速度快、操作简单、检测准确率高等优点已被广

收稿日期: 2020-12-01, 修订日期: 2021-03-20

基金项目: 国家“十三五”重点研发计划项目(2017YFD0301106), 国家自然科学基金项目(52075149), 现代农业装备与技术教育部重点实验室和江苏省农业装备与智能化高技术重点实验室开放基金项目(JNZ201901), 河南省高等学校青年骨干教师培养计划项目(2017GGJS062), 河南省高等教育教学改革研究与实践项目(研究生教育)成果(2019SJGLX063Y)资助

作者简介: 张 伏, 1978 年生, 河南科技大学农业装备工程学院教授 e-mail: zhangfu30@126.com

* 通讯作者 e-mail: keding19@163.com

泛应用于水果、农作物^[2]、药物^[3-4]、茶叶^[5]等检测。受采集环境和样本的影响,采集到的光谱数据包含了各种各样的外界噪声,包括基线漂移、样本不均匀和光散射等。因此,对所获数据进行适当预处理以消除噪声污染对后期模型分析尤为重要。潘庆磊等^[6]搭建高光谱图像采集系统研究孵化 0~12 d 的种蛋胚胎圆头、中间、尖头 3 个部位的感兴趣区域,对原始数据搭建模型进行分析比较;祝志慧等^[7]以紫外-可见光波段为有效波段研究鸡胚雌雄鉴别,其采用多元散射校正对数据进行预处理后,筛选特征波段建立模型;刘燕德等^[8]采用近红外漫反射研究鸡蛋品质,采用多元散射校正、最大-最小归一化、一阶和二阶导数、基线校正、一阶导数+基线校正,消除常数偏移量等预处理方法,发现一阶导数+基线校正的预处理方法较好。

本文在深入研究近红外光谱预处理的基础上,以种鸡蛋为研究对象,为找到适合种鸡蛋性别鉴定的预处理方法,比较了多种预处理方法对偏最小二乘判别和线性判别模型的影响。

1 实验部分

1.1 材料

实验样本为 96 个壳色相近、表面无裂纹的新鲜种鸡蛋,对蛋壳表面清理,采用新洁尔灭溶液消毒后将样本标号,且分别将鸡蛋的钝端、锐端和赤道[见图 1(a,b)]位置标记为 1, 2 和 3,如图 1 所示。采集胚胎发育过程中的种蛋表面漫反射光谱,样本采集后集中放置在温度为 37.8 °C、湿度为 60% 的孵化机中,且孵化机的温度和湿度根据变温孵化法调节。待孵化完成后,对雏鸡翻肛鉴定性别[见图 2(a,b)],对无法准确鉴定结果的雏鸡进行解剖检测,通过雄性睾丸和雌性卵巢的形态变化进行区分,如图 2(a)公鸡性腺,(b)母鸡性腺。

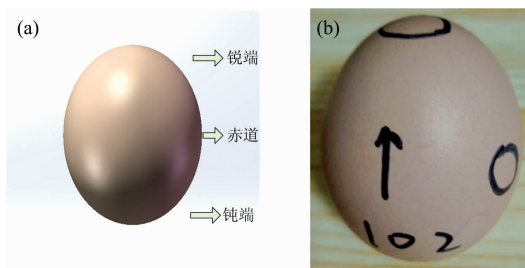


图 1 种鸡蛋测量位置图

(a): 测量位置示意图; (b): 实际测量位置图

Fig. 1 The picture of measurement position

(a): Schematic diagram of measuring position;

(b): Actual measurement position picture

1.2 仪器

实验用可见/近红外光谱检测系统如图 3 所示,系统主要包括一台光源(150 W 卤素灯)、波长范围为 345~1 041 nm 的 USB 4000 光纤光谱仪(Ocean Optics)、装有与光谱仪配套采集软件 Spectrasuite 的计算机和一根 Y 型光纤。采用

便携式光源,其入射光强度可调,计算机与光谱仪间通过 USB 数据线连接。所使用的孵化机型号为 WQ-192(德州市通达孵化设备制造厂,山东德州)。

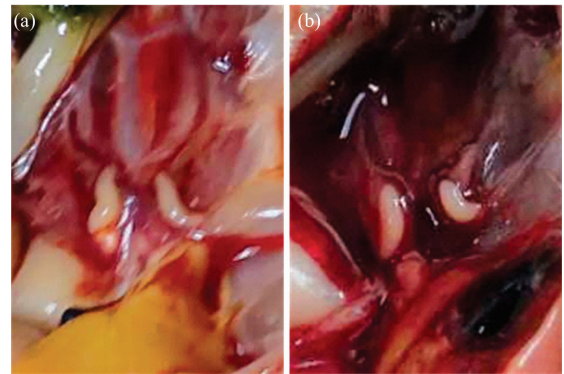


图 2 雏鸡性腺图

(a): 公鸡性腺; (b): 母鸡性腺

Fig. 2 The picture of chicken gonad

(a): The gonad of cocks; (b): The gonad of hens

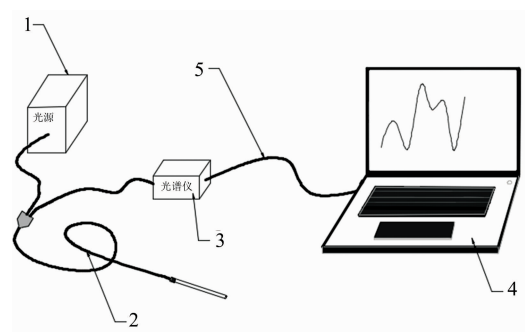


图 3 可见/近红外光谱检测系统

1: 光源; 2: Y 型光纤; 3: 光纤光谱仪;

4: 计算机; 5: 数据线

Fig. 3 The detection system of visible/nearinfrared spectroscopy

1: Light; 2: Y-type fiber; 3: Fiber spectrometer;

4: Computer; 5: USB

1.3 方法

使用检测系统采集 345~1 041 nm 波长范围的种蛋样品蛋壳所标记的 3 个不同位置的漫反射光谱。采集数据过程在黑暗环境下进行,采集时将光纤探头在不损坏样本的前提下紧贴样本表面以避免除入射光外其他光源的影响。采集数据前先将设备预热 30 min,采集软件 Spectrasuite 的积分时间设置为 4 ms,采集样本数据前,将探头紧贴校准白板,调节光源使最大反射率为最大量程的 80%,采集参考光谱,而后采集暗光谱。为消除暗电流的影响,需进行光谱校正,样本反射率的计算公式为

$$R = \frac{I - I_A}{I_C - I_A} \quad (1)$$

式(1)中, R 为所测样本反射率; I 为样本的反射光谱强度; I_C 为参考光谱的反射光谱强度; I_A 为暗光谱的反射光谱强度; (I, I_C, I_A 均无单位)。

从放入孵化机时开始,每隔 24 h 采集一次光谱数据,采集时待光谱曲线稳定后保存所测数据。经光谱校正后,对数据进行预处理,减弱数据中无关信息变量对光谱的影响,提高模型的稳健性和预测能力^[9]。对胚胎发育 216, 240, 264, 288 和 312 h 的数据进行分析,主要是因为发育 216 h 后,胚胎各种器官已包入体腔。

2 结果与讨论

2.1 统计结果

对 96 个种蛋进行观察统计发现,共有 46 个为雄性蛋,50 个母鸡蛋。分析时按照 2 : 1 的比例划分训练集和测试集,本实验将雄性种蛋的类别变量设为 1,雌性种蛋的类别变量设为 -1。

表 1 样本划分结果

Table 1 Samples set division results

Data Set	The number of socks	The number of hens
Training set	32	32
Test set	14	18

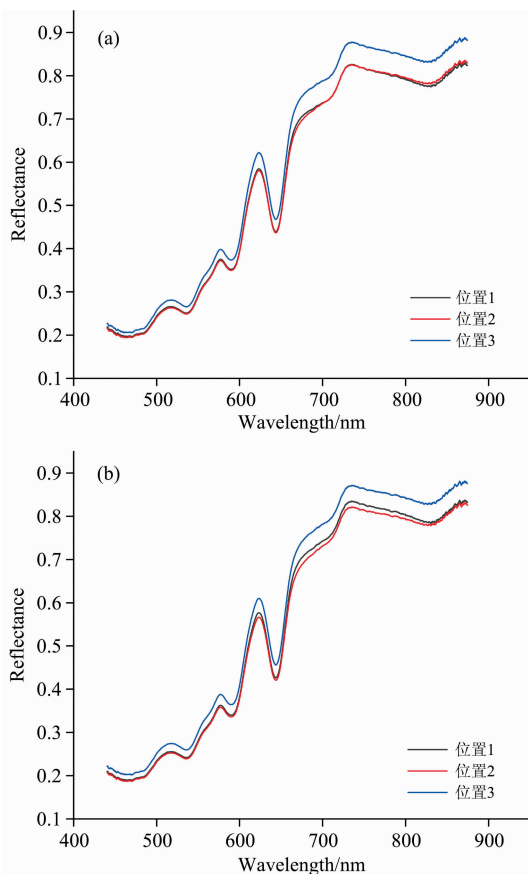


图 4 样本平均原始光谱图

(a): 雄性种蛋平均原始光谱; (b): 雌性种蛋平均原始光谱

Fig. 4 The average raw NIR spectra of samples

(a): Average raw NIR spectra of sock eggs;

(b): Average raw NIR spectra of hen eggs

2.2 数据处理

雄性种蛋和雌性种蛋的平均原始光谱曲线如图 4 所示。由于光谱两端的噪声大,信噪比低且信息少,故选取 440.27 ~ 874.6 nm 中间区域内的光谱曲线进行分析。

2.3 建模方法比较

采用偏小二乘法判别 PLS-DA 和线性判别分析 LDA 模型,使用训练集建模,预测集进行预测,得出预测集判别准确率,求取相应的准确率变化,216, 240, 264, 288 和 312 h 预测集准确率数据见表 2—表 6。

表 2 216 h 不同位置、不同预处理方法的预测集判别准确率变化

Table 2 The variation accuracy of different position and pretreatment method after 216 hours' hatching

Δ Accuracy/%	Position 1		Position 2		Position 3	
	PLS-DA	LDA	PLS-DA	LDA	PLS-DA	LDA
Normalize	-3.03	15.16	-6.06	-6.06	18.19	21.22
2nd Derivative	0	6.06	27.28	15.16	12.13	6.06
SNV	9.09	27.28	3.03	-9.09	18.19	21.22
MSC	9.09	27.28	0	-9.09	18.19	21.22
Detrend	0	21.22	6.06	3.03	3.03	9.09
Spectroscopic	0	18.19	9.09	24.25	6.06	-

表 3 240 h 不同位置、不同预处理方法的预测集判别准确率变化

Table 3 The variation accuracy of different position and pretreatment method after 240 hours' hatching

Δ Accuracy/%	Position 1		Position 2		Position 3	
	PLS-DA	LDA	PLS-DA	LDA	PLS-DA	LDA
Normalize	21.87	40.62	-18.75	-34.37	-3.13	-3.12
2nd Derivative	-9.38	15.62	-3.13	-25.00	-15.63	0
SNV	18.75	34.37	6.25	3.13	0	3.13
MSC	18.75	34.37	6.25	3.13	0	3.13
Detrend	12.50	34.37	-12.50	-25.00	-6.25	0
Spectroscopic	9.37	25.00	-3.13	-9.37	-9.38	-12.50

表 4 264 h 不同位置、不同预处理方法的预测集判别准确率变化

Table 4 The variation accuracy of different position and pretreatment method after 264 hours' hatching

Δ Accuracy/%	Position 1		Position 2		Position 3	
	PLS-DA	LDA	PLS-DA	LDA	PLS-DA	LDA
Normalize	0	-3.12	-25.00	-37.50	12.50	9.38
2nd Derivative	6.25	-12.50	-6.25	-6.25	15.62	9.38
SNV	3.13	-3.12	-12.50	-31.25	15.62	12.50
MSC	3.13	-3.12	28.13	-31.25	15.62	12.50
Detrend	3.13	-3.12	-18.75	-31.25	9.37	-3.12
Spectroscopic	-3.12	-21.87	-6.25	9.37	-3.13	-

216~312 h 内的准确率变化情况如图 5 所示,图中黄色柱形表示 PLS-DA 模型的准确率变化,红色柱形表示 LDA 模型的准确率变化,短斜线分别表示无法预测胚胎发育 216,

表 5 288 h 不同位置、不同预处理方法的
预测集判别准确率变化

Table 5 The variation accuracy of different position and pretreatment method after 288 hours' hatching

ΔAccuracy/%	Position 1		Position 2		Position 3	
	PLS-DA	LDA	PLS-DA	LDA	PLS-DA	LDA
Normalize	28.13	18.75	6.25	-3.13	15.62	25.00
2nd Derivative	21.88	18.75	15.62	15.62	21.87	21.87
SNV	28.13	28.13	21.87	3.12	25.00	25.00
MSC	28.13	28.13	21.87	3.12	25.00	25.00
Detrend	25.00	25.00	34.37	43.75	12.50	28.12
Spectroscopic	18.75	21.88	37.50	34.37	21.87	28.12

表 6 312 h 不同位置、不同预处理方法的
预测集判别准确率变化

Table 6 The variation accuracy of different position and pretreatment method after 312 hours' hatching

ΔAccuracy/%	Position 1		Position 2		Position 3	
	PLS-DA	LDA	PLS-DA	LDA	PLS-DA	LDA
Normalize	3.12	6.25	-28.12	-31.25	12.50	12.50
2nd Derivative	15.62	0	-12.50	-12.50	21.87	9.37
SNV	25.00	15.63	-21.87	-28.13	9.37	3.12
MSC	25.00	15.63	-21.87	-28.13	9.37	3.12
Detrend	15.62	15.63	-15.62	-21.88	25.00	18.75
Spectroscopic	15.62	15.63	-12.50	-21.88	18.75	-

264 和 312 h 所采数据经 Spectroscopic 预处理后所建模型的判别准确率。由图 5 知，胚胎发育 288 h 时，预处理可提高 35 个模型的准确率，降低 1 个模型的准确率；264 h 的处理

效果在分析时间内最差，降低 19 个模型的判别准确率；312 h 的处理效果次之，预处理降低 12 个模型的判别准确率。

Normalize 是对数据进行归一化处理，将其限制在一定范围内，由图 5 可见其对模型判别准确率影响不确定。SNV 和 MSC 均是消除漫反射数据采集过程中光谱差异的方法，本研究中分析比较不同时间的 SNV 和 MSC 预处理，发现除胚胎发育 264 h 时位置 2 处的 PLS-DA 模型外，两者对模型判别的准确率的影响相同，由于存在实验误差，可认为 SNV 和 MSC 预处理对不同时间、不同位置的模型有相同作用。导数法虽可有效消除基线和其他背景的干扰，但其引入噪声，降低信噪比。本研究中采用二阶导数法对数据进行探索，发现其可能对模型无作用效果，判别准确率与原始数据相比没有明显差异。比较胚胎发育 288 h 时的预处理影响，发现 Detrend 和 Spectroscopic 两种预处理方法可显著提高判别效果，其中 Detrend 是一种去趋势的方法，可消除获取数据时产生偏移的影响，而 Spectroscopic 是将反射率单位转换为吸光度的一种方法，对比两种预处理方法在不同位置的处理效果可知，其对位置 2 即钝端处的效果更优。但经 Spectroscopic 预处理后可能无法获取模型的预测集预测准确率，所以 Detrend 是一种较优的预处理方法。由表 5 见，此时 PLS-DA 模型的变化判别准确率为 34.37%，LDA 模型的变化判别准确率为 43.75%，故利用 LDA 模型对种蛋性别鉴定易获得较佳的判别效果。

综上，对胚胎发育 288 h 的数据进行预处理可提高模型的判别准确率，其中采集位置 2 即种蛋钝端处数据后，进行 Detrend 去趋势预处理，利用 LDA 模型分析种蛋性别效果较好。

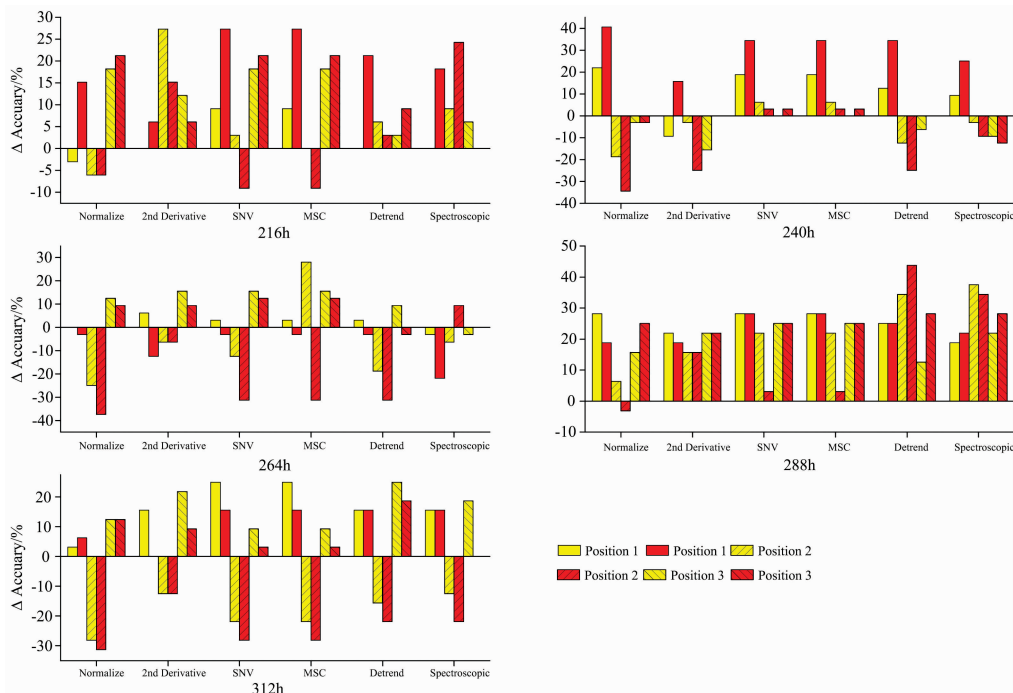


图 5 216~312 h 内的准确率变化

Fig. 5 The variation accuracy of 216~312 hours' hatching

3 结 论

(1)研究了基于可见/近红外光谱的受精蛋的性别判别位置和判别时间,通过对光谱数据的截取,最终确定 440.27~874.6 nm 波段的光谱为有效光谱,通过对该段光谱的研究分析可以对种鸡蛋进行有效研究。

(2)研究了不同时间、不同预处理两种模型的预测集准确率变化,发现胚胎发育 288 h 时,预处理可有效提高 35 个模型的准确率;264 h 的处理降低了 19 个模型的判别准确率;312 h 的预处理降低了 12 个模型的判别准确率。288 h 时的预处理效果较好,264 h 时的预处理效果分析时间内最差。

(3)研究了不同预处理方式间的差别及其对所建模型的预测效果影响。发现 Normalize 对模型判别准确率影响不确

定;SNV 和 MSC 预处理对不同时间、不同位置的模型有相同作用效果;二阶导数法可能对模型无作用效果,判别准确率与原始数据相比没有发生变化;Detrend 和 Spectroscopic 两种预处理方法可显著提高判别效果,且对位置 2 即钝端处的效果更优,但经 Spectroscopic 预处理后可能无法获取模型的预测集准确率,所以 Detrend 是一种较优的预处理方法。

(4)比较了种蛋孵化 288 h 时,位置 2 即钝端处的模型预测效果。此时,PLS-DA 模型的变化判别准确率为 34.37%,LDA 模型的变化判别准确率为 43.75%,LDA 模型比 PLS-DA 模型的变化判别准确率高 9.38%,故利用 LDA 模型对种蛋性别鉴定易获得较佳的判别效果。

由于近红外光谱分析技术具有快速、对样本无损等特点,该研究对近红外光谱技术在孵化中期的性别鉴定时的数据采集位置和数据预处理方法进行了比较,提供了一种新的思路。

References

- [1] YEZI Qiao-jing, YANG Ya-ru, ZHANG Ning-ning, et al(叶子巧婧,杨雅茹,张宁宁,等). Heilongjiang Animal Science and Veterinary Medicine(黑龙江畜牧兽医), 2014, (19): 62.
- [2] MENG Qing-yan, HE Jian-guo, LIU Gui-shan, et al(孟庆琰,何建国,刘贵珊,等). Food Science and Technology(食品科技), 2015, 40(3): 287.
- [3] YANG Yue, YANG Liu-chang, JI Xiao-liang, et al(杨越,杨留长,纪晓亮,等). Journal of Instrumental Analysis(分析测试学报), 2020, 39(11): 1311.
- [4] WAN Na, LIN Huan-yu, WU Zhen-feng, et al(万娜,林环玉,伍振峰,等). Chinese Traditional and Herbal Drugs(中草药), 2020, 51(17): 4425.
- [5] WANG Sheng-peng, TENG Jing, ZHENG Peng-cheng, et al(王胜鹏,滕靖,郑鹏程,等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2020, 36(8): 271.
- [6] PAN Lei-qing, ZHANG Wei, YU Min-li, et al(潘磊庆,张伟,于敏莉,等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2016, 32(1): 181.
- [7] ZHU Zhi-hui, HONG Qi, WU Lin-feng, et al(祝志慧,洪琪,吴林峰,等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2019, 39(9): 2780.
- [8] LIU Yan-de, ZHOU Yan-rui, PENG Yan-yin(刘燕德,周延睿,彭彦颖). Optics and Precision Engineering(光学精密工程), 2013, 21(1): 40.
- [9] CHU Xiao-li, YUAN Hong-fu, LU Wan-zhen(褚小立,袁洪福,陆婉珍). Progress in Chemistry(化学进展), 2004, 16(4): 528.

Study on the Influence of Different Pretreatment Methods on Gender Determination of Multiposition

ZHANG Fu^{1, 2, 3}, CUI Xia-hua¹, DING Ke^{4*}, ZHANG Ya-kun¹, WANG Yong-xian¹, PAN Xiao-qing⁵

1. College of Agricultural Equipment Engineering, Henan University of Science and Technology, Luoyang 471003, China
2. Key Laboratory of Modern Agricultural Equipment and Technology, Ministry of Education, Jiangsu University, Zhenjiang 212013, China
3. Collaborative Innovation Center of Machinery Equipment Advanced Manufacturing of Henan Province, Luoyang 471003, China
4. College of Animal Science and Technology, Henan University of Science and Technology, Luoyang 471023, China
5. Institute of Animal Husbandry, Jiangsu Academy of Agricultural Sciences, Nanjing 210014, China

Abstract Due to the impact of swine fever, the demand for eggs which is an important substitute for pork, has increased significantly, and the laying hens breeding industry has also gradually developed and expanded to meet people's demands. Therefore, it is of great significance for the development of layer breeding industry that how to judge gender at the stage of chick and even embryo development accurately and conveniently. To this, 96 fresh seed eggs with similar shell color and no cracks on the surface were selected, and the visible/near-infrared diffuse reflection spectrum was used as the research object, investigated the influence of data collection location and spectral pretreatment method on the qualitative model of gender identification of seed eggs. Diffuse reflectance spectral intensity was collected at three different positions on the surface: blunt end, sharp end and the equator. After correction, 440.27~874.6 nm was selected as the effective spectral band for analysis. The spectral intensity was calculated according to 2:1 divided into a training set and test set, the proportion of the normalized (Normalize), the second Derivative (2nd Derivative), standard normal variable transformation (SNV), multiple scatter correction (MSC), to trend method (Detrend), spectral transformation method (Spectroscopic), a total of six kinds of pretreatment of PLS-DA model and LDA models' prediction accuracy were analyzed, then compared with the original data (Raw) prediction accuracy, the changes of accuracy were obtained. Through comprehensive analysis of spectral data collected 216, 240, 264, 288 and 312 h after incubation and egg gender information at different positions, it was found that the pretreatment effect was the best at 288 h after embryo development, and the accuracy of 35 models was effectively improved. The pretreatment effect at 264 h was the worst in the analysis time, and its treatment reduced the accuracy of 19 models. The pretreatment of 312 h reduced the discriminant accuracy of 12 models. Detrend and Spectroscopic, two kinds of pretreatment method, could significantly improve the effect of discrimination, but the Spectroscopic model may not be able to predict; SNV and MSC had the same effect on the model, Normalize's effect on the model could not be determined. The accuracy of 2nd Derivative treatment was uncertain, which is sometimes consistent with the effect of original data modeling. The comprehensive experimental results showed that the preprocessing used LDA model of 288 h embryo development data could effectively improve the discriminant accuracy of the model, among which the Detrend preprocessing of the data at the blunt end of the egg was good. The results provided a reference for establishing an early and rapid detection model based on visible/near-infrared gender information in egg species.

Keywords Visible/near infrared spectroscopy; Diffuse reflection; Hatching egg; Pretreatment method; Judgement position; Gender Determination

(Received Dec. 1, 2020; accepted Mar. 20, 2021)

* Corresponding author