

基于紫外-可见透射光谱技术和极限学习机的早期鸡胚雌雄识别

祝志慧^{1,2}, 洪琪^{1,2}, 吴林峰^{1,2}, 王巧华^{1,2}, 马美湖^{3*}

1. 华中农业大学工学院, 湖北 武汉 430070
2. 农业部长江中下游农业装备重点实验室, 湖北 武汉 430070
3. 华中农业大学食品科学技术学院, 湖北 武汉 430070

摘要 为了对鸡种蛋胚胎进行雌雄识别, 探究利用紫外-可见-近红外透射光谱进行鸡胚雌雄识别的可行性, 搭建了鸡种蛋透射光谱检测系统, 采用横向和竖向大头朝上2种放置方式获取210枚鸡种蛋孵化0~15 d的光谱, 光谱范围为360~1 000 nm。构建极限学习机(ELM)鸡胚雌雄识别模型, 通过比较不同放置方式和孵化天数下模型的识别准确率, 发现竖向放置且孵化第7 d的识别效果最好; 将竖向放置孵化第7 d的光谱初步分为紫外(360~380 nm)、可见光(380~780 nm)、近红外(780~1 000 nm)、紫外-可见光(360~780 nm)和全波段(360~1 000 nm)5个不同的波段范围来分析, 预测集准确率分别为82.86%, 77.14%, 75.71%, 84.29%和81.43%, 筛选出360~780 nm的紫外-可见光波段为有效波段; 在紫外-可见光(360~780 nm)波段, 采用多元散射校正(MSC)去噪, 并用竞争性自适应重加权采样算法(CARS)和连续投影算法(SPA)筛选特征波长降维, 建立不经筛选特征波长、CARS筛选特征波长和SPA筛选特征波长的3种ELM模型。其中不经筛选特征波长的ELM模型识别效果最好, 但输入变量最多, 隐含层神经元为680且激活函数为sig时, 预测集准确率为84.29%。SPA筛选特征波长的ELM模型识别效果次之, 输入变量有9个, 隐含层神经元为840且激活函数为hardlim时, 预测集准确率为81.43%。CARS筛选特征波长的ELM模型识别效果最差, 输入变量有27个, 隐含层神经元为100且激活函数为sig时, 预测集准确率为78.57%; 用遗传算法(GA)优化ELM模型的权值变量和隐含层阈值, 不经筛选特征波长建立的GA-ELM模型, 预测集准确率为87.14%, SPA筛选特征波长建立的GA-ELM模型, 预测集准确率为87.14%, CARS筛选特征波长建立的GA-ELM模型, 预测集准确率为81.43%。紫外-可见光波段不经筛选特征波长的GA-ELM模型识别效果和经SPA筛选特征波长的GA-ELM模型相同, 表明SPA筛选的特征波长变量能够有效反映360~780 nm波段的信息, SPA使用的变量数仅占紫外-可见光波段的2.14%, 因此, 雌雄识别最佳模型为紫外-可见光波段经SPA筛选特征波长的GA-ELM模型, 预测集准确率为87.14%, 其中, 雌性识别率为88.57%, 雄性识别率为85.71%, 单个样本平均判别时间0.080 ms。结果表明紫外-可见透射光谱技术和ELM模型为孵化早期鸡胚蛋雌雄识别提供了一种可行方法。

关键词 种蛋; 鸡胚; 雌雄; 识别; 紫外-可见-近红外光谱; 极限学习机

中图分类号: O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)09-2780-08

引言

在家禽繁育中, 鸡蛋孵化是一个耗时、耗能的过程, 大约需要21 d^[1]。鸡种蛋胚胎在自然情况下孵化, 雌雄比例大致为1:1, 但不同行业对雌雏和雄雏的需求量各不相同。在蛋鸡育种行业中, 雌性的需求量远大于雄性; 而在肉品行业

中, 因为雄性的发育周期短、饲料利用率高, 雄性的需求量相比于雌性要占极大优势。在生产实际中, 每年约有70亿只一日龄幼雏被安乐死^[2]。因此, 对鸡种蛋胚胎进行性别鉴定可以指导实际生产, 按照市场需求调整鸡种蛋孵化比例, 不仅可以节省设备和能源, 减少孵化成本, 还能避免社会伦理道德等问题。

在国内外已有的报道中, 鸡种蛋的雌雄识别主要有基于

收稿日期: 2018-07-31, 修订日期: 2018-12-02

基金项目: 中央高校基本科研业务费(2662017PY057), 公益性行业(农业)科研专项(201303084)资助

作者简介: 祝志慧, 1975年生, 华中农业大学工学院副教授 e-mail: zzh@mail.hzau.edu.cn

* 通讯联系人 e-mail: mameihuhn@163.com

内容物的破口检测和基于图像或高光谱的无损检测。Weissmann 等^[3]在孵化的第 9 d, 通过酶免疫测定分析来自尿囊液中的硫酸雌酮(E1S)含量进行鸡胚性别鉴定, 准确率达 90.70%。Roberta Galli 等^[4-6]在鸡种蛋钝端开一个小口, 用近红外激光照射获得强荧光和弱拉曼信号, 将血液中的血红蛋白作为荧光源, 通过 PC 分数的监督分类, 可以确定孵化 3.5 d 时的鸡胚性别, 正确率高达 93%。基于内容物的检测属于破坏性检测, 对后期鸡胚孵化有影响, 且操作步骤繁多, 不适用于工业生产。祝志慧等^[7]利用图像处理技术提取鸡种蛋血线特征, 构建深度置信网络识别鸡胚性别, 准确率达 83.33%。潘磊庆等^[1]采集鸡种蛋高光谱透射图像, 获取感兴趣区域的响应信号构建鸡胚性别鉴定模型, 发现孵化第 10 d 种蛋中间部位具有最好的识别效果, 准确率为 82.86%; Doreen Gohler 等^[8]利用特殊的鸡品种呈现羽毛色的性二型现象, 用高光谱相机收集 400~1 000 nm 光谱范围内的透射光, 采用主成分分析和线性判别分析确定孵化 14 d 的鸡胚性别, 准确率约为 97%。然而, 高光谱设备昂贵, 数据量大, 不适合工业生产。

紫外-可见-近红外透射光谱是一种非破坏性的检测方法。光谱能够透过鸡蛋外壳表面, 被内部物质吸收, 并能反映在光谱图中。因此, 本文探究利用紫外-可见-近红外透射光谱进行鸡胚雌雄识别的可行性。

1 实验部分

1.1 材料

材料为白壳京粉 1 号蛋, 购于荆州峪口禽业有限公司, 共 210 枚, 其中雌雄各 105 枚。鸡种蛋在 37 °C、5% 新洁尔灭原溶液中浸泡 3 min 进行消毒处理、编号, 静置干燥后放入孵化箱进行孵化, 其中孵化温度和湿度分别为 37.8 °C 和 63.5%。

1.2 光谱检测系统

鸡种蛋透射光谱检测装置如图 1 所示。硬件主要由海洋光学 USB2000+ 型光纤光谱仪、海洋光学 84UV-25 准直透镜、南京鸿照聚焦透镜、玻璃光纤、广州标旗 LS-3000 可调卤素灯光源(100 W)和计算机构成。胚胎放置方式分为横向放置和竖向大头朝上放置 2 种, 采集孵化第 0~15 d 的光谱。

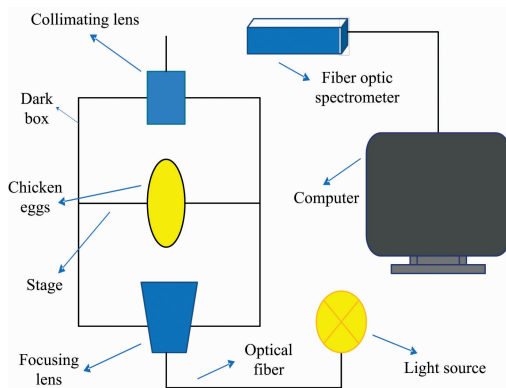


图 1 鸡种蛋透射光谱检测系统

Fig. 1 Chicken egg transmission spectrum detection system

光谱采集波段设置为 360~1 000 nm, 积分时间为 40 ms, 平均次数为 3 以提高数据的稳定性, 平滑宽度为 5 以匹配系统的分辨率。

1.3 雌雄胚胎的物理判别

胚胎发育的 4~8 d 性腺开始转化成早期卵巢和睾丸, 在胚胎发育至 8 d 后, 可以通过睾丸和卵巢的形态进行性别区分, 雌性卵巢两侧发育不平衡, 雄性睾丸两侧发育对称^[9]。为准确获得鸡种蛋胚胎的性别信息, 为后期鉴别模型提供判别依据, 在鸡种蛋孵化的第 15 d, 人工破蛋用解剖法判断胚胎性别, 通过肉眼观察雄性睾丸和雌性卵巢的形态进行区分, 作为鸡种蛋实际性别。图 2 为孵化第 15 d 鸡种蛋胚胎解剖图, 图 2(a)为雌性, 图 2(b)为雄性。

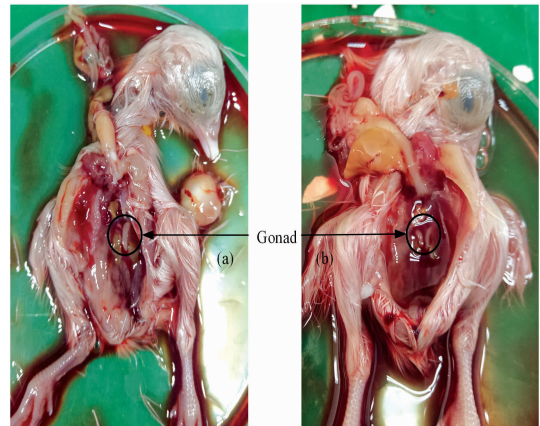


图 2 孵化第 15 d 雌性(a)和雄性(b)鸡种蛋胚胎解剖图

Fig. 2 Anatomical map of female (a) and male (b) chicken embryos on the 15th day of hatching

1.4 数据处理

为避免训练集样本分布不均匀, 采用 Kennard-Stone 法将光谱差异较大的样本选入训练集, 保证有代表性的样本全部进入训练集^[10]。本试验总计 210 枚样本, 按 2:1 划分训练集和预测集, 训练样本 140 个(雌雄各 70 个), 预测样本 70 个(雌雄各 35 个)。

1.4.1 光谱范围划分

为了寻找适合雌雄鉴别的光谱波段, 将 360~1 000 nm 范围的光谱分为紫外波段(360~380 nm)、可见光波段(380~780 nm)、近红外波段(780~1 000 nm)、紫外-可见光波段(360~780 nm)和全波段(360~1 000 nm)5 个不同光谱范围来分析, 去除部分冗余信息, 对光谱进行初步降维。

1.4.2 光谱预处理

采集到的鸡胚光谱数据夹杂仪器操作、样品背景和杂散光等引起的噪声和无关信息。本文采用的预处理方法为多元散射校正(multiplication scatter correction, MSC)。MSC 可以有效消除或减小由于样品粒径分布不均匀造成的杂散光问题, 降低光谱数据的噪声^[10]。

1.4.3 光谱数据降维算法

采集的光谱波段不论是可见光波段(380~780 nm)、近红外波段(780~1 000 nm)、紫外-可见光波段(360~780 nm), 还是全波段(360~1 000 nm), 光谱范围都较宽波长数

目也较多,部分波段存在较强的相关性,光谱波段内除了包含反映内部成分差异的信息外还含有大量冗余信息,分类时可能会产生维数灾难现象,且会影响分类的精度。因此,选用 CARS 和 SPA 这 2 种常见方法对初选波段的光谱数据进行筛选,选出反映雌雄差异的特征波长,优化模型的输入变量。

(1) 竞争性自适应重加权采样算法(competitive adaptive reweighted sampling, CARS)采用“适者生存”原则,对光谱数据进行多次筛选将 PLS 模型中回归系数绝对值大的波长点选出,通过对比每次筛选变量产生的新子集的交叉验证均方差值(RMSECV),RMSECV 值最小的那个模型所对应的波长点被选为最优变量。该算法可以克服变量选择中变量组合爆炸问题^[11]。

(2) 连续投影算法(successive projection algorithm, SPA)是一种前向选择算法,初选一个变量,然后迭代入新的变量,通过设置不同的变量范围数,根据交叉验证均方根误差 RMSECV 值,选出预先设定好的变量数目。该算法能从全部变量中选取含有最少冗余信息的变量组,保证更少变量数所包含的样本信息最全,可以有效解决信息重叠和共线性的问题^[12]。

1.4.4 定性判别建模算法

极限学习机(extreme learning machine, ELM)是常用的定性判别建模方法。极限学习机是一种以单隐层前馈神经网络为基础的新算法,具有学习速度快、泛化能力强等优点,解决了传统神经网络参数选取复杂、易陷入局部最优的缺点,在模式识别和非线性拟合方面具有明显优势^[13]。ELM 中随机产生输入层与隐含层的连接权值和隐含层神经元的阈值且无需调整,只需设置隐含层神经元个数和隐含层神经元的激活函数。本文在 3 种不同激活函数 sig, sine 和 hardlim 下,让隐含层神经元个数从 20 到 1 500 以 20 步长进行寻优,寻找测试误差最低时的隐含层神经元个数和激活函数。

2 结果与讨论

2.1 鸡种蛋光谱分析、放置方式和最佳鉴别天数的确定

2.1.1 不同孵化天数鸡胚的光谱曲线变化情况

鸡胚原始平均光谱随孵化天数的变化如图 3。可以看到,无论是横向放置还是竖向放置,鸡胚光谱曲线均在前端存在很强的吸收峰,在前端光谱值的大小随天数的变化没有规律。在 500~780 nm 波段随着孵化天数的增加,波形没有太大变化,只是透过率逐渐降低。说明孵化过程中,鸡种蛋内部吸收峰在 500~780 nm 波段的某些物质转化成了鸡胚发育所需的其他物质,从而浓度逐渐降低导致光谱响应值降低。

2.1.2 性别鉴定的最佳放置方式和鉴别天数的确定

为确定鸡种蛋胚胎性别鉴定的最佳放置方式和鉴别天数,采用横向和竖向大头朝上 2 种鸡胚放置方式,将采集到的 0~15 d 的全波段原始光谱数据建立 ELM 模型,结果见图 4。准确率为预测集准确率。图 4 表明以横向放置的方式采集的光谱数据,在鸡种蛋胚胎孵化过程中,建立的 ELM

模型最高的识别准确率在第 1, 9 和 13 d, 预测集准确率为 75.71%。以竖向大头朝上放置的方式采集的光谱数据,在鸡种蛋胚胎孵化过程中,建立的 ELM 模型最高的识别准确率在第 7 d, 预测集准确率为 81.43%。

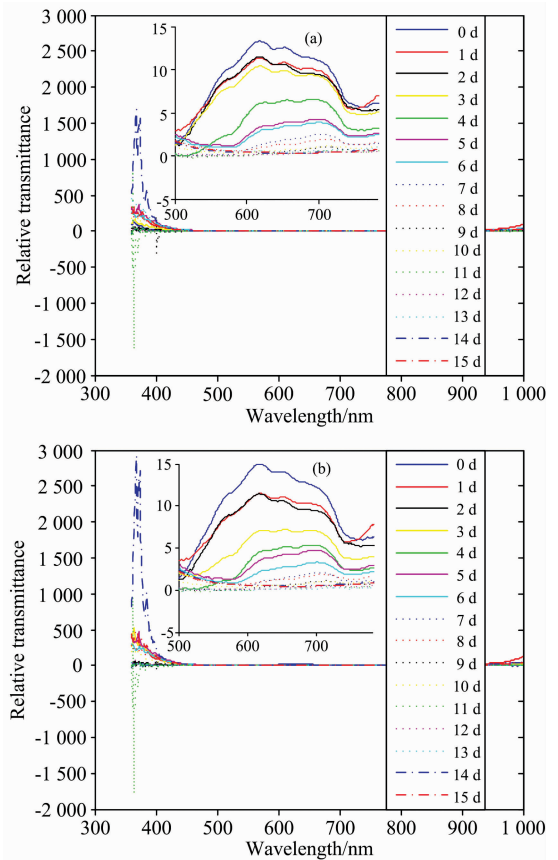


图 3 横向放置(a)和竖向放置(b)鸡胚光谱随孵化天数变化

Fig. 3 Chicken embryo spectrum changes with hatching days under horizontal placement (a) and vertical placement (b)

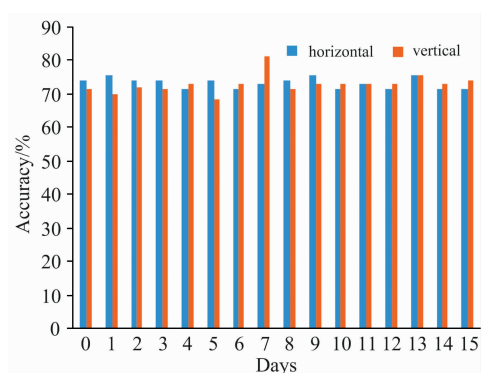


图 4 预测集准确率随孵化天数增加的变化
Fig. 4 Predictive set accuracy changes with increasing hatching days

从上述结果可以发现,竖向放置且孵化第 7 d 更容易区分鸡种蛋胚胎的性别。因此,选择竖向放置且孵化第 7 d 的

光谱数据进行重点分析。

2.2 竖向放置孵化第 7 d 鸡胚识别模型的建立

2.2.1 雌雄鸡胚光谱特性

鸡种蛋竖向放置孵化第 7 d 采集到的雌雄鸡胚原始平均光谱如图 5。可以看出雌雄鸡胚原始平均光谱曲线随波长变化的规律大体一致，只是存在一些数值上的细小差异，很难通过直接观察来区分雌雄。此外，雌雄鸡胚光谱在紫外-可见光波段(360~780 nm)比在近红外波段(780~1 000 nm)内含有更丰富的波峰波谷信息。

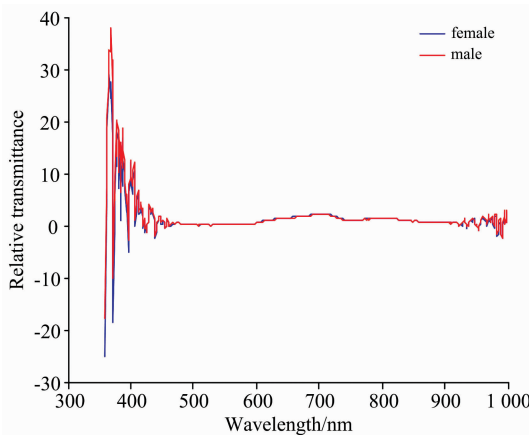


图 5 孵化第 7 d 雌雄鸡胚平均光谱

Fig. 5 Average spectrum of male and female chicken embryos on the 7th day of hatching

2.2.2 波段初选

为了寻找适合雌雄鉴别的光谱波段，将 360~1 000 nm 范围的光谱分为紫外波段(360~380 nm)、可见光波段(380~780 nm)、近红外波段(780~1 000 nm)、紫外-可见光波段(360~780 nm)和全波段(360~1 000 nm)5 个不同光谱范围并建模，建模结果如表 1。在 5 个波段范围中，近红外波段的识别率最低；紫外波段和紫外-可见光波段相比于全波段，模型预测集准确率均有所提升，其中紫外-可见光波段的模型识别率最高为 84.29%。因此，选择波段范围为 360~780 nm 的紫外-可见光来对种蛋进行雌雄鉴别。

表 1 不同波段范围下 ELM 模型判别结果

Table 1 Discriminant results of ELM models in different band ranges

波段范围/nm	训练集准确率/%	预测集准确率/%
360~380	100	82.86
380~780	100	77.14
780~1 000	100	75.71
360~780	100	84.29
360~1 000	100	81.43

2.2.3 光谱预处理

采用多元散射校正(MSC)对紫外-可见光(360~780 nm)波段的光谱数据进行降噪处理，图 6 为原始光谱与 MSC 预处理后的光谱图。可以看到 MSC 在保留原始光谱主要信息

的同时，有效地消除了因纵向漂移产生的光谱差异。

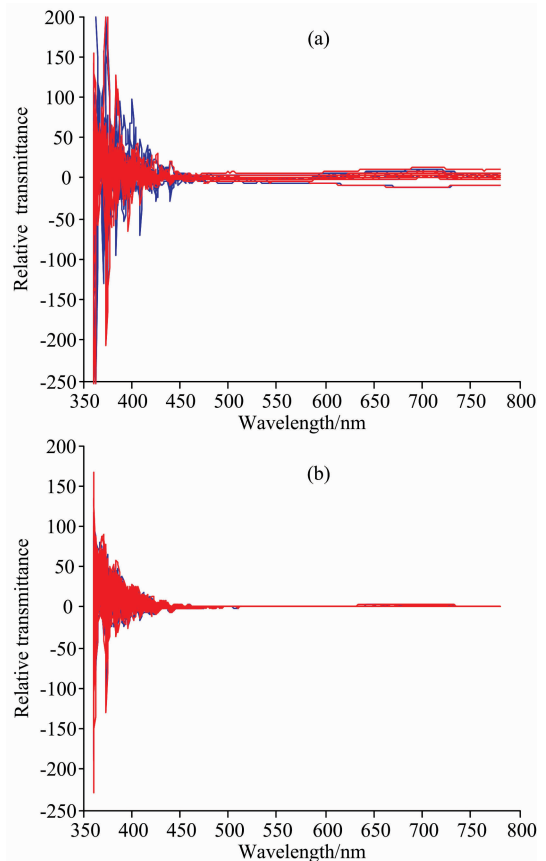


图 6 全部雌雄鸡胚原始(a)和 MSC 预处理后(b)的光谱

Fig. 6 Original (a) and MSC pretreated (b) spectra of all male and female chicken embryos

2.2.4 光谱数据降维处理

选择的紫外-可见光波段(360~780 nm)，仍然包含很多无用信息，需要找出反映雌雄差异的特征波长点。选用 CARS 和 SPA 这 2 种常见方法对光谱数据进行特征变量筛选。

(1)利用竞争性自适应重加权采样算法(CARS)进行采样时，采样次数设定为 50 次。第一次所有变量都用于建模，第 N 次时只使用两个变量。图 7 为波段筛选过程，其中横坐标均为采样次数。从图 7(a)可以看到选用的变量个数先急剧减小后缓慢趋于稳定，这是指数衰减函数的作用；图 7(b)中随着采样次数的增加交叉验证均方差(RMSECV)逐渐减小后增大，在第 26 次采样时达到最小值，这表明第 26 次采样后剔除的波长点与鸡胚性别鉴定的相关性大；图 7(c)中各线是随着采样次数的增加回归系数的变化，图中的“*”线为 RMSECV 最小值处。利用 CARS 算法对紫外-可见光波段的光谱数据进行筛选，选出的波长有 27 个。图 9 表示 CARS 所筛选的特征波长在紫外-可见光波段的分布状况，占紫外-可见光波段的 6.43%。

(2)连续投影算法(SPA)通过设置波长数的范围 1~50，得到 9 个特征波长变量，筛选过程如图 8。从图 8 可以看到 RMSECV 随变量数目增加大体呈现逐渐减小的趋势，并在 9

个波长时 RMSECV 最低为 0.447 73, 说明 9 个特征波长包含了雌性识别的更多有效信息。利用 SPA 算法对紫外-可见光波段的光谱数据进行筛选, 图 9 表示所筛选的特征波长在紫外-可见光波段中的分布状况, 占紫外-可见光波段的 2.14%。

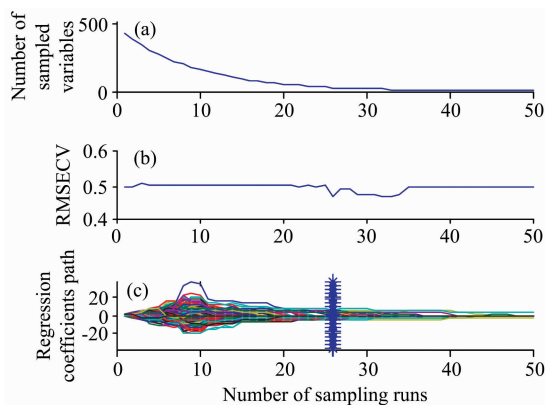


图 7 CARS 变量筛选过程

Fig. 7 CARS variable screening process

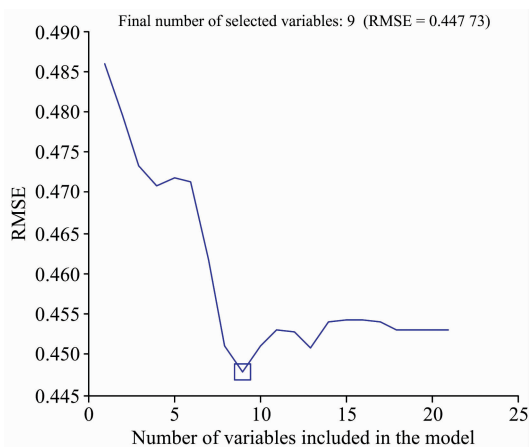


图 8 SPA 变量筛选过程

Fig. 8 SPA variable screening process

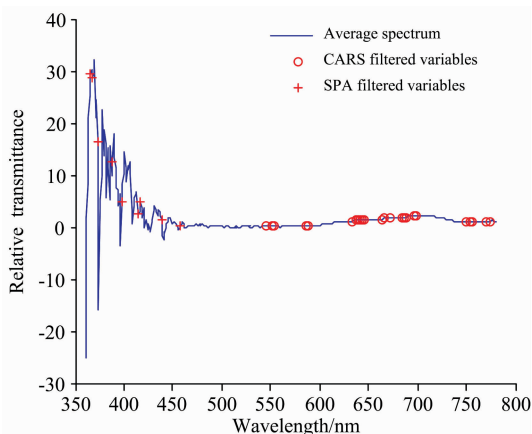


图 9 CARS 和 SPA 筛选的变量

Fig. 9 CARS and SPA filtered variables

2.2.5 ELM 定性判别分析模型

在紫外-可见光波段(360~780 nm)经过 MSC 预处理后, 运用极限学习机(ELM)分别建立不经筛选特征波长的判别模型、基于 CARS 筛选特征波长的判别模型和基于 SPA 筛选特征波长的判别模型。极限学习机是一种以单隐含层前馈神经网络为基础的新算法, 输入层与隐含层的连接权值和隐含层神经元的阈值随机产生, 只需设置隐含层神经元个数和

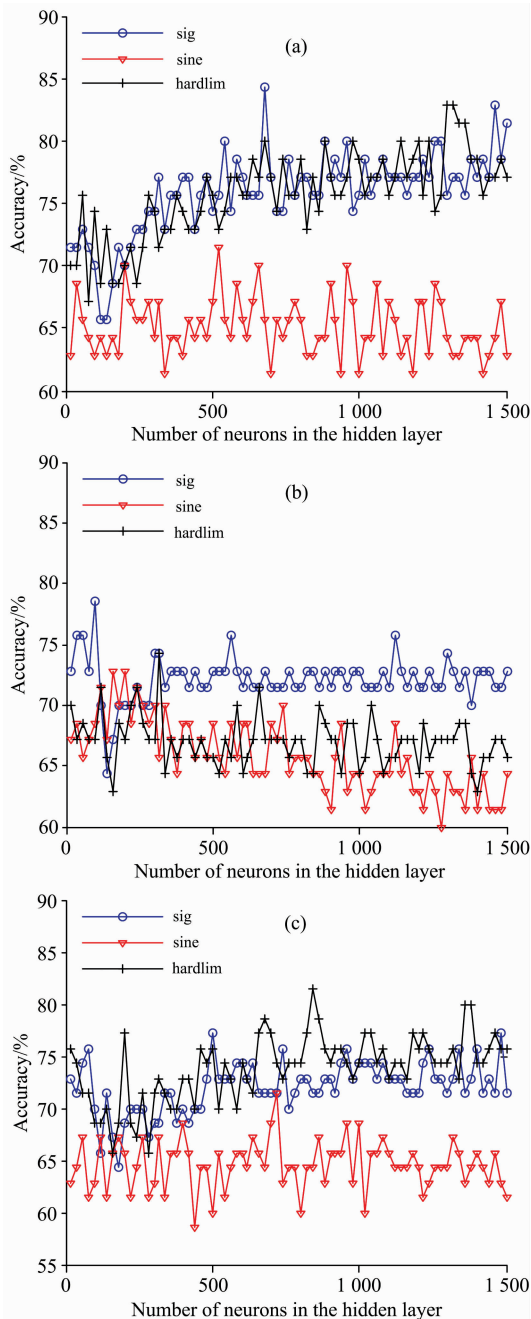


图 10 不筛选(a)、CARS 筛选(b)和 SPA 筛选(c) 隐含层神经元数和激活函数对 ELM 模型性能的影响

Fig. 10 Non-screening (a), CARS screening (b) and SPA screening (c) Influence of number of neurons in the hidden layer and activation function on the performance of the ELM model

隐含层神经元的激活函数。为确定模型的最佳隐含层神经元个数和隐含层神经元激活函数，本文在 3 种不同激活函数下，让隐含层神经元个数从 20 到 1 500 以 20 步长进行寻优，图 10 为 3 种判别模型的寻优过程。

图 10(a)显示了不经筛选特征波长的 ELM 模型，隐含层神经元个数为 680 时，以 sig 作为激活函数预测集准确率最高。因此在紫外-可见光波段不经筛选特征波长的光谱数据建立的 ELM 模型中，选用 sig 作为激活函数且隐含层神经元个数设为 680。图 10(b)显示了基于 CARS 筛选特征波长的 ELM 模型，隐含层神经元个数为 100 时，以 sig 作为激活函数预测集准确率最高。因此在紫外-可见光波段经 CARS 筛选特征波长的光谱数据建立的 ELM 模型中，选用 sig 作为激活函数且隐含层神经元个数设为 100。图 10(c)显示了基于 SPA 筛选特征波长的 ELM 模型，隐含层神经元个数为 840 时，以 hardlim 作为激活函数预测集准确率最高。因此在紫外-可见光波段经 SPA 筛选特征波长的光谱数据建立的 ELM 模型中，选用 hardlim 作为激活函数且隐含层神经元个数设为 840。

表 2 是不同输入变量下 ELM 模型的判别结果。从表 2 中可以看出不经筛选特征波长的光谱数据建模识别效果最优，预测集准确率为 84.29%，其输入变量较多；基于 SPA 筛选特征波长的模型识别效果次之，预测集准确率为 81.43%，其输入变量为筛选的 9 个特征波长；基于 CARS 筛选特征波长建模识别效果最差，预测集准确率为 78.57%，其输入变量为筛选的 27 个特征波长。

表 2 不同输入变量下 ELM 模型判别结果
Table 2 ELM model discrimination results under different input variables

筛选方法	输入变量	隐含层神经元	激活函数	训练集准确率/%	预测集准确率/%
无	420	680	sig	100	84.29
CARS	27	100	sig	91.43	78.57
SPA	9	840	hardlim	100	81.43

2.2.6 判别模型的优化

由于 ELM 随机产生输入权值和隐含层阈值，使得模型

预测精度不高且存在泛化能力较差的问题。因此本文选用遗传算法(genetic algorithm, GA)优化 ELM 的输入权值和隐含层阈值。算法参数设定如下：初始种群规模为 30，最大遗传代数 200，交叉概率为 0.6，变异概率为 0.01。优化过程如图 11。

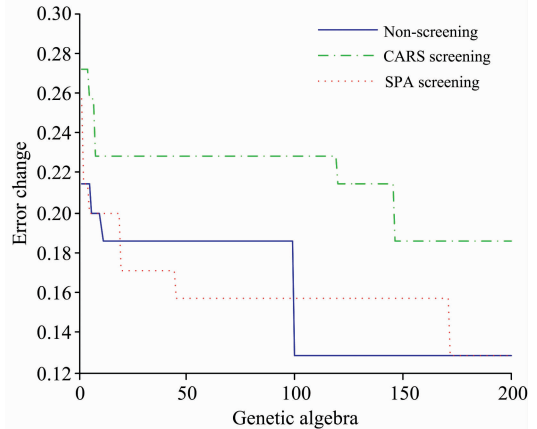


图 11 GA 优化不筛选、CARS 筛选和 SPA 筛选的 ELM 模型过程

Fig. 11 GA optimized ELM model of non-screening, CARS screening and SPA screening

GA 优化后的建模效果如表 3。基于紫外-可见光波段不经筛选特征波长的光谱数据建模，优化后预测集准确率从 84.29% 上升到 87.14%，其中有 6 个雌胚被判错，3 个雄胚被判错，雌性识别率为 82.86%，雄性识别率为 91.43%。以预测集单个样本的平均判别时间作为模型判别时间，优化后模型判别时间为 0.105 ms。基于紫外-可见光波段 CARS 筛选的 27 个特征波长建模，优化后预测集准确率从 78.57% 上升到 81.43%，其中有 6 个雌胚被判错，7 个雄胚被判错，雌性识别率为 82.86%，雄性识别率为 80.00%，优化后模型判别时间为 0.071 ms。基于紫外-可见光波段 SPA 筛选的 9 个特征波长建模，优化后预测集准确率从 81.43% 上升到 87.14%，其中有 4 个雌胚被判错，5 个雄胚被判错，雌性识别率为 88.57%，雄性识别率为 85.71%，优化后模型判别时间为 0.080 ms。

表 3 GA 优化前后 ELM 模型判别结果的对比

Table 3 Comparison of discriminant results of ELM model before and after GA optimization

筛选方法	ELM						GA-ELM						判别时间 /ms
	训练集		错误样本		预测集		训练集		错误样本		预测集		
	准确率/%	雌	雄	准确率/%	雌	雄	准确率/%	雌	雄	准确率/%	雌	雄	
无	100	0	0	84.29	9	2	100	0	0	87.14	6	3	0.105
CARS	91.43	6	6	78.57	10	5	96.43	2	3	81.43	6	7	0.071
SPA	100	0	0	81.43	8	5	100	0	0	87.14	4	5	0.080

结果表明紫外-可见光波段不经筛选特征波长的 GA-ELM 模型识别效果和经 SPA 筛选特征波长的 GA-ELM 模型相同，预测集准确率均为 87.14%，表明 SPA 筛选的特征波长能有效反映 360~780 nm 波段的信息。SPA 使用的变量

数仅占紫外-可见光波段的 2.14%，大大减少了模型输入变量，因此，雌雄识别最佳判别模型为紫外-可见光波段经 SPA 筛选特征波长的 GA-ELM 模型，预测集准确率为 87.14%，模型判别时间为 0.080 ms。结果表明紫外-可见光透射光谱

技术为孵化早期鸡胚蛋雌雄识别提供了一种可行方法。

3 结 论

为了对鸡种蛋胚胎进行雌雄识别,采用光谱技术研究了孵化期内鸡胚雌雄的无损鉴别方法。通过试验分析得出以下结论:

(1)对于横向和竖向大头向上 2 种放置方式获取的鸡种蛋 0~15 d 光谱数据,建立了极限学习机(ELM)鸡种蛋胚胎识别模型。其中竖向放置孵化 7 d 的鸡种蛋胚胎光谱数据,相比于其他孵化天数,建立的 ELM 模型具有更优的识别效果。因此,选用孵化第 7 d 竖向放置采集的光谱数据进行雌雄识别。

(2)基于竖向放置孵化 7 d 的鸡种蛋光谱,将光谱初步分为紫外(360~380 nm)、可见光(380~780 nm)、近红外(780~1 000 nm)、紫外-可见光(360~780 nm)和全波段(360~1 000 nm)5 个不同光谱范围来分析,结果表明 360~780 nm 的紫外-可见光波段的光谱建模效果最好;采用 MSC 进行预处理去噪,并用 CARS、SPA 筛选特征波长建模,对比分析

后得到紫外-可见光波段不经筛选特征波长的 ELM 模型判别效果最好,预测集准确率为 84.29%;紫外-可见光波段经 SPA 筛选特征波长的 ELM 模型判别效果次之,预测集准确率为 81.43%;紫外-可见光波段经 CARS 筛选特征波长的 ELM 模型判别效果最差,预测集准确率为 78.57%。

(3)用 GA 优化 ELM 模型的权值变量和隐含层神经元阈值,发现紫外-可见光波段不经筛选特征波长的 GA-ELM 模型识别效果和经 SPA 筛选特征波长的 GA-ELM 模型相同,预测集准确率均为 87.14%,但 SPA 使用的变量数仅占紫外-可见光波段的 2.14%。因此,雌雄识别最佳模型为紫外-可见光波段经 SPA 筛选特征波长的 GA-ELM 模型,模型准确率为 87.14%,其中有 4 个雌胚被判错,5 个雄胚被判错,雌性识别率为 88.57%,雄性识别率为 85.71%。单个样本平均判别时间 0.080 ms。

研究结果表明紫外-可见透射光谱技术和 ELM 模型为孵化早期鸡胚蛋雌雄识别提供了一种可行方法。但准确率有待进一步提高,在以后的研究中,可以改善光源,降低光谱在紫外波段的噪声,同时可以选择深度学习算法进一步来提高准确率;使用更多的样本数量,提高模型的通用性。

References

- [1] 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.
- [2] HAN Hao-yue, Jackie Linden(韩浩月, Jackie Linden). Animal Science Abroad(国外畜牧学), 2017, 37(1): 71.
- [3] Weissmann A, Reitemeier S, Hahn A, et al. Theriogenology, 2013, 80(3): 199.
- [4] Galli R, Koch E, Preusse G, et al. Current Directions in Biomedical Engineering, 2017, 3(2): 131.
- [5] Galli R, Preusse G, Uckermann O, et al. Analytical Chemistry, 2016, 88(17): 8657.
- [6] Galli R, Preusse G, Uckermann O, et al. Analytical & Bioanalytical Chemistry, 2017, 409(5): 1185.
- [7] ZHU Zhi-hui, TANG Yong, HONG Qi, et al(祝志慧, 汤勇, 洪琪, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2018, 34(6): 197.
- [8] Göhler D, Fischer B, Meissner S. Poult. Sci., 2017, 96(1): 1.
- [9] FENG Yan-ping(俸艳萍). Huazhong Agricultural University(华中农业大学), 2007.
- [10] ZHANG Xiao-chao, WU Jing-zhu, XU Yun, et al(张小超, 吴静珠, 徐云, 等). Beijing: Electronics Industry Press(北京: 电子工业出版社), 2012.
- [11] ZHANG Hua-xiu, LI Xiao-ning, FAN Wei, et al(张华秀, 李晓宁, 范伟, 等). Journal of Instrumental Analysis(分析测试学报), 2010, 29(5): 430.
- [12] Anderson S S, Telma W L, Daniel V L, et al. Computer Technology and Application, 2013, 4(9): 466.
- [13] Huang G, Zhu Q, Siew C. Neurocomputing, 2006, 70(1): 489.

Early Identification of Male and Female Embryos Based on UV/Vis Transmission Spectroscopy and Extreme Learning Machine

ZHU Zhi-hui^{1,2}, HONG Qi^{1,2}, WU Lin-feng^{1,2}, WANG Qiao-hua^{1,2}, MA Mei-hu^{3*}

1. College of Engineering, Huazhong Agricultural University, Wuhan 430070, China

2. Key Laboratory of Agricultural Equipment, Ministry of Agriculture, Wuhan 430070, China

3. College of Food Science and Technology, Huazhong Agricultural University, Wuhan 430070, China

Abstract In order to identify male and female embryos of chicken eggs, the feasibility of using UV/Vis/NIR transmission spectrum to identify the male and female embryos is explored. The transmission spectrum detection system of chicken eggs is established with blunt end vertically placed upwards and horizontally placed separately to obtain the 0~15 d spectrum (ranging from 360 to 1 000 nm) of 210 hatched eggs. The identification model of the embryo learning male and female of the extreme learning machine (ELM) is constructed. By comparing the identification accuracy of different placement and the number of hatching days, it is found that the recognition effect of the vertical placement hatching on the 7th day is the best. The spectrum of the 7th day of vertical incubation is initially divided into ultraviolet (360~380 nm), visible light (380~780 nm), near-infrared (780~1 000 nm), ultraviolet/visible (360~780 nm) and full-band (360~1 000 nm). Five different band ranges are analyzed, and the prediction set accuracy rates are 82.86%, 77.14%, 75.71%, 84.29%, and 81.43%, respectively. The ultraviolet/visible bands of 360~780 nm are selected as effective bands; In the ultraviolet/visible (360~780 nm) band, Multiplicative scatter correction (MSC) is used to denoise, and the characteristic wavelength reduction is selected by Competitive adaptive reweighted sampling (CARS) and Successive projection algorithm (SPA). Three kinds of wavelengths without screening, CARS screening characteristic wavelength and SPA screening characteristic wavelength are established ELM model. Among them, the ELM model without screening characteristic wavelengths has the best recognition effect, but the input variables are the most. When the hidden layer neuron is 680 and the activation function is sig, the prediction set accuracy is 84.29%. The ELM model of the SPA screening characteristic wavelength has the second recognition effect, and there are 9 input variables. When the hidden layer neurons are 840 and the activation function is hardlim, the prediction set accuracy is 81.43%. The ELM model with the CARS screening characteristic wavelength has the worst recognition effect, and there are 27 input variables. When the hidden layer neurons are 100 and the activation function is sig, the prediction set accuracy is 78.57%. Using Genetic algorithm (GA) to optimize the weight variable and hidden layer threshold of ELM model, the prediction set accuracy rate is 87.14%, 87.14% and 81.43% separately under the condition of the GA-ELM model established without screening the characteristic wavelength, the GA-ELM model established by SPA screening characteristic wavelength, and the GA-ELM model established by the CARS screening characteristic wavelength. The recognition effect of GA-ELM model in the ultraviolet/visible band without screening characteristic wavelength is the same as that in the GA-ELM model with SPA screening characteristic wavelength, which indicates that the characteristic wavelength variable of SPA screening can effectively reflect the information of 360~780 nm band. The number of variables used by the SPA is only 2.14% of the ultraviolet/visible range. Therefore, the best model for male and female identification is the GA-ELM model for screening characteristic wavelengths with SPA in the ultraviolet/visible range. The accuracy of the prediction set is 87.14%, of which the female recognition rate is 88.57%, the male recognition rate is 85.71%, and the average discrimination time of a single sample is 0.080 ms. The results show that UV/Vis transmission spectroscopy and ELM model provide a feasible method for the identification of chicken embryo eggs in early hatching.

Keywords Egg; Chicken embryo; Male and female; Recognition; UV

(Received Jul. 31, 2018; accepted Dec. 2, 2018)

* Corresponding author