

基于人工神经网络和二元逻辑回归的甜玉米种子生活力检测模型研究

刘敏洁^{1,2} 许暄^{1,2} 王建华² 孙群^{2*} 向春阳^{1*}

(1. 天津农学院, 天津 300384;

2. 中国农业大学 农学院/农业部农作物种子全程技术研究北京创新中心/北京市作物遗传改良重点实验室, 北京 100193)

摘要 为探索快速、高效检测甜玉米种子生活力的方法,利用机器视觉技术(Seed Identification)批量快速提取金菲甜玉米种子的 Red(红基色)、Green(绿基色)、Blue(蓝基色)、Hues(色相)、Saturation(饱和度)、Brightness(亮度)、Light(明度)、 a (红色至绿色的范围)、 b (蓝色至黄色的范围)、灰度、宽度、长度和投影面积等物理特征参数,通过单粒发芽试验确定每粒种子的生活力,然后采用人工神经网络和二元逻辑回归结合主成分分析进行建模。结果表明:1) a 值、 b 值、Saturation 和投影面积与种子的活力均存在极显著或显著相关,且变异系数相对较大,其中当 $a \leq 3$ 时,发芽率可从 72.7% 提升至 77.6%, 获选率达到 79.4%; 投影面积 $\leq 77.31 \text{ mm}^2$ 时,发芽率可提升至 73.7%, 获选率 87.6%; 2) 用 13 个物理指标标准化后直接进行人工神经网络建模,双隐藏层(训练集:测试集=6:4)建模,模型整体预测正确率为 74.2%, 优质种子获选率达到 93.8%, 发芽率可提升至 76.9%; 3) 经二元逻辑回归模型预测发芽率为 74.5%, 但神经网络模型稳定性优于二元逻辑回归建模。

关键词 甜玉米; 机器视觉技术; 主成分分析; 人工神经网络; 二元逻辑回归

中图分类号 S513 **文章编号** 1007-4333(2018)07-0001-10 **文献标志码** A

Seed viability testing model of sweet corn based on artificial neural network and binary logistic regression

LIU Minjie^{1,2}, XU Xuan^{1,2}, WANG Jianhua², SUN Qun^{2*}, XIANG Chunyang^{1*}

(1. Tianjin Agricultural University, Tianjin 300384, China;

(2. College of Agronomy and Biotechnology/Beijing Innovation Center of Crop Seeds Full Technologies Research of Ministry/Beijing Key Laboratory of Crop Genetic Improvement, China Agricultural University, Beijing 100193, China)

Abstract In order to test the vitality of sweet corn seed quickly and efficiently, several physical characteristics (red, green, blue, hues, saturation, brightness, light, a , b , width, length and projected area) of a certain cultivar of sweet corn, Jinfei, were obtained in batch and quickly went through image recognition technology (Seed Identification software). Each seed viability was confirmed through single seed germination test, and a model for seed viability discrimination of Jinfei was established by applying principle component analysis, artificial neural network and binary logistic regression. The results showed that: 1) a , b , saturation, width and projection area all had significant correlation with the seed vigor and relatively high variation coefficient. For the seeds with $a \leq 3$, the germination percentage was increased from 72.7% to 77.6%, and the selected rate reached to 79.4%. For the seeds with projection area $\leq 77.31 \text{ mm}^2$, the germination percentage can be increased to 73.7% and the selected rate reached 87.6%. 2) Based on 13 standardized physical indicators, artificial neural network model with double hidden layers (training set : testing set = 6 : 4) possessed an overall predicting accuracy at 74.2%, germination percentage also reached to 76.9%, with

收稿日期: 2017-07-22

基金项目: 北京科委项目(Z151100001015004)

第一作者: 刘敏洁, 硕士研究生, E-mail: 1053510830@qq.com

通讯作者: 向春阳, 教授, 主要从事特种玉米遗传育种与栽培研究, E-mail: xiang5918@sina.com

孙群, 副教授, 主要从事种子加工检验研究, E-mail: sqcau@126.com

93.8% quality seed selection. 3) The germination percentage of binary logistic regression was 74.5%, but the model stability of Artificial Neural Network model was better than binary logistic regression. These results would provide references for further research and implementation of quality seed sorting technology.

Keywords sweet corn; machine vision technology; Principal Component Analysis; Artificial Neural Network; Binary Logic Regression

甜玉米因其独特的水果般香甜鲜脆的口感和所富含的多种营养而深受人们喜爱。随着甜玉米需求量的增大,其表现出的种子活力低和田间出苗差成为甜玉米推广及发展的主要限制因素。种子生活力检测尤为重要。但是离体胚测定法和四唑染色等传统种子生活力检测方法具有周期长且种子损耗率高等缺点,无法满足日益增长的甜玉米种业市场的需求。因此,建立适用于检测甜玉米种子生活力的高效且无损的检测鉴定技术是提高甜玉米种子活力的有效途径之一。

机器视觉技术是计算机技术和图像识别、处理技术相结合的多领域新型交叉技术^[1],其作为一项新型无损鉴定筛选种子生活力高低的技术,具有非接触、高效和低成本特点,近几年来已在我国得到迅速地发展^[2]。很多学者从种子的纯度分析、品种鉴别、质量检测及种子精选等方面进行研究^[3-7],但对甜玉米种子生活力的机器视觉检测技术研究还较少。

主成分分析是利用降维的思维,可以将多个指标在不丢失其全部信息的前提下转化为几个较少的新指标,从而使多变量问题分析得到简化^[8-9]。因此在分析多因子问题时主成分分析是有效的化繁为简的方法。

人工神经网络是由多个非常简单的处理单元相互连接构成的一种运算模型。广泛应用于医疗、金融和交通等领域^[10-12],在农学中,神经网络也得到了广泛应用。李奔等^[13]用人工神经网络对所建模型识别玉米品种的研究中,3个品种总识别率达到93.3%。孙志超等^[14]采用神经网络所模拟的对超级玉米形状的研究模型效果表现良好。目前,将人工神经网络所建模型用于种子生活力检测的研究较少。

二元逻辑回归是一种以二分类变量为因变量的线性回归分析方法^[15]。目前,逻辑回归常被用来预测和判别某种疾病或危险的发生。苏光明等^[16]采用二元逻辑回归鉴定苹果汁掺假研究中发现其模型的精度和稳定性分别为87.4%和94.5%,该模型判别效果良好,可用于苹果汁的质量控制。孙传恒

等^[17]在昆虫学的研究中,二元逻辑回归模型被认为可以很好地识别各个光强度对棉铃虫反应的特性。目前,二元逻辑回归在农学中的应用也比较少。本研究以金菲甜玉米种子为试验材料,通过 Seed Identification 软件识别并采集不同生活力甜玉米种子的物理指标,通过标准发芽试验测得每一粒种子的苗长、根长和鲜重作为单粒种子的活力指标,结合相关性分析、主成分分析、人工神经网络及二元逻辑回归建立金菲甜玉米种子基于颜色或形态指标的分级精选模型,旨在为快速、无损、高效检测甜玉米种子的生活力提供理论参考。

1 材料与方法

1.1 试验材料

试验材料为金菲甜玉米种子,于2016年在甘肃收获。

1.2 试验方法

1.2.1 种子物理特征参数提取

随机选取300粒完整无损的金菲甜玉米种子,整齐摆放于清华紫光 unisanD6810 型扫描仪玻璃板上,横竖各10粒种子,每100粒扫描为一张图片。扫描仪分辨率设置为300 dpi,扫描后图片保存为png格式。所得图片用 Photoshop cs5 软件去除黑色背景,以确保种子被更好的识别。用 Seed Identification 软件打开预处理过的图片,进行物理特性识别,包括每粒种子的 Red(红基色)、Green(绿基色)、Blue(蓝基色)、Light(亮度)、 a (红色至绿色的范围)、 b (蓝色至黄色的范围)、Hues(色相)、Saturation(饱和度)、Brightness(明度)、灰度、长度、宽度和投影面积值(图1),数据导出并保存到 Excel 工作表。导出保存带序号种子图像。

1.2.2 标准发芽试验

将扫描后的种子进行卷纸发芽,每张卷纸摆放20粒种子,胚根均朝下统一摆放,记录编号,并在顶端套上密封袋防止水分流失,将卷纸置于25℃的全光照培养室进行发芽。第7天记录种子发芽数,发芽结束时测量幼苗的苗长、根长和鲜重(图2)。

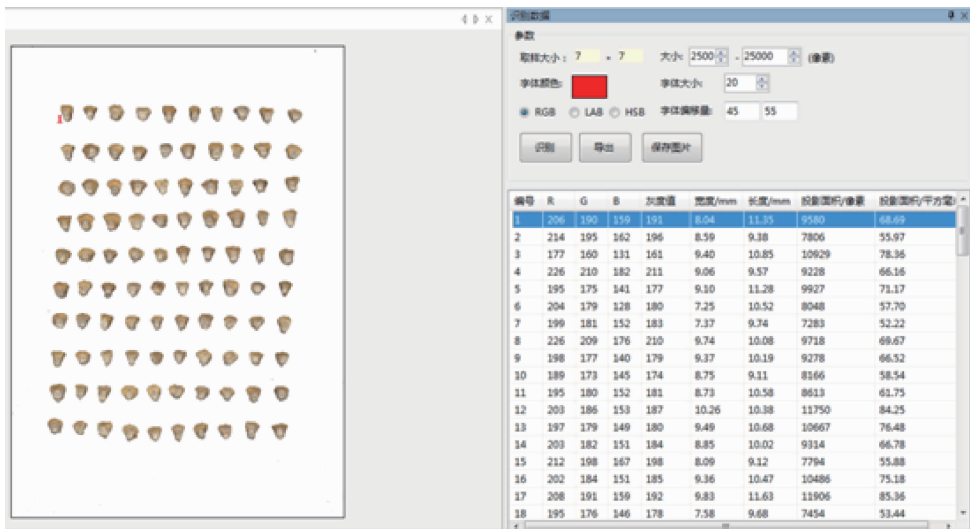


图 1 Seed Identification 软件识别界面

Fig. 1 Screen shot of identification interface of Seed Identification software



标尺=6 cm; Scale bar=6 cm

图 2 金菲甜玉米种子发芽第 7 天幼苗图像

Fig. 2 Seedling image on germination on day 7 of Jinfei sweet corn seeds

1.3 数据处理与分析

1.3.1 数据处理

在做神经网络和二元逻辑回归分析前,首先对 300 粒金菲甜玉米种子进行二分类,将不能发芽的死种子和不能正常发育为幼苗的种子定义为无生活力的种子,用 0 表示;将能够成长发芽成苗的种子分为有生活力的种子,标记为 1。并在分析前对各变量进行数据标准化处理。

1.3.2 数据分析

1) 单指标精选: 选用通过 Seed Identification 软件扫描获取的 13 个物理指标在 SPSS 21.0 中进行相关性分析。选用与活力指标有显著相关性且变异

系数较大的 4 个指标,在 Excel 2013 中利用频数分布函数分别将其等差距分为 4 个区间,计算每个区间内发芽率和获选率等值,并与原始发芽率比较,选择高于原始发芽率的区间确定为精选区间。

2) 神经网络建模分析: 分别选用标准化的 13 个物理指标、主成分分析确定的 4 个主成分和 6 个与活力指标显著相关且变异系数较大的物理指标作为输入层神经元、种子分类(0,1)作为输出层神经元,利用 SPSS 21.0 分析软件进行 BP 神经网络建模,训练集和测试集为随机选择,其样本数量比例分别设为 6:4、7:3 和 8:2,即分别选用原始数据的 60%、70% 和 80% 构建神经网络模型,40%、30% 和 20% 的数据用来评估模型性能。其中,隐藏层和输出层激活函数分别为双曲正切和 Softmax。因为考虑到隐藏层的个数对神经网络建模结果有影响,因此试验还讨论了相同变量在同一参数设定下双隐藏层所建模型的预测效果,其中两层单位数分别为 8 和 6。本试验中所建立模型均为 30 次运行后的平均结果。

3) 二元逻辑回归建模分析: 分别选用经过标准化的 13 个物理指标、主成分分析确定的 4 个主成分和 6 个与活力指标有显著相关性的物理指标利用 SPSS 21.0 分析软件进行二元逻辑回归建模分析。其中利用 SPSS 21.0 中计算变量中的 RV. Bernoulli 函数随机将 300 样本进行随机分区,设定用 67% 样本进行建模,33% 样本用来检验模型效果。

2 结果与分析

2.1 金菲种子各项物理指标描述性统计和相关性分析

经标准发芽试验测得总共 300 粒种子其中发芽并长成正常幼苗种子数 218 粒,不发芽种子 82 粒,

发芽率 72.7%。表 1 中可看出金菲甜玉米种子的 a 、 b 、Hues、Saturation、宽度和投影面积均与种子的活力指标(幼苗苗长、根长和鲜重)呈极显著或显著相关,其中 a 、 b 值、Saturation 和投影面积的变异系数相对较大,说明这几个物理指标可能对种子活力影响较明显,用于单个物理指标精选。

表 1 金菲物理指标描述性统计及其活力的相关性分析

Table 1 The descriptive statistics and correlation analysis of the physical indexes of Jinfei sweet corn seeds

项目 Item	平均值 Mean	变异系数/% Coefficient of variance	与苗长相关系数 Correlation coefficient with seeding length	与根长相关系数 Correlation coefficient with root length	与苗鲜重相关系数 Correlation coefficient with fresh weight
红基色 Red	197.84	7.33	-0.073	-0.077	-0.077
绿基色 Green	180.00	8.44	-0.033	-0.032	-0.030
蓝基色 Blue	149.16	10.87	0.007	0.014	0.024
亮度 Light	74.22	7.41	-0.038	-0.039	-0.039
a	3.07	30.26	-0.194**	-0.206**	-0.213**
b	18.42	14.92	-0.145*	-0.172**	-0.198**
色相 Hues	38.08	3.87	0.143*	0.160**	0.127*
饱和度 Saturation	24.71	16.78	-0.115*	-0.135*	-0.155**
明度 Brightness	77.58	7.37	-0.079	-0.083	-0.081
灰度 Grey	181.30	8.26	-0.039	-0.038	-0.035
宽度/mm Width	8.84	8.40	-0.183**	-0.129*	-0.143*
长度/mm Length	10.27	8.76	-0.111	-0.097	-0.101
投影面积/mm ² Projected area	66.47	14.97	-0.171**	-0.145*	-0.151**

注: a 表示从红色至绿色的范围; b 表示从黄色至蓝色的范围。**表示在 0.01 水平(双侧)上显著相关;*表示在 0.05 水平(双侧)上显著相关。下表同。

Note: a means the range between red and green; b means the range between yellow and blue. ** showed the significant correlation at 0.01 level (bilateral); * represents the significant correlation at 0.05 level. The same below.

2.2 金菲种子物理指标精选分级区间选择

通过选用已筛选出的 a 值、 b 值、投影面积和 Saturation 这 4 个指标对金菲甜玉米种子进行单指标精选(表 2),其中 $a \leq 3$ 时,发芽率可提升至 77.6%,获选率达到 79.4%;投影面积 $\leq 77.31 \text{ mm}^2$ 时,发芽率可提升至 73.7%,获选率 87.6%,但平均苗长、平均根长及平均鲜重提升效果不太明显。鉴于此,本研究

进一步采用主成分分析、人工神经网络和二元逻辑回归分析对金菲甜玉米种子物理指标进行建模分析。

2.3 各项物理指标主成分分析结果

如表 3 所示,用 SPSS 21.0 对金菲种子的 13 个物理参数进行主成分分析,共提取出 4 个主成分,累积贡献率达 94.22%。基本代表金菲种子所有颜色、形态信息。

表 2 金菲甜玉米种子各物理指标筛选分析结果

Table 2 Screening and analysis of physical parameters of Jinfei sweet corn seeds

指标 Index	区间 Interval	获选率/% Selected rate	发芽率/% Germination percentage	平均苗长/cm Averageseedling length	平均根长/cm Average root length	平均鲜重/g Average fresh weight
对照 CK		100.0	72.7	6.1	9.9	0.33
a	$a \leq 2$	32.6	84.5	7.1	12.2	0.42
	$2 < a \leq 3$	46.8	73.4	6.2	9.8	0.33
	$3 < a \leq 4$	14.7	58.2	4.9	7.5	0.25
	$a > 4$	6.0	61.9	4.5	7.7	0.24
	$a \leq 3$ 精选	79.4	77.6	6.5	10.7	0.36
b	$b \leq 15$	12.4	81.8	6.3	11.5	0.41
	$15 < b \leq 20$	98.6	74.9	6.5	10.3	0.34
	$20 < b \leq 25$	12.8	58.3	4.4	7.4	0.25
	$b > 25$	0.9	50.0	2.0	4.6	0.12
	$b \leq 20$ 精选	86.2	75.8	6.4	10.5	0.35
投影面积/mm ² Projected area	投影面积 ≤ 67.31	60.1	74.0	7.2	11.0	0.38
	$67.31 < \text{投影面积} \leq 77.31$	27.5	73.2	6.1	10.8	0.36
	$77.31 < \text{投影面积} \leq 87.31$	9.6	70.0	4.9	8.3	0.26
	投影面积 > 87.31	2.8	54.5	3.3	3.7	0.14
	投影面积 ≤ 77.3 精选	87.6	73.7	6.3	10.4	0.35
饱和度 S Saturation	$S \leq 20$	15.1	78.6	6.1	10.6	0.38
	$20 < S \leq 25$	52.8	74.7	6.4	10.4	0.35
	$25 < S \leq 30$	26.1	71.3	6.1	9.5	0.31
	$S > 30$	6.0	54.2	3.8	6.5	0.21
	$S \leq 25$ 精选	67.9	75.5	6.3	10.5	0.36

注：获选率 = 精选后有生活力种子数 / 精选前有生活力种子数 × 100%。

Note: Selected rate = viable seed after selection / viable seed before selection × 100%.

表 3 金菲甜玉米种子特征根和累计贡献率

Table 3 Characteristic roots and accumulative contribution rate of Jinfei sweet corn seeds

成份 Ingredient	初始特征值 Initial eigenvalue			提取平方和载入 Extract sum of squares load		
	特征根 Characteristic root	方差 Variance	累积贡献率/% Accumulated tribute offer rate	特征根 Characteristic root	方差 Variance	累积贡献率/% Accumulated tribute offer rate
1	6.816	52.432	52.432	6.816	52.432	52.432
2	2.317	17.821	70.253	2.317	17.821	70.253
3	2.034	15.647	85.900	2.034	15.647	85.900
4	1.082	8.324	94.224	1.082	8.324	94.224
5	0.626	4.818	99.042			
6	0.064	0.494	99.537			
7	0.045	0.344	99.881			
8	0.010	0.078	99.959			
9	0.003	0.020	99.979			
10	0.002	0.013	99.992			
11	0.001	0.005	99.997			
12	0.000	0.002	99.999			
13	0.000	0.001	100.000			

得出4个主成分表达式如下:

$$F_1 = 0.921ZX_1 + 0.980ZX_2 + 0.988ZX_3 + 0.970ZX_4 - 0.600ZX_5 - 0.401ZX_6 - 0.454ZX_7 - 0.744ZX_8 + 0.912ZX_9 + 0.974ZX_{10} + 0.109ZX_{11}$$

$$F_2 = 0.832ZX_{11} + 0.806ZX_{12} + 0.983ZX_{13}$$

$$F_3 = 0.388ZX_1 + 0.199ZX_2 + 0.237ZX_4 + 0.691ZX_5 + 0.817ZX_6 - 0.240ZX_7 + 0.616ZX_8 + 0.386ZX_9 + 0.225ZX_{10}$$

$$F_4 = 0.388ZX_1 + 0.199ZX_2 + 0.237ZX_4 + 0.691ZX_5 + 0.817ZX_6 - 0.240ZX_7 + 0.616ZX_8 + 0.386ZX_9 + 0.225ZX_{10}$$

式中: $F_1 \sim F_4$ 表示4个新的主成分; $ZX_1 \sim ZX_{13}$ 分别是 Red、Green、Blue、Light、 a 、 b 、Hues、Saturation、Brightness、灰度、宽度、长度和投影面积这13个物理指标经标准化处理后的数值。

2.4 人工神经网络建模分析

由表4可以看出,单隐藏层情况下,在测试集和训练集样本数量分区比例为6:4时,3种类别金菲种子神经网络模型T1类训练集和测试集模型识别

率最接近,模型表现稳定。且该模型整体预测正确率最高,为73.1%,经模型筛选后发芽率可提高至74.6%。当测试集和训练集样本数量分区比例为7:3和8:2时,相比其他两类模型,T1类模型仍表现出很好的预测效果。其中测试集和训练集样本数按8:2分区时所建模型,其测试集预测准确率高于训练集为77.3%,且模型稳定性差。T1类模型在测试集和训练集样本数量分区比例7:3时,模型整体预测率为73.6%,且经过该模型筛选,发芽率可提高3.2%。其次为T3类模型,即选用主成分分析提取的4个主成分作为协变量进行神经网络建模,在测试集和训练集样本数量分区比例为7:3时,可将发芽率提高2.6%,其模型整体预测准确率为72.6%。T2类模型,即选用与金菲种子活力有显著相关性的6个物理指标(a 、 b 、Hues、Saturation、宽度和投影面积)进行人工神经网络单隐藏层建模,模型整体预测准确率较其他两类模型低,最高仅为72.7%,且发芽率提升效果不理想,不建议选用此类指标进行神经网络建模。

表4 金菲甜玉米种子BP神经网络建成模型的识别准确率

Table 4 Recognition accuracy of BP neural network of Jinfei sweet corn seeds

样本类别 Sample category	样本数量比 (训练集:测试集) Sample size ratio (Training set:Testing set)	训练集识别率/%		测试集识别率/%	
		Recognition rate of training set		Recognition rate of testing set	
		单隐藏层 Single hidden layer	双隐藏层 double Hidden layers	单隐藏层 Single hidden layer	双隐藏层 double Hidden layers
T1	6:4	74.1	74.7	73.1	74.2
	7:3	75.5	75.9	73.6	72.5
	8:2	73.8	73.4	77.3	72.1
T2	6:4	73.2	73.3	71.7	70.1
	7:3	72.4	73.2	72.7	71.0
	8:2	73.3	71.8	69.6	76.1
T3	6:4	74.4	74.0	72.2	72.7
	7:3	73.9	73.3	72.6	71.2
	8:2	74.4	73.9	72.1	71.9

注:T1,选用13个物理指标进行神经网络建模;T2,选用与金菲种子活力显著相关的6个物理指标(a 、 b 、Hues、Saturation、宽度和投影面积)进行建模;T3,选用主成分分析提取的4个主成分进行建模。

Note:T1,13 physical indexes are used to model the neural network;T2,6 physical indexes (a , b , Hues, Saturation, width and projection area) related to the vigor of Jinfei are selected for modeling;T3,principal component analysis is used to extract the 4 principal components for modeling.

在双隐藏层情况下,综合模型稳定性及整体预测率,表现最好的为 T1 类模型,在测试集和训练集样本数量分区比例为 6 : 4 时模型表现最优,其测试集预测准确率为 74.7%、训练集预测准确率为 74.2%,测试集和训练集预测结果接近,证明模型预测效果比较稳定。经该模型预测,发芽率可提升至 76.9%,提高 4.2%。其他两类模型在不同样本数量分区比例所建模型与单隐藏层所建模型相比,其模型整体预测率均有所下降。

ROC 曲线所围成的面积可以进一步验证预测模型的效果,ROC 曲线下面积越接近 1,说明模型预测准确率越高,预测效果也越好^[18]。

由表 5 可以看出,T1 类模型在双隐藏层、测试集和训练集样本数量分区比例为 6 : 4 时所建模型的 ROC 曲线下面积最大,为 0.719;同样条件下进行单隐藏层神经网络建模时 ROC 曲线围成面积为 0.708。进一步说明 T1 类模型双隐藏层、样本数量分区比例为 6 : 4 时所建模型最优,具有良好的拟合度和较高的识别准确率。

表 5 金菲种子神经网络分析的 ROC 曲线所围成面积

Table 5 The area of the ROC curve of neural network of Jinfei sweet corn seeds

类别 Category	样本数量比 (训练集 : 测试集) Sample size ratio (Training set : Testing set)	ROC 围成的面积 The area of the ROC curve		
		0	1	
单隐藏层 Single hidden layer	T1	6 : 4	0.708	0.708
		7 : 3	0.710	0.710
		8 : 2	0.707	0.707
	T2	6 : 4	0.646	0.646
		7 : 3	0.682	0.682
		8 : 2	0.632	0.632
	T3	6 : 4	0.684	0.684
		7 : 3	0.696	0.696
		8 : 2	0.691	0.691
双隐藏层 Double hidden layer	T1	6 : 4	0.719	0.719
		7 : 3	0.712	0.712
		8 : 2	0.709	0.709
	T2	6 : 4	0.658	0.658
		7 : 3	0.659	0.659
		8 : 2	0.662	0.662
	T3	6 : 4	0.689	0.689
		7 : 3	0.678	0.678
		8 : 2	0.654	0.654

注: T1、T2、T3 同表 4;1 代表有生活力种子,0 代表无生活力种子。

Note: T1, T2 and T3 are the same as in Table 4; 1 represents viable seeds, 0 represents nonviable seeds.

2.5 二元逻辑回归分析结果

本研究提取不同类型的金菲种子物理参数变量采用 SPSS 21.0 进行二元逻辑回归分析,由表 6 可以看出,选用 13 个物理指标建立的二元逻辑回归模型 Hosmer 和 Lemeshow 检验 $P > 0.10$,说明适合进行逻辑回归建模,且模型整体预测率为 70.3%;用与种子活力指标相关的 6 个物理指标所建的逻辑回归模型 Hosmer 和 Lemeshow 检验 $P > 0.10$,模

型拟合度较好,对优质种子识别率为 91.5%,但对劣质种子识别率仅为 15.3%,因此经模型筛选对发芽率提升并没有多大贡献;选用 4 个主成分建模的 Hosmer 和 Lemeshow 检验 $P > 0.10$,达到 0.548,模型整体预测率为 71.2%,其中对于优质种子的预测率达到 91.5%,但是对劣质种子的鉴别模型表现较差。3 种模型相比较,用 4 个主成分建模效果较好。

表 6 金菲种子 Hosmer 和 Lemeshow 检验以及预测效果

Table 6 Hosmer and Lemeshow test and prediction results of Jinfei sweet corn seeds

类别 Category	卡方 Chi-square	自由度 Degree of freedom	显著性 Significance	预测率/% Prediction rate		
				劣质种子 Poor seeds	优质种子 High quality seed	整体 Overall
T1	9.34	8	0.315	18.6	90.2	70.3
T2	9.44	8	0.305	15.3	91.5	70.3
T3	6.89	8	0.548	18.6	91.5	71.2

2.6 3 种精选方法综合比较结果

由表 7 可知,相较于原始发芽率 72.7%,经 3 种方法精选后金菲甜玉米种子发芽率均有不同幅度提升,其中基于 BP 神经网络建模发芽率最高,为

76.9%,且获选率高于其他 2 种方法。综合考虑发芽率提升幅度、获选率及模型识别正确率,BP 神经网络建模分析法,即选用 13 个物理指标进行双隐藏层、测试集:训练集=6:4 建模,精选效果最佳。

表 7 金菲甜玉米种子 3 种精选方法的比较

Table 7 Comparison of 3 selection methods of Jinfei sweet corn seeds %

方法 Method	发芽率 Germination percentage	获选率 Selected rate	模型稳定性 Model stability
单指标精选	75.8	86.2	
人工神经网络模型	76.9	93.8	99.3
二元逻辑回归模型	74.5	91.5	92.1

注:神经网络模型稳定性=测试集识别率/训练集识别率 $\times 100\%$;二元逻辑回归模型稳定性=未选定案例识别率/选定案例识别率 $\times 100\%$ 。

Note:The model stability of Neural Network=recognition rate of testing set/recognition rate of training set $\times 100\%$;The model stability of Binary Logistic Regression=recognition rate of unselected case/selected case $\times 100\%$.

3 结论与讨论

本研究通过 Seed Identification 软件识别提取的金菲甜玉米种子物理性状及结合单粒标准发芽试验获得的活力指标基于 BP 人工神经网络分析建立

甜玉米种子生活力预测模型,结果发现,采用 13 个物理指标 (Red、Green、Blue、Light、a、b、Hues、Saturation、Brightness、灰度、宽度、长度和投影面积)的标准化数据进行神经网络双隐藏层建模效果最好,其中测试集和训练集样本数量比例为 6:4

时,所建模型整体预测效果最好,经模型筛选后发芽率从 72.7% 提高到 76.9%, 获选率达到 93.8%, 且模型稳定,效果优于其他方法。

农作物种子的质量好坏一定程度上直接影响产量和农民收益。因此,在生产中确保种子质量最关键。如何快速、无损和高效检测种子质量优劣,是国内外学者持续探索和关注的研究课题。目前实际生产中对于大批量种子的生活力检测及精选多靠传统检测方法和人为经验,通过机器视觉技术识别提取的种子物理参数结合不同数据分析方法建立种子生活力预测、筛选模型,可以实现种子生活力快速、高效和精准化检测分级,能有效降低成本及提高效率,在实际农业生产中具有重要意义。本研究基于神经网络和二元逻辑回归、结合 Seed Identification 软件建立的甜玉米种子生活力预测模型,对研究甜玉米种子生活力自动化检测系统提供理论依据,对促进机器视觉技术及数据建模在甜玉米种子生活力检测方面具有积极意义。

本研究尝试采用 SPSS 21.0 建立神经网络和二元逻辑回归模型,将其运用于种子生活力预测,模型预测效果较好。但是由于本研究所有数据来自一个甜玉米品种,结果存在一定局限性,研究结果推广性还需进一步验证。

参考文献 References

- [1] 李振. 基于机器视觉技术的蔬菜种子质量检测方法研究及系统设计[D]. 合肥:安徽大学,2016
Li Z. Research on vegetable seed quality detection method and system design based on machine vision technology[D]. Hefei: Anhui University, 2016 (in Chinese)
- [2] 叶凤林,李琳,杨丽明,王建华,孙群. 应用机器视觉技术筛选射干种子精选指标的研究[J]. 中国农业大学学报,2016,21(8): 119-124
Ye F L, Li L, Yang L M, Wang J H, Sun Q. Study on screening selection indexes of Shegan seed by machine vision technique [J]. *Journal of China Agricultural University*. 2016, 21(8): 119-124 (in Chinese)
- [3] Anvarkhah S, Hajehhosseini M K, Davariedalatpanah A, Mohassel M H R. Medicinal plant seed identification using machine vision[J]. *Seed Science and Technology*, 2013, 41(1): 107-120(14)
- [4] Pablo M. Granitto, Pablo F. Verdes, H. Alejandro Ceccatto. Large-scale investigation of weed seed identification by machine vision[J]. *Computers and Electronics in Agriculture*. 2004, 47(1): 15-24.
- [5] 程洪,史智兴,尹辉娟,冯娟,李亚男. 基于机器视觉的多个玉米籽粒胚部特征检测[J]. 农业工程学报,2013,29(19):145-151.
Cheng H, Shi Z, Yin H, Feng J, Li Y. Detection of multi-corn kernel embryos characteristic using machine vision [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2013, 29(19): 145-151 (in Chinese)
- [6] 彭江南,谢宗铭,杨丽明,孙宝启,王建华,孙群. 基于 Seed Identification 软件的棉籽机器视觉快速精选[J]. 农业工程学报,2013,29(23):147-152
Peng J N, Xie Z M, Yang L M, Sun B Q, Wang J H, Sun Q. Rapid selection of cottonseed machine vision based on Seed Identification software [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2013, 29(23): 147-152 (in Chinese)
- [7] 贾佳,王建华,谢宗铭,杨丽明,孙宝启,孙群. 计算机图像识别技术在小麦种子精选中的应用[J]. 中国农业大学学报,2014, 19(5):180-186
Jia J, Wang J H, Xie Z M, Yang L M, Sun B Q, Sun Q. Application of computer image identification technology in wheat seed selection [J]. *Journal of China Agricultural University*, 2014, 19(5): 180-186 (in Chinese)
- [8] 董海合,李风华,张旭,楼辰军,钱芳,杨兆顺. 主成分分析在甜玉米育种中的应用[J]. 玉米科学,2005,13(3):23-26.
Dong H H, Li F H, Zhang X, Lou C J, Qian F, Yang Z S. Application of Principal Component Analysis in Sweet Corn Breeding[J]. *Journal of Maize Sciences*, 2005, 13(3): 23-26 (in Chinese)
- [9] 李小胜,陈珍珍. 如何正确应用 SPSS 软件做主成分分析[J]. 统计研究,2010,27(8):105-108
Li X S, Chen Z Z. How to correctly apply SPSS software to make the principal component analysis[J]. *Statistical Research*, 2010, 27(8): 105-108 (in Chinese)
- [10] 绳慧峰,刘晴,许苹,陈春林. 神经网络在医疗风险预测中的应用研究[J]. 中国卫生质量管理,2017,24(4):15-17
Sheng H F, Liu Q, Xu P, Chen C L. The application of artificial neural network in the prediction of medical risk[J]. *Chinese Health Quality Management*, 2017, 24(4): 15-17 (in Chinese)
- [11] 王晓东,薛宏智,贾雯超. 基于 BP 神经网络的股票涨跌预测模型[J]. 价值工程,2010,29(31):47-49
Wang X D, Xue H Z, Jia W C. Prediction Model of Stock's rising and falling based on BP neural network[J]. *Value Engineering*, 2010, 29(31): 47-49 (in Chinese)
- [12] 安杰,董龙洋. 基于 BP 神经网络的道路交通安全预测模型研究[J]. 公路与汽运,2014(3):63-67
An J, Dong L Y. Research on road traffic safety prediction model based on BP neural network[J]. *Highways & Automotive Applications*, 2014(3): 63-67 (in Chinese)
- [13] 李奔. 基于图像颜色特征的玉米种子神经网络分类研究[C]. 丹东:第九届全国信息获取与处理学术会议, 2011
Li B. Neural network classification of corn seeds based on image color characteristics[C]. Dandong: The Ninth National

- Conference on information acquisition and processing, 2011 (in Chinese)
- [14] 孙志超, 荆绍凌, 刘文国, 张志军, 周小辉, 任军, 岳尧海, 张建新. BP神经网络模型对超级玉米性状的研究[J]. 玉米科学, 2008, 16(6): 42-43, 47
- Sun Z C, Jin S L, Liu W G, Zhang Z J, Zhou X H, Ren J, Yue R H, Zhang J X. Study on super corn characters by BP neural network model[J]. *Journal of Maize Sciences*, 2008, 16(6): 42-43, 47 (in Chinese)
- [15] 杜谦, 范文, 李凯, 杨德宏, 吕佼佼. 二元 Logistic 回归和信息量模型在地质灾害分区中的应用[J]. 灾害学, 2017, 32(2): 220-226
- Du Q, Fan W, Li K, Yang D H, Lv J J. Application of two element Logistic regression and information model in regionalization of geological disasters[J]. *Journal of Catastrophology*, 2017, 32(2): 220-226 (in Chinese)
- [16] Su G M, Gao H Y, Wang Z F, Liao X J, Zhang Y, Zhang M W, Hu X S, Wu J H. Identification model of apple juice based on binary logistic regression analysis [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2011, 27(6): 349-356
- [17] 孙传恒, 唐启义. Logistic 回归模型及其在昆虫学中的应用[J]. 昆虫知识, 2004, 41(6): 599-602
- Sun C H, Tang Q Y. Logistic Regression Model and its application in entomology [J]. *Chinese Bulletin of Entomology*, 2004, 41(6): 599-602 (in Chinese)
- [18] 万柏坤, 薛召军, 李佳, 王瑞平. 应用 ROC 曲线优选模式分类算法[J]. 自然科学进展, 2006, 16(11): 1511-1516
- Wan B K, Xue Z J, Li J, Wang R P. Application of ROC curve optimization model classification algorithm [J]. *Progress in Natural Science*, 2006, 16(11): 1511-1516 (in Chinese)

责任编辑: 吕晓梅