

基于特征性脂肪酸和甘油三酯指标的油茶籽油掺伪定性鉴别模型对比分析

孙婷婷^{1,2}, 刘剑波³, 任佳丽^{1,2}, 钟海雁^{1,2}, 周波^{1,2}

(1. 林产可食资源安全与加工利用湖南省重点实验室, 长沙 410004; 2. 中南林业科技大学 食品科学与工程学院, 长沙 410004; 3. 岳阳市检验检测中心食品药品检验所, 湖南 岳阳 414000)

摘要:为解决油茶籽油掺伪其他植物油的定性鉴别问题,在油茶籽油中分别掺入大豆油、花生油、葵花籽油、棉籽油、葡萄籽油、菜籽油、棕榈油和米糠油,设置高和低两种不同掺伪梯度,基于14个特征性脂肪酸和甘油三酯指标,运用Python语言构建并对比分析了二分类决策树模型、多分类决策树模型和多层感知机人工神经网络(MLP-ANN)模型用于油茶籽油掺伪定性鉴别的效果。结果表明:高和低掺伪梯度下,二分类决策树模型对油茶籽油掺伪其他植物油的定性鉴别的准确率均达到0.95以上;多分类决策树模型的精确率和准确率在高掺伪梯度下均达到了0.95,但在低掺伪梯度下仅为0.90;在高和低掺伪梯度下,MLP-ANN模型对油茶籽油掺伪定性鉴别的平均精确率均达到0.98,准确率分别达到0.97和0.98。相比于决策树模型,MLP-ANN模型能很好地实现油茶籽油掺伪定性鉴别。

关键词:油茶籽油;决策树模型;多层感知机人工神经网络模型;定性鉴别;脂肪酸;甘油三酯

中图分类号:TS227;O657

文献标识码:A

文章编号:1003-7969(2023)01-0066-08

Comparative analysis of qualitative identification models for oil-tea camellia seed oil adulteration based on characteristic fatty acid and triglyceride

SUN Tingting^{1,2}, LIU Jianbo³, REN Jiali^{1,2}, ZHONG Haiyan^{1,2}, ZHOU Bo^{1,2}

(1. Hunan Key Laboratory of Forestry Edible Sources Safety and Processing, Changsha 410004, China;

2. School of Food Science and Engineering, Central South University of Forestry and Technology,

Changsha 410004, China; 3. Food and Drug Inspection Institute of Yueyang City

Inspection and Testing Center, Yueyang 414000, Hunan, China)

Abstract: In order to solve the qualitative identification problem of adulterated oil-tea camellia seed oil with other vegetable oils, soybean oil, peanut oil, sunflower seed oil, cottonseed oil, grape seed oil, rapeseed oil, palm oil and rice bran oil were mixed into oil-tea camellia seed oil respectively, two different adulteration gradients of high and low were set up, and based on characteristic fatty acid and triglyceride indicators, the effects of the binary decision tree model, multi-classification decision tree model and multilayer perceptron artificial neural network (MLP-ANN) model for qualitative identification of adulterated oil-tea camellia seed oil were compared and analysed using Python language. The results showed that the accuracy of the binary decision tree model for qualitative identification of oil-tea camellia seed oil adulterated with other vegetable oils under high and low

adulteration gradients was above 0.95. The accuracy and precision of the multi-classification decision tree model reached 0.95 at high adulteration gradient, but only 0.90 at low adulteration gradient. Under high and low adulteration gradients, the average precision of MLP-ANN model for qualitative identification of adulterated oil-tea camellia seed oil reached

收稿日期:2021-10-26;修回日期:2022-09-30

基金项目:湖南省市场监督管理局科技计划项目(2020KJJH55);湖南省林业科技创新基金项目(XLK202101-02);中央引导地方科技发展专项资金区域创新体系建设专项(2020ZYQ036)

作者简介:孙婷婷(1994),女,硕士,研究方向为粮食、油脂与植物蛋白工程(E-mail)726153254@qq.com。

通信作者:周波,教授(E-mail)zhbo78@126.com。

0.98, and the accuracy reached 0.97 and 0.98 respectively. Compared with the decision tree model, the MLP-ANN model can well realize the qualitative identification of adulterated oil-tea camellia seed oil.

Key words: oil-tea camellia seed oil; decision tree model; multilayer perceptron artificial neural network model; qualitative identification; fatty acid; triglyceride

目前快速准确鉴别食用植物油掺伪的研究,大部分是首先利用先进的实验仪器和检测技术获得大量复杂的结构化测量数据,然后根据鉴别需求和问题的性质,采用不同的机器学习算法来挖掘和提取出掺伪鉴别所需的特征信息^[1]。其中检测技术和方法包括常规理化检测法^[2]、色谱法^[3-4]、核磁共振法^[5]、近红外光谱法^[6-7]、拉曼光谱法^[8]、荧光光谱法^[9-10]、电子鼻技术^[11]及稳定同位素质谱法^[12]等。相对而言,色谱法具有灵敏度高、选择性强、分析速度快、操作简便、样品用量少等优势,是目前广泛应用在植物油掺伪鉴别中的方法。

食品的掺伪鉴别问题分为两类,分别是掺伪定性鉴别和掺伪定量预测,二者分别属于分类问题和回归问题。目前用于油脂掺伪鉴别模型建立的机器学习算法主要包括主成分分析法(PCA)^[13]、判别分析法(DA)^[14-15]、支持向量机(SVM)^[16]、随机森林(RF)^[17-18]、人工神经网络(ANN)^[19-22]和偏最小二乘回归算法(PLSR)^[23-25]等。PCA和DA应用广泛,原理简单,易于操作,但具有鉴别准确率较低,对低掺伪样本的鉴别能力差,非线性分类能力差且存在主观性等缺点;SVM对低掺伪样本鉴别能力较强,但训练操作耗时且复杂,计算复杂度较高;RF鉴别准确率较高,可识别对掺伪鉴别有重要影响的特征,但原理复杂,计算复杂度较高,且只适合解决定性鉴别问题;ANN鉴别准确率高,对低掺伪样本的鉴别能力强,可用于定性鉴别和定量鉴别,但训练操作耗时且复杂,计算复杂度较高,模型可解释性差,训练所需数据规模大;PLSR综合考虑了多种特征性物质与掺伪量的映射关系,但其只能拟合线性相关关系,掺伪量与特征性物质含量间的非线性因素会影响鉴别结果。此外,掺伪鉴别方法的准确率不仅取决于所采用的机器学习算法,在很大程度上还受到实验样本数据的影响,如当实验样本的覆盖性较弱时,训练集无法充分反映掺伪样本的全部特征,在此基础上训练得到的不同机器学习算法模型的鉴别准确性会受到影响,无法对训练集未覆盖的样本做出准确的掺伪鉴别;同时,在没有充分训练实验样本的前提下,训练得到的机器学习算法模型对于新出

现的新品种油脂(化合物组成种类和含量不同)的检测精度受到影响^[26]。

油茶籽油是由山茶科(Theaceae)山茶属(*Camellia*)的植物种子制备而成的一种营养丰富的食用植物油^[27],主要生产地在湖南、江西、浙江、广西和贵州等地,与橄榄油、棕榈油和椰子油并称为世界四大木本油脂^[28]。油茶籽油具有较高的营养功能价值和商品价格,导致市场上用低价食用植物油掺伪油茶籽油的现象较普遍^[29],现有用于掺伪的植物油有大豆油、花生油、葵花籽油、棉籽油、菜籽油等。这一现象严重损害了油茶籽油生产者和消费者的利益,所以建立快速、精准的检测技术和方法来鉴别油茶籽油掺伪是保障我国油茶籽油生产和销售市场正常秩序以及实现油茶籽油高质量发展的必然需求。

为使得研究成果契合油茶籽油掺伪鉴别的实际应用需求,本文综合考虑了当前市场上现有的8种用于掺伪油茶籽油的食用植物油,包括大豆油、花生油、葵花籽油、棉籽油、葡萄籽油、菜籽油、棕榈油、米糠油。同时,由于脂肪酸和甘油三酯是掺伪油茶籽油鉴别相关研究中常用的特征性指标,故本文基于脂肪酸和甘油三酯数据开展研究。综上,本文基于脂肪酸和甘油三酯的色谱数据,运用Python语言建立并对比分析决策树(Decision tree)和多层感知机人工神经网络(MLP-ANN)模型用于掺伪油茶籽油的定性鉴别效果,以期鉴别掺伪油茶籽油纯度及定量分析调和油茶籽油配比提供理论基础。

1 数据采集

设计油茶籽油($n=53$)中分别掺入大豆油、花生油、葵花籽油、棉籽油、葡萄籽油、菜籽油、棕榈油、米糠油的掺伪模型,设计2个掺伪梯度,自定为高掺伪梯度(10%、15%、20%、40%、60%、80%)和低掺伪梯度(2%、4%、6%、8%、10%)。

本文中油茶籽油及掺伪其他植物油油茶籽油的脂肪酸和甘油三酯指标基于本实验室已报道^[30-31]数据。实验中采用的14个特征性脂肪酸和甘油三酯指标包括棕榈酸、硬脂酸、油酸、亚油酸、亚麻酸、棕榈酸/硬脂酸、油酸/硬脂酸、亚油酸/油酸以及当量碳数40(ECN40)、当量碳数42(ECN42)、当量碳

数 44 (ECN44)、当量碳数 46 (ECN46)、当量碳数 48 (ECN48)、当量碳数 50 (ECN50) 的甘油三酯。

2 模型建立及验证

2.1 决策树 (Decision tree) 模型

本文将二分类决策树模型和多分类决策树模型分别应用于油茶籽油掺伪定性鉴别的研究。二分类决策树模型中样本的标签共分为 2 类, 分别是数据标签值为“0”的掺伪油茶籽油和数据标签值为“1”

的纯油茶籽油。多分类决策树模型中样本的掺伪种类标签共分为 9 类, 分别是纯油茶籽油、掺大豆油、掺花生油、掺葵花籽油、掺棉籽油、掺葡萄籽油、掺菜籽油、掺棕榈油、掺米糠油的油茶籽油。各类样本数据数量及标签设置见表 1。利用 Python 语言下 scikit - learn 库的 tree. export_graphviz 函数和 decision tree classifier 函数对所构建的二分类和多分类决策树模型进行建模和可视化展示。

表 1 各类样本数量及标签设置

掺伪种类	二分类决策树模型			多分类决策树模型		
	标签值	高掺伪样本数量	低掺伪样本数量	标签值	高掺伪样本数量	低掺伪样本数量
纯油茶籽油	1	53	53	0	53	53
掺大豆油	0	318	265	1	318	265
掺花生油	0	318	265	2	318	265
掺葵花籽油	0	318	265	3	318	265
掺棉籽油	0	318	265	4	318	265
掺葡萄籽油	0	318	265	5	318	265
掺菜籽油	0	318	265	6	318	265
掺棕榈油	0	318	265	7	318	265
掺米糠油	0	318	265	8	318	265
合计		2 597	2 173		2 597	2 173

2.2 多层感知机人工神经网络 (MLP - ANN) 模型

本文构建基于反向传播算法训练的 MLP - ANN 模型的网络结构为 1 层输入层 (含 14 个神经元)、3 层隐藏层 (各层均含 24 个神经元) 和 1 层输出层 (含 9 个神经元)。此模型的样本数量及标签设置与多分类决策树模型的设置一致。

2.3 模型评价指标选取

按照 5 - 折交叉验证方法对模型开展训练和验证。交叉验证可避免由于数据划分方式导致的模型偏差, 有效地避免模型出现过拟合和欠拟合的情况, 更为客观地评价模型的准确性和可靠性。5 - 折交叉验证方法是将数据集随机分为数量均等的 5 份, 依次取其中 1 份数据作为验证集以验证模型的准确性, 其余 4 份数据用于训练模型。以上过程共进行 5 次, 使每份数据都能用于验证模型准确性, 最终所得模型准确性评价结果建立在整个数据集之上。

采用常见的分类模型评价指标如精确率、准确率、召回率、F - 分数等对模型的精度做出评价。精确率代表预测为正的样本中有多少是真正的正样本; 准确率代表预测正确的样本占所有样本的比例; 召回率代表样本中的正例有多少被预测正确; F - 分数是兼顾了精确率和召回率的评估指标。按公式(1) ~ 公式(4) 计算精确率 (P)、准确率 (A)、召回率 (R)、F - 分数 (F_1)。

$$P = P_T / (P_T + P_F) \quad (1)$$

$$A = (P_T + N_T) / (P_T + N_T + P_F + N_F) \quad (2)$$

$$R = P_T / (P_T + N_F) \quad (3)$$

$$F_1 = 2PR / (P + R) = 2P_T / (2P_T + P_F + N_F) \quad (4)$$

式中: P_T 是被预测正确的正样本的数量; P_F 是被错判为正的负样本数量; N_T 是被预测正确的负样本的数量; N_F 是被错判为负的正样本数量。本文中, 预测正确是指预测标签与实际标签一致, 预测错误是指预测标签与实际标签不一致。

2.4 数据处理

所有模型和算法均基于 Python 3.7 编程语言在 PyCharm 2018 IDE 平台 [JetBrains (Prague), Czech Republic] 进行程序编写。

3 模型鉴别分析

3.1 决策树掺伪鉴别模型

3.1.1 决策树掺伪鉴别模型精度分析

不同掺伪梯度下各二分类决策树模型的掺伪鉴别精度指标见表 2。由表 2 可知: 二分类决策树模型对高掺伪梯度和低掺伪梯度下油茶籽油样本掺伪其他植物油的定性鉴别效果较好, 准确率都达到 0.95 以上; 特别是高掺伪梯度下, 定性鉴别掺米糠油和棕榈油的油茶籽油的精确率和准确率均达到 1.00, 但高掺伪梯度下二分类决策树模型对掺葵花籽油和大豆油的油茶籽油的定性鉴别效果较差, 准确率均为 0.96; 而低掺伪梯度下二分类决策树模型对掺葵花籽油和棉籽油的油茶籽油的定性鉴别效果较差, 准确率均为 0.97。

表2 不同掺伪梯度下各二分类决策树模型的掺伪鉴别精度指标

掺伪种类	掺伪梯度	标签值	精确率	召回率	F-分数	准确率
掺大豆油	高	0	0.98	0.98	0.98	0.96
		1	0.84	0.83	0.84	
		平均值	0.96	0.96	0.96	
	低	0	1.00	1.00	1.00	0.99
		1	0.98	0.97	0.98	
		平均值	0.99	0.99	0.99	
掺花生油	高	0	1.00	1.00	1.00	1.00
		1	0.99	0.98	0.99	
		平均值	1.00	1.00	1.00	
	低	0	1.00	0.99	0.99	0.99
		1	0.96	0.97	0.96	
		平均值	0.99	0.99	0.99	
掺菜籽油	高	0	1.00	1.00	1.00	1.00
		1	0.98	1.00	0.99	
		平均值	1.00	1.00	1.00	
	低	0	0.99	0.99	0.99	0.98
		1	0.95	0.91	0.93	
		平均值	0.98	0.98	0.98	
掺葵花籽油	高	0	0.97	0.98	0.97	0.96
		1	0.83	0.81	0.82	
		平均值	0.96	0.96	0.96	
	低	0	0.98	0.98	0.98	0.97
		1	0.87	0.87	0.87	
		平均值	0.97	0.97	0.97	
掺米糠油	高	0	1.00	1.00	1.00	1.00
		1	1.00	0.99	1.00	
		平均值	1.00	1.00	1.00	
	低	0	0.99	0.99	0.99	0.98
		1	0.92	0.93	0.92	
		平均值	0.98	0.98	0.98	
掺棉籽油	高	0	1.00	1.00	1.00	1.00
		1	0.98	0.98	0.98	
		平均值	1.00	1.00	1.00	
	低	0	0.98	0.99	0.98	0.97
		1	0.90	0.85	0.87	
		平均值	0.97	0.97	0.97	
掺葡萄籽油	高	0	0.99	1.00	0.99	0.99
		1	0.98	0.95	0.96	
		平均值	0.99	0.99	0.99	
	低	0	0.99	0.99	0.99	0.99
		1	0.96	0.92	0.94	
		平均值	0.99	0.99	0.99	
掺棕榈油	高	0	1.00	1.00	1.00	1.00
		1	1.00	1.00	1.00	
		平均值	1.00	1.00	1.00	
	低	0	1.00	1.00	1.00	0.99
		1	0.97	0.97	0.97	
		平均值	0.99	0.99	0.99	

不同掺伪梯度下多分类决策树模型的掺伪鉴别精度指标见表3。由表3可知,高掺伪梯度和低掺伪梯度下,多分类决策树模型对油茶籽油掺伪油脂种类的定性鉴别精确率、准确率都较高,分别达到了0.95、0.95和0.90、0.90,但稍低于二分类决策树模型,这可能是由于多分类决策树模型需要给出未知样本的具体掺伪种类,分类数量共9类,问题更复杂,模型复杂度和训练难度都更高。

由表3还可知,高掺伪梯度下多分类决策树模型对掺大豆油和葵花籽油的油茶籽油的定性鉴别精确率较低,分别为0.84和0.85,这与二分类决策树模型的定性鉴别结果一致。相比于高掺伪梯度,低掺伪梯度下多分类决策树模型对于纯油茶籽油以及掺葵花籽油、棕榈油和棉籽油的油茶籽油的定性鉴别的精准度较差,精确率分别为0.40、0.81、0.87和0.88,其中对纯油茶籽油的鉴别能力最差,精确率低于0.5,这说明在低掺伪梯度下以脂肪酸和甘油三酯来鉴别掺伪其他种类植物油的油茶籽油的可行性较低。

表3 不同掺伪梯度下多分类决策树模型的掺伪鉴别精度指标

掺伪梯度	掺伪种类	精确率	召回率	F-分数	准确率
高	纯油茶籽油	0.98	0.98	0.98	0.95
	掺大豆油	0.84	0.86	0.85	
	掺花生油	1.00	1.00	1.00	
	掺葵花籽油	0.85	0.81	0.83	
	掺棉籽油	0.98	0.99	0.98	
	掺葡萄籽油	0.98	0.99	0.99	
	掺菜籽油	0.99	0.99	0.99	
	掺棕榈油	0.99	1.00	1.00	
	掺米糠油	0.99	0.99	0.99	
平均值	0.95	0.95	0.95		
低	纯油茶籽油	0.40	0.36	0.38	0.90
	掺大豆油	0.96	0.98	0.97	
	掺花生油	0.95	0.95	0.95	
	掺葵花籽油	0.81	0.88	0.85	
	掺棉籽油	0.88	0.86	0.87	
	掺葡萄籽油	0.90	0.88	0.89	
	掺菜籽油	0.98	0.95	0.96	
	掺棕榈油	0.87	0.90	0.88	
	掺米糠油	0.96	0.93	0.94	
平均值	0.90	0.90	0.90		

3.1.2 决策树掺伪鉴别模型可视化混淆矩阵分析
不同掺伪梯度下各二分类决策树模型的混淆矩

阵取值见表4。由表4可知,高掺伪梯度和低掺伪梯度下各二分类决策树模型对油茶籽油掺伪其他植物油具有较好的鉴别能力,但高掺伪梯度下掺大豆油和葵花籽油的油茶籽油样本中,有小部分样本被错误分类为掺伪其他类植物油,在低掺伪梯度下掺棉籽油和葵花籽油的油茶籽油样本中,有小部分样本被错误分类为掺伪其他类植物油。

表4 不同掺伪梯度下各二分类决策树模型的混淆矩阵取值

掺伪种类	掺伪梯度	真实标签	预测标签	
			掺伪其他油	掺伪该类油
掺大豆油	高	掺伪其他油	2 175	51
		掺伪该类油	53	265
	低	掺伪其他油	1 850	5
		掺伪该类油	7	258
掺花生油	高	掺伪其他油	2 223	3
		掺伪该类油	5	313
	低	掺伪其他油	1 843	12
		掺伪该类油	8	257
掺菜籽油	高	掺伪其他油	2 221	5
		掺伪该类油	1	317
	低	掺伪其他油	1 842	13
		掺伪该类油	23	242
掺葵花籽油	高	掺伪其他油	2 173	53
		掺伪该类油	59	259
	低	掺伪其他油	1 819	36
		掺伪该类油	34	231
掺米糠油	高	掺伪其他油	2 225	1
		掺伪该类油	2	316
	低	掺伪其他油	1 833	22
		掺伪该类油	19	246
掺棉籽油	高	掺伪其他油	2 220	6
		掺伪该类油	5	313
	低	掺伪其他油	1 829	26
		掺伪该类油	40	225
掺葡萄籽油	高	掺伪其他油	2 219	7
		掺伪该类油	16	302
	低	掺伪其他油	1 844	11
		掺伪该类油	20	245
掺棕榈油	高	掺伪其他油	2 226	0
		掺伪该类油	0	318
	低	掺伪其他油	1 848	7
		掺伪该类油	9	256

不同掺伪梯度下多分类决策树模型在 5 - 折交叉验证下的可视化混淆矩阵如图 1 所示。

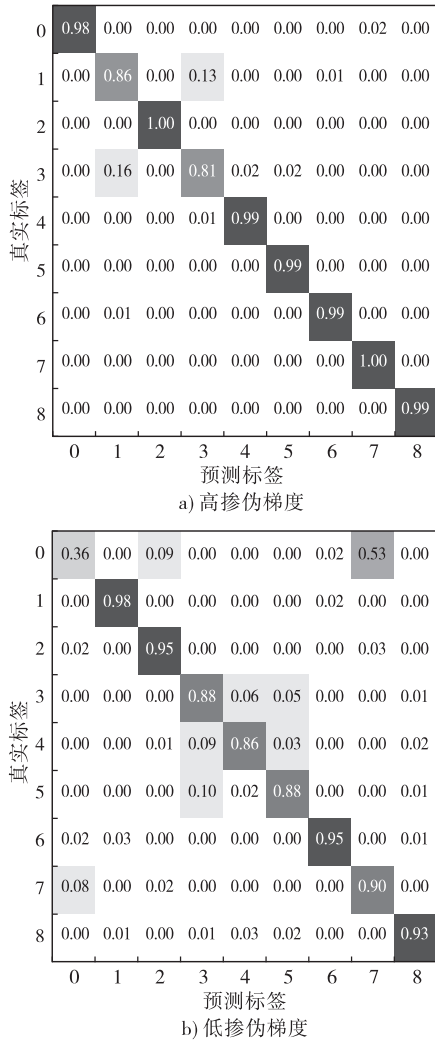


图 1 不同掺伪梯度下多分类决策树模型的可视化混淆矩阵

由图 1a 可知,高掺伪梯度下,多分类决策树模

型的定性鉴别精准度较高,但有 41 个掺大豆油的油茶籽油样本被误判为掺葵花籽油,有 50 个掺葵花籽油的油茶籽油样本被误判为掺大豆油。由图 1b 可知,低掺伪梯度下,多分类决策树模型的定性鉴别精准度较差,有 28 个纯油茶籽油样本被误判为掺棕榈油,有 27 个掺葡萄籽油的油茶籽油样本被误判为掺葵花籽油,有 24 个掺棉籽油的油茶籽油样本被误判为掺葵花籽油,有 21 个掺棕榈油的油茶籽油样本被误判为纯油茶籽油,有 17 个掺葵花籽油的油茶籽油样本被误判为掺棉籽油,有 12 个掺葵花籽油的油茶籽油样本被误判为掺葡萄籽油。

以上结果表明,在高掺伪梯度和低掺伪梯度下,决策树模型对油茶籽油掺伪其他植物油的定性判断准确率均较高,但二分类决策树模型在实际应用中存在模型数量较多、各决策相互独立、分类结论相矛盾和冲突等问题。本文所建多分类决策树模型可直接给出油茶籽油样本掺伪种类分类结论,但模型在低掺伪梯度下的定性鉴别精确率较低,且在实际应用中,其训练难度和模型复杂度高于二分类决策树模型,说明多分类决策树模型不适合解决低浓度梯度多元掺伪定性鉴别问题。

3.2 MLP - ANN 掺伪鉴别模型

不同掺伪梯度下 MLP - ANN 模型在 5 - 折交叉验证下的定性鉴别精度指标见表 5。不同掺伪梯度下 MLP - ANN 模型的可视化混淆矩阵见图 2。由表 5 可知,高掺伪梯度和低掺伪梯度下,MLP - ANN 模型对油茶籽油掺伪种类的鉴别精确率均较高,平均精确率均达到 0.98,高于多分类决策树模型,但该模型对高掺伪梯度下掺大豆油和葵花籽油的油茶籽油鉴别精确率较低,分别为 0.86 和 0.93。

表 5 不同掺伪梯度下 MLP - ANN 模型的定性鉴别精度指标

掺伪种类	高掺伪梯度			低掺伪梯度		
	精确率	准确率	F - 分数	精确率	准确率	F - 分数
纯油茶籽油	1.00	0.98	0.99	0.87	0.87	0.87
掺大豆油	0.86	0.94	0.90	1.00	1.00	1.00
掺花生油	1.00	1.00	1.00	1.00	1.00	1.00
掺葵花籽油	0.93	0.85	0.89	1.00	1.00	1.00
掺棉籽油	1.00	1.00	1.00	1.00	1.00	1.00
掺葡萄籽油	1.00	1.00	1.00	1.00	1.00	1.00
掺菜籽油	1.00	1.00	1.00	1.00	1.00	1.00
掺棕榈油	1.00	1.00	1.00	0.97	0.97	0.97
掺米糠油	1.00	1.00	1.00	1.00	1.00	1.00
平均值	0.98	0.97	0.98	0.98	0.98	0.98

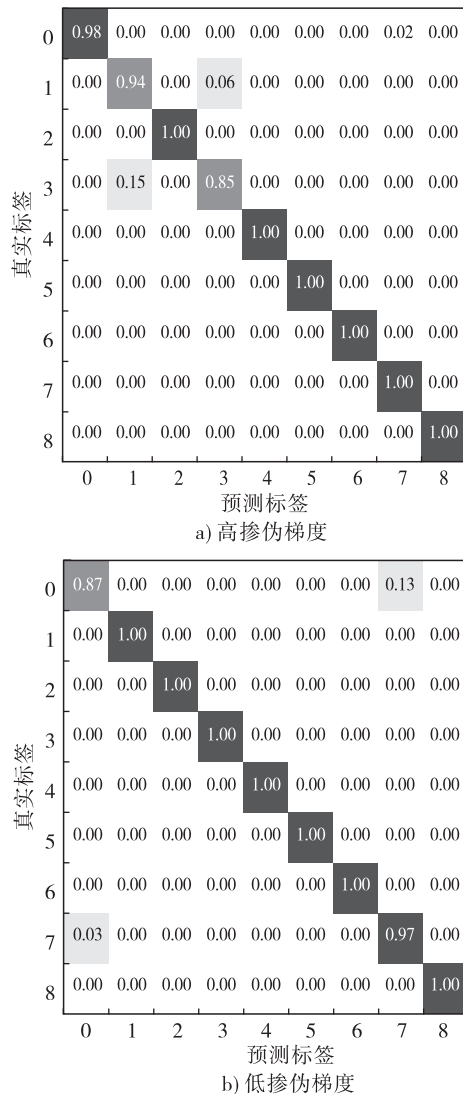


图2 不同掺伪梯度下 MLP-ANN 模型的可视化混淆矩阵

由图 2a 可知,高掺伪梯度下,MLP-ANN 模型对大部分样本的掺伪定性鉴别正确,少量掺伪样本定性鉴别错误,如 53 个纯油茶籽油样本中有 1 个样本被误判为掺棕榈油(准确率 0.98),318 个掺大豆油的油茶籽油样本中有 20 个样本被误判为掺葵花籽油(准确率 0.94),318 个掺葵花籽油的油茶籽油样本中有 47 个样本被误判为掺大豆油(准确率 0.85),MLP-ANN 模型对掺花生油、棉籽油、葡萄籽油、菜籽油、棕榈油、米糠油的油茶籽油样本的定性鉴别准确率都达到了 1.00。

由图 2b 可知,低掺伪梯度下,MLP-ANN 模型对大部分样本的掺伪定性鉴别正确,少量掺伪样本定性鉴别错误,如 53 个纯油茶籽油样本中有 7 个样本被误判为掺棕榈油(准确率 0.87),265 个掺棕榈油的油茶籽油的样本中有 7 个样本被误判为纯油茶籽油(准确率 0.97),在 265 个掺葵花籽油、花生油和葡萄籽油的油茶籽油样本中分别都只有 1 个样本被误判为掺米糠油、菜籽油和葵花籽油,准确率都达

到 1.00,MLP-ANN 模型对掺大豆油、棉籽油、菜籽油、米糠油的油茶籽油样本的定性鉴别准确率都达到了 1.00。

以上结果表明,在高和低掺伪梯度下,与决策树模型相比,MLP-ANN 模型的整体精准率和准确率有很大提升,达到了良好的油茶籽油掺伪种类鉴别效果。

4 结论

本文基于 14 个特征性脂肪酸和甘油三酯指标,以决策树和人工神经网络机器学习算法构建了三种掺伪鉴别模型(二分类决策树模型、多分类决策树模型和多层感知机人工神经网络模型)用于油茶籽油掺伪其他植物油的定性鉴别,三种模型都能很好地实现掺伪油茶籽油的定性鉴别,多层感知机人工神经网络模型最佳,多分类决策树模型和二分类决策树模型次之。另外,本文在实验中发现,掺伪鉴别模型对掺伪不同种类植物油的油茶籽油样本的鉴别效果有所不同,如高掺伪梯度下掺大豆油或葵花籽油的油茶籽油样本的鉴别准确率与掺其他种类植物油的油茶籽油相比较低。猜测原因可能在于不同种类植物油的脂肪酸和甘油三酯组成不同,当掺伪油脂的组成与油茶籽油相似时则掺伪难以鉴别,当掺伪油脂的组成与油茶籽油差别较大时则掺伪鉴别效果较好。

参考文献:

- [1] MUNCK L, NØRGAARD L, ENGELSEN S B, et al. Chemometrics in food science: a demonstration of the feasibility of a highly exploratory, inductive evaluation strategy of fundamental scientific significance [J]. Chemometr Intell Lab, 1998, 44(1/2): 31-60.
- [2] 赵瑜亮,仲山民. 油茶籽油与常见食用油的理化指标分析比较研究[J]. 安徽农业科学, 2014, 42(32): 11434-11436, 11439.
- [3] 余佳荣,晁燕,谭利娟. 基于气相色谱法的湖南油茶籽油杂判定[J]. 中南林业科技大学学报, 2014, 34(5): 88-91.
- [4] 杨柳,吴翠蓉,朱杰丽,等. 顶空-气质联用法鉴别油茶籽油真伪[J]. 中国粮油学报, 2012, 27(11): 105-109, 113.
- [5] ZHU M T, SHI T, CHEN Y, et al. Prediction of fatty acid composition in camellia oil by ^1H NMR combined with PLS regression[J]. Food Chem, 2019, 297(5): 339-346.
- [6] WANG L, LEE F S C, WANG X R, et al. Feasibility study of quantifying and discriminating soybean oil adulteration in camellia oils by attenuated total reflectance MIR and fiber optic diffuse reflectance NIR [J]. Food Chem, 2006, 95(3): 529-536.

- [7] LI S F, ZHU X R, ZHANG J H, et al. Authentication of pure camellia oil by using near infrared spectroscopy and pattern recognition techniques[J]. *J Food Sci*, 2012, 77(4): C374 – C380.
- [8] 邓平建,梁裕,杨冬燕,等. 基于拉曼光谱-聚类分析快速鉴别掺伪茶籽油[J]. *中国粮油学报*, 2016, 31(4): 72 – 75, 80.
- [9] 孙艳辉,吴霖生,翁长晟,等. 应用同步荧光光谱和支持向量机快速鉴别油茶籽油真伪[J]. *食品工业科技*, 2012, 33(4): 52 – 55.
- [10] WANG T, WU H L, LONG W J, et al. Rapid identification and quantification of cheaper vegetable oil adulteration in camellia oil by using excitation – emission matrix fluorescence spectroscopy combined with chemometrics[J]. *Food Chem*, 2019, 293(9): 348 – 357.
- [11] 鲁小利,王俊. 仿生电子鼻在芝麻油掺伪检测中的应用研究[J]. *粮食与油脂*, 2016, 29(6): 75 – 77.
- [12] 靳欣欣,潘立刚,李安. 稳定同位素质谱法鉴别芝麻油中掺杂大豆油、玉米油的研究[J]. *中国油脂*, 2020, 45(3): 32 – 37.
- [13] BRO R, SMILDE A K. Principal component analysis[J]. *Anal Methods – UK*, 2014, 6(9): 2812 – 2831.
- [14] BALLABIO D, CONSONNI V. Classification tools in chemistry. Part 1: linear models. PLS – DA[J]. *Anal Methods – UK*, 2013, 5(16): 3790 – 3798.
- [15] CHEN H, GENG D, CHEN T, et al. Second – derivative laser – induced fluorescence spectroscopy combined with chemometrics for authentication of the adulteration of camellia oil [J]. *CYTA – J Food*, 2018, 16(1): 747 – 754.
- [16] HSU C W, LIN C J. A comparison of methods for multiclass support vector machines [J]. *IEEE Trans Neural Netw*, 2002, 13(2): 415 – 425.
- [17] AI F F, BIN J, ZHANG Z M, et al. Application of random forests to select premium quality vegetable oils by their fatty acid composition[J]. *Food Chem*, 2014, 143(1): 472 – 478.
- [18] XU B C, ZHANG L X, WANG H, et al. Characterization and authentication of four important edible oils using free phytosterol profiles established by GC – GC – TOF/MS [J]. *Anal Methods – UK*, 2014, 6(17): 6860 – 6870.
- [19] BASHEER I A, HAJMEER M. Artificial neural networks: fundamentals, computing, design, and application[J]. *J Microbiol Meth*, 2000, 43(1): 3 – 31.
- [20] IZQUIERDO M, LASTRA – MEJIAS M, GONZALEZ – FLORES E, et al. Deep thermal imaging to compute the adulteration state of extra virgin olive oil[J/OL]. *Comput Electron Agric*, 2020, 171: 105290 [2021 – 10 – 26]. <https://doi.org/10.1016/j.compag.2020.105290>.
- [21] RASHVAND M, AKBARNIA A. The feasibility of using image processing and artificial neural network for detecting the adulteration of sesame oil [J]. *AIMS Agric Food*, 2019, 4(2): 237 – 243.
- [22] CAPAR T D, KAVUNCUOGLU H, YALCIN H, et al. Rheological analysis for detection of extra virgin olive oil adulteration with vegetable oils: predicting oil type by artificial neural network[J]. *Qual Assur Saf Crop*, 2019, 11(8): 687 – 699.
- [23] 石婷,陈倩,闫小丽,等. 核磁共振氢谱结合化学计量学快速检测掺假茶油[J]. *食品科学*, 2018, 39(22): 241 – 248.
- [24] 张东生,金青哲,王兴国,等. 基于脂肪酸组成甄别油茶籽油掺伪的研究[J]. *中国粮油学报*, 2015, 30(1): 124 – 128.
- [25] CHU X, WANG W, LI C Y, et al. Identifying camellia oil adulteration with selected vegetable oils by characteristic near – infrared spectral regions[J/OL]. *J Innov Opt Heal Sci*, 2017, 11(2): 1850006 [2021 – 10 – 26]. <https://doi.org/10.1142/S1793545818500062>.
- [26] 孙婷婷,刘剑波,沈银梅,等. 基于机器学习算法的食用植物油掺伪鉴别的研究进展[J]. *中国油脂*, 2021, 46(3): 103 – 108, 113.
- [27] 陈永忠,邓绍宏,陈隆升,等. 油茶产业发展新论[J]. *南京林业大学学报(自然科学版)*, 2020, 44(1): 1 – 10.
- [28] 国家林业和草原局. 中国林业和草原年鉴2019[M]. 北京:中国林业出版社, 2019.
- [29] 赵淑娟,郭平,万建春,等. 山茶油掺伪鉴别技术研究进展[J]. *食品工业*, 2018, 39(5): 300 – 303.
- [30] LIU J B, LIAO Z P, SUN T Y, et al. Fatty acids, triacylglycerol and Sn – 2 fatty acids distributions variations in seed oil from camellia cultivars[J]. *J Food Nutr Sci*, 2018, 6(6): 143 – 153.
- [31] 冯纳. 茶油脂肪酸和甘油三酯的分析及掺伪识别初步研究[D]. 长沙:中南林业科技大学, 2016.