

# 基于气相色谱法结合化学计量学识别微量植物油的比较研究

胡 昆,张成龙,杨瑞琴

(中国人民公安大学 侦查学院,北京 100032)

**摘要:**为准确识别法庭科学领域中微量油脂物证,给涉及微量植物油物证鉴定的相关案件提供技术支持,以遗留在不同载体上并在4、25、38℃下分别放置1、3、7、14、30、45、60 d的8种微量植物油(亚麻籽油、油茶籽油、菜籽油、玉米油、花生油、芝麻油、大豆油、葵花籽油)为研究对象,利用气相色谱技术测定其脂肪酸组成,以5种主要脂肪酸(十六烷酸、硬脂酸、油酸、亚油酸、亚麻酸)作为识别指标,结合化学计量学方法构建Fisher判别分析、卷积神经网络和随机森林3种植物油识别模型。结果表明:Fisher判别分析、卷积神经网络和随机森林3种模型均能实现对8种植物油的准确识别,其中随机森林模型能评估各脂肪酸对分类结果的重要性,且识别准确率最高,达98.2%。综上,随机森林模型参数设置简单,识别准确率高,有效解决了微量植物油种类识别困难的问题。

**关键词:**植物油;微量;脂肪酸;气相色谱;化学计量学

中图分类号:TS225.1;TS227 文献标识码:A 文章编号:1003-7969(2025)02-0123-05

## Comparative study on the identification of trace vegetable oils based on gas chromatography combined with chemometrics

HU Kun, ZHANG Chenglong, YANG Ruiqin

(School of Investigation, People's Public Security University of China, Beijing 100032, China)

**Abstract:** In order to accurately identify trace oil evidence in the field of forensic science, and provide technical support for relevant cases involving the identification of trace amounts of vegetable oil evidence, eight trace vegetable oils (flaxseed oil, oil-tea camellia seed oil, rapeseed oil, corn oil, peanut oil, sesame seed oil, soybean oil, and sunflower seed oil) left on different carriers and stored at 4, 25, 38℃ for 1, 3, 7, 14, 30, 45 d, and 60 d respectively were used as research object. The fatty acid composition was determined by gas chromatography, and five main fatty acids (hexadecanoic acid, stearic acid, oleic acid, linoleic acid, and linolenic acid) from eight vegetable oils were selected as identification indicators to construct three vegetable oil recognition models (Fisher discriminant analysis, convolutional neural network, and random forest) using chemometrics methods. The results showed that Fisher discriminant analysis, convolutional neural network, and random forest models could all achieve accurate recognition of eight vegetable oils, among which the random forest model could evaluate the importance of each fatty acid to the classification results, and the recognition accuracy was the highest, reaching 98.2%. In conclusion, the random forest model has simple parameter settings and high recognition accuracy, and can effectively solve the problem of difficult identification types of trace vegetable oils.

**Key words:** vegetable oil; trace; fatty acid; gas chromatography; chemometrics

收稿日期:2023-09-18;修回日期:2024-10-14

基金项目:中国人民公安大学刑事科学技术双一流创新研究专项(2023SYL06)

作者简介:胡 昆(1998),男,硕士研究生,研究方向为理化检验(E-mail)761982623@qq.com。

通信作者:杨瑞琴,教授(E-mail)Yangrui@ppsuc.edu.cn。

植物油是人体脂肪的主要来源,其影响人体膳食结构平衡、内分泌稳定和代谢综合征发生等<sup>[1]</sup>。

目前,我国常见的植物油有大豆油、花生油、玉米油、菜籽油、芝麻油、葵花籽油、油茶籽油和亚麻籽油等。脂肪酸是植物油中最重要的质量指标,气相色谱法因效能高、选择性强、操作简便等特点,是目前脂肪酸测定中应用最广泛的检测方法之一<sup>[2-3]</sup>。化学计量学方法是处理多变量数据的重要工具,其中主成分分析、聚类分析、判别分析等统计方法在多变量样本分类领域有着广泛应用。气相色谱法结合化学计量学可有效识别植物油种类。林晨等<sup>[4]</sup>运用气相色谱法对4种植物油的脂肪酸组成及相对含量进行分析,并结合化学计量学方法进行建模分析,预报率最高达92.6%,较好地实现了不同植物油的准确鉴别。Zhang等<sup>[5]</sup>通过气相色谱-质谱法检测5种植物油的脂肪酸组成及相对含量,结合有监督和无监督多变量统计方法建立了5种植物油掺假判别模型,结果表明,所建立的模型可用于检测5种植物油中掺入10%以上的其他油脂。

目前,基于脂肪酸组成及含量识别植物油种类的研究通常会利用检测到的所有脂肪酸作为识别依据,而在法庭科学领域,由于案件现场微量植物油物证中含量较低的脂肪酸的检测存在偶然性和不确定性,若仍将此类脂肪酸作为植物油的识别依据可能会产生错误的分类结果<sup>[6]</sup>。因此,本文以4种载体上于不同温度下放置不同时间的8种微量植物油为研究对象,以通过气相色谱法测定的5种主要脂肪酸〔十六烷酸(C16:0)、硬脂酸(C18:0)、油酸(C18:1)、亚油酸(C18:2)、亚麻酸(C18:3)〕作为识别指标,结合化学计量学方法构建植物油识别模型,以得到更加准确的识别结果,满足法庭科学领域对案件现场中微量植物油物证识别的实践需要。

## 1 材料与方法

### 1.1 实验材料

亚麻籽油、油茶籽油、菜籽油、玉米油、花生油、芝麻油、大豆油、葵花籽油,均购自淘宝商家。氢氧化钾、甲醇,分析纯;正己烷,色谱纯;26种脂肪酸甲酯混合标准品,中国计量科学研究院。

7890A气相色谱仪〔配备氢火焰离子化检测器(FID)〕,美国安捷伦科技公司。

### 1.2 实验方法

#### 1.2.1 样本准备

8种植物油(亚麻籽油、油茶籽油、菜籽油、玉米油、花生油、芝麻油、大豆油、葵花籽油)各取50  $\mu\text{L}$  分别滴加至直径2 cm的铁片、塑料片、布片、纸片上,每种载体上的每种植物油样本各42份(共1344个植物油样本),等量分别放置于4  $^{\circ}\text{C}$  冰箱、25  $^{\circ}\text{C}$  烘箱、

38  $^{\circ}\text{C}$  烘箱中1、3、7、14、30、45、60 d。将载体取出用10 mL正己烷浸泡5 min得到植物油正己烷溶液。

#### 1.2.2 脂肪酸组成的测定

随机取8种植物油样本(亚麻籽油10个、油茶籽油11个、菜籽油10个、玉米油12个、花生油11个、芝麻油9个、大豆油11个、葵花籽油12个)测定脂肪酸组成,平行测定3次,结果以“平均值 $\pm$ 标准偏差”表示。

甲酯化:取1.2.1中得到的植物油正己烷溶液于锥形瓶中,加入10 mL 0.4 mol/L 氢氧化钾-甲醇溶液,振荡30 s,室温反应30 min后,加入10 mL超纯水终止反应并振荡30 s,取上清液上机测定。

气相色谱条件:Agilent DB-225ms色谱柱(60 m  $\times$  0.25 mm  $\times$  0.25  $\mu\text{m}$ );载气为高纯氮气(99.999%);升温程序为柱温箱初始温度120  $^{\circ}\text{C}$ ,保持1 min,以3  $^{\circ}\text{C}/\text{min}$  的速率升温至175  $^{\circ}\text{C}$ ,保持15 min,以5  $^{\circ}\text{C}/\text{min}$  升温至235  $^{\circ}\text{C}$ ,保持30 min;进样口温度240  $^{\circ}\text{C}$ ;分流比10:1,流速1 mL/min;FID检测器温度260  $^{\circ}\text{C}$ ;进样量1  $\mu\text{L}$ 。

根据脂肪酸甲酯混合标准品的保留时间定性,采用峰面积归一化法定量。

#### 1.2.3 植物油识别模型构建

##### 1.2.3.1 Fisher判别分析模型

应用SPSS22.0软件将8种植物油中5种主要脂肪酸数据作为判别分析的自变量,8种植物油作为分组变量,定义范围为1~8,进行判别模型的拟合,输出数据得到典则判别函数特征值和典则判别函数系数,构建判别函数。

##### 1.2.3.2 卷积神经网络(Convolutional neural network, CNN)模型

CNN的基本结构如图1所示,主要由输入层、卷积层、池化层、全连接层和输出层构成<sup>[7]</sup>,其中卷积层是实现CNN特征提取功能的核心,相当于特征提取器,池化层能够压缩数据和参数的量。

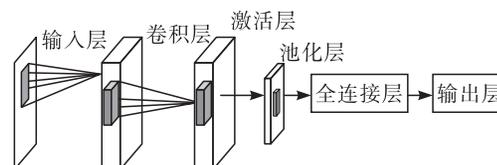


图1 CNN基本结构

Fig.1 Basic structure of CNN

采用Matlab软件将8种植物油中C16:0、C18:0、C18:1、C18:2、C18:3数据导入CNN中,采用75%的样本作为训练集,25%的样本作为测试集,并优化参数,建立分类识别模型。本次构建的CNN网

络结构由 1 个输入层、2 个卷积层、1 个激活层、2 个池化层、1 个全连接层和 1 个输出层组成,采用批归一化加快模型收敛速度,激活函数选择 Relu。设置初始学习率为 0.001,学习率下降因子为 0.1,最大迭代次数为 1 000,最终模型训练时间为 53 s。

### 1.2.3.3 随机森林算法(Random forest, RF)模型

采用 Matlab 软件搭建 RF 分类模型,将 8 种植

物油样本中的 C16:0、C18:0、C18:1、C18:2、C18:3 数据导入 RF 模型,选择 75% 的样本作为训练集,25% 的样本作为测试集,设置决策树数目为 50,最小叶子数位 1。

## 2 结果与讨论

### 2.1 8 种植物油的脂肪酸组成及含量

8 种植物油的脂肪酸组成及含量见表 1。

表 1 8 种植物油主要脂肪酸组成及含量

脂肪酸	亚麻籽油	油茶籽油	菜籽油	玉米油	花生油	芝麻油	大豆油	葵花籽油
十六烷酸(C16:0)	5.79 ± 0.03	7.67 ± 0.12	4.83 ± 0.07	12.71 ± 0.17	10.01 ± 0.22	8.80 ± 0.11	10.83 ± 0.08	6.03 ± 0.09
十六碳一烯酸(C16:1)	0.06 ± 0.02	0.12 ± 0.03	0.26 ± 0.02	0.08 ± 0.02	0.07 ± 0.02	0.11 ± 0.01	0.09 ± 0.02	0.81 ± 0.04
十七烷酸(C17:0)	0.05 ± 0.01	-	0.05 ± 0.01	0.06 ± 0.03	0.04 ± 0.02	-	0.13 ± 0.03	-
十七碳一烯酸(C17:1)	0.03 ± 0.01	0.07 ± 0.02	0.06 ± 0.01	-	0.04 ± 0.01	-	0.04 ± 0.01	-
硬脂酸(C18:0)	4.39 ± 0.05	2.23 ± 0.01	1.83 ± 0.01	1.59 ± 0.04	3.42 ± 0.06	5.79 ± 0.11	4.55 ± 0.08	3.73 ± 0.01
油酸(C18:1)	19.59 ± 0.05	78.46 ± 0.39	55.97 ± 0.41	26.33 ± 0.09	44.67 ± 0.08	38.42 ± 0.42	23.93 ± 0.31	30.02 ± 0.03
亚油酸(C18:2)	15.90 ± 0.04	8.92 ± 0.17	21.86 ± 0.23	57.79 ± 0.08	34.89 ± 0.21	45.71 ± 0.18	53.98 ± 0.23	56.84 ± 0.19
亚麻酸(C18:3)	53.62 ± 0.13	0.22 ± 0.02	8.32 ± 0.04	0.58 ± 0.02	0.08 ± 0.01	0.24 ± 0.02	4.97 ± 0.07	0.76 ± 0.04
二十烷酸(C20:0)	0.14 ± 0.02	0.09 ± 0.01	0.58 ± 0.02	0.40 ± 0.01	1.26 ± 0.05	0.56 ± 0.01	0.55 ± 0.03	0.27 ± 0.03
二十碳一烯酸(C20:1)	0.13 ± 0.01	0.48 ± 0.03	2.01 ± 0.04	0.24 ± 0.03	0.99 ± 0.01	0.11 ± 0.02	0.28 ± 0.03	0.16 ± 0.02
二十二烷酸(C22:0)	0.14 ± 0.01	0.18 ± 0.02	0.35 ± 0.02	0.13 ± 0.02	2.88 ± 0.03	0.22 ± 0.02	0.43 ± 0.02	1.03 ± 0.04
二十二碳一烯酸(C22:1)	-	0.08 ± 0.03	3.80 ± 0.12	-	0.07 ± 0.02	-	-	-
二十四烷酸(C24:0)	-	0.76 ± 0.07	-	-	1.47 ± 0.06	-	0.15 ± 0.03	0.24 ± 0.03
二十四碳一烯酸(C24:1)	-	0.65 ± 0.03	-	-	-	-	-	-

注: - 表示脂肪酸含量低于 0.01% 或未检出

Note: "-" Indicates that the fatty acid content is below 0.01% or not detected

从表 1 可以看出,8 种植物油共检测出 6 种饱和脂肪酸、6 种单不饱和脂肪酸和 2 种多不饱和脂肪酸,其中有 9 种脂肪酸为 8 种植物油的共有成分,分别为 C16:0、C16:1、C18:0、C18:1、C18:2、C18:3、C20:0、C20:1、C22:0。8 种植物油中含量较高的脂肪酸包括 C16:0、C18:0、C18:1、C18:2 和 C18:3 5 种脂肪酸。由于微量植物油物证中含量较低的脂肪酸的检测存在偶然性和不确定性,基于 8 种植物油中含量较高的 5 种脂肪酸(C16:0、C18:0、C18:1、C18:2、C18:3)作为识别指标来确定植物油的种类,以实现更加准确的识别结果。

### 2.2 Fisher 判别分析模型

Fisher 判别分析是一种处理概率分布未知的分类问题的线性判别分析方法,其基本思想是将所有数据投影到某一方向上,使得不同类别样本间差异最大化的同时实现同一类别内样本间差异的最小化。8 种植物油中 5 种主要脂肪酸的典则判别函数特征值见表 2。

由表 2 可知,5 种脂肪酸前 5 个典则判别函数的累积方差贡献率为 100%,表明 5 种脂肪酸所得模型能够对 8 种植物油进行判别。经分析得到 5 个

判别函数:  $Y_1 = 0.312X_1 + 0.421X_2 - 0.213X_3 + 0.128X_4 + 0.026X_5 - 0.452$ ,  $Y_2 = -0.379X_1 + 0.779X_2 + 0.005X_3 - 0.029X_4 + 0.241X_5 - 0.718$ ,  $Y_3 = 0.540X_1 - 0.585X_2 - 0.026X_3 - 0.057X_4 + 0.081X_5 - 0.186$ ,  $Y_4 = 0.190X_1 + 0.397X_2 + 0.059X_3 + 0.024X_4 + 0.012X_5 - 6.859$ ,  $Y_5 = 0.056X_1 + 0.010X_2 + 0.132X_3 + 0.131X_4 + 0.129X_5 - 10.914$ ,式中,  $X_1$ 、 $X_2$ 、 $X_3$ 、 $X_4$ 、 $X_5$  分别为 C16:0、C18:0、C18:1、C18:2、C18:3 含量。

表 2 典则判别函数特征值

Table 2 Eigenvalues of the canonical discriminant function

函数	特征值	方差 贡献率/%	累积方差 贡献率/%	典型 相关性
1	32.865	55.50	55.50	0.99
2	23.844	40.27	95.77	0.98
3	2.024	3.42	99.19	0.82
4	0.417	0.70	99.90	0.54
5	0.061	0.10	100.00	0.24

注:使用了前 5 个典则判别函数

Note: The first 5 canonical discriminant functions are used

由表 2 可知,前 2 个典则判别函数的累积方差贡献率达到了 95.77%,说明前 2 个函数可以解释样本分类,故选取函数 1 和函数 2 建立联合分布图,结果见图 2。

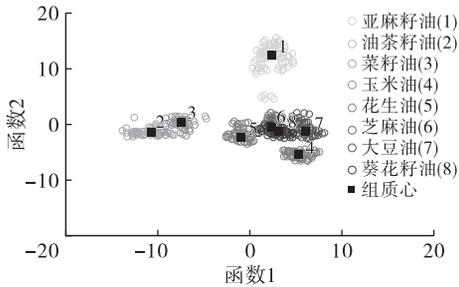


图2 联合分布图

Fig. 2 Joint distribution plot

由图2可知,除芝麻油和葵花籽油的质心有部分重合,其他植物油的质心分布较为分散。具体判别分类结果见表3。

表3 判别分类结果

Table 3 Discriminant classification results

组别	不同组别的识别结果								总计	准确率/%
	1	2	3	4	5	6	7	8		
1	155	0	0	0	0	5	0	0	160	96.9
2	0	159	6	0	3	0	0	0	168	94.6
3	0	1	167	0	0	0	0	0	168	99.4
4	0	0	0	168	0	0	0	0	168	100.0
5	0	0	0	0	168	0	0	0	168	100.0
6	0	0	0	0	1	163	0	4	168	97.0
7	0	0	0	0	0	2	166	0	168	98.8
8	0	0	0	0	4	21	2	141	168	83.9

注:组别1~8分别代表亚麻籽油、油茶籽油、菜籽油、玉米油、花生油、芝麻油、大豆油、葵花籽油。下同

Note: Groups 1 - 8 represent flaxseed oil, oil - tea camellia seed oil, rapeseed oil, corn oil, peanut oil, sesame seed oil, soybean oil, and sunflower seed oil, respectively. The same below

由表3可以看出,该模型对大部分植物油均实现了较准确的识别,整体识别准确率达96.3%,但葵花籽油的识别准确率仅83.9%。由于Fisher判别分析需要依托方差贡献率排名靠前的2个函数进行分析,当累积方差贡献率不足或脂肪酸含量接近时,会导致原始数据部分信息损失,在联合分布图上存在重叠以及部分样本识别准确率较低的情况,因此需要进一步建立识别准确率更高的识别模型。

2.3 CNN模型

CNN是一种多层前馈神经网络,是通过多层感知机(Multilayer perceptron, MLP)演变而来,具备良好的自学习特征能力,强大的并行处理能力,以及优秀的鲁棒性。CNN训练集和测试集混淆矩阵分别见图3和图4。

由图3和图4可知,CNN模型整体训练集和测试集准确率分别为99.0%、97.9%,8种植物油各自识别准确率均在92%以上,模型整体准确率高于Fisher判别分析识别模型,但在网络构建的复杂程度和参数

设置方面难度远高于Fisher判别分析识别模型。

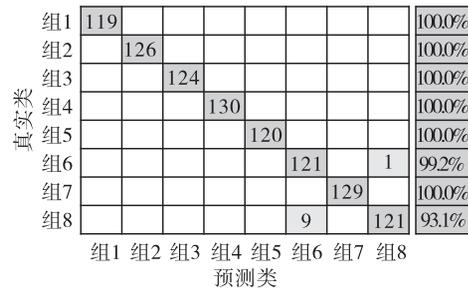


图3 CNN训练集混淆矩阵

Fig. 3 Confusion matrix of CNN training set

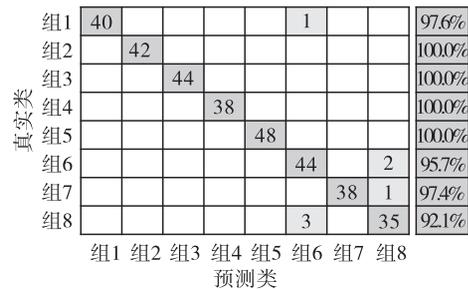


图4 CNN测试集混淆矩阵

Fig. 4 Confusion matrix of CNN test set

2.4 RF模型

RF是一种基于分类回归树的集成算法,即在变量(列)的使用和数据(行)的使用上进行随机化,生成很多分类树,再汇总分类树的结果。RF在运算量没有显著提高的前提下提高了预测精度。RF对多元共线性不敏感,对缺失数据和非平衡的数据比较稳健,可以很好地预测多达几千个解释变量的作用<sup>[8]</sup>。

RF误差曲线如图5所示,5种脂肪酸在分类过程中的重要性大小如图6所示。

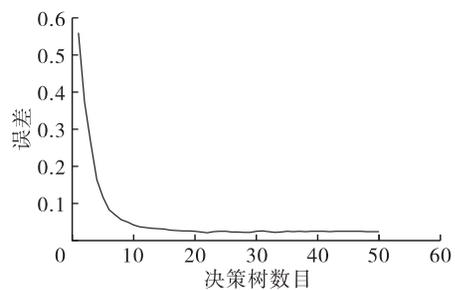


图5 RF误差曲线

Fig. 5 RF error curve

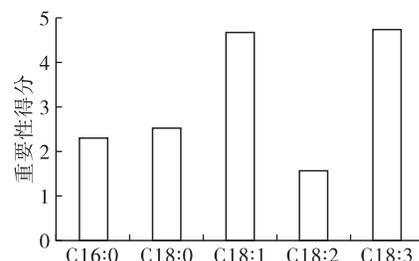


图6 5种脂肪酸的重要性大小

Fig. 6 Importance of five fatty acids

由图 5 可以看出,随着决策树数目的增加,误差逐渐减小。由图 6 可以看出,C18:1 和 C18:3 是 8 种植物油识别最重要的 2 个指标。

RF 训练集和预测集混淆矩阵如图 7 和图 8 所示。

真实类	组1	129							100.0%	
	组2		116						100.0%	
	组3			124					100.0%	
	组4				128				100.0%	
	组5					130			100.0%	
	组6						131		100.0%	
	组7							126	100.0%	
	组8								116	100.0%
		组1	组2	组3	组4	组5	组6	组7	组8	
		预测类								

图 7 RF 训练集混淆矩阵

Fig. 7 Confusion matrix of RF training set

真实类	组1	31							100.0%	
	组2		52						100.0%	
	组3			44					100.0%	
	组4				40				100.0%	
	组5					38			100.0%	
	组6						36	1	97.3%	
	组7							42	100.0%	
	组8		1	1	1	4		46	88.5%	
		组1	组2	组3	组4	组5	组6	组7	组8	
		预测类								

图 8 RF 测试集混淆矩阵

Fig. 8 Confusion matrix of RF test set

由图 7 和图 8 可以看出,8 种植物油训练集识别准确率均达到 100%,测试集识别准确率也达到 98.2%。该 RF 模型整体准确率高于一 Fisher 判别分析模型和 CNN 模型,并且训练模型参数设置简单,在实现 8 种植物油准确识别的同时,还能够评估各脂肪酸对分类结果的重要性。

### 3 结 论

本文立足于法庭科学领域中微量油脂物证识别的实践需要,主要针对微量植物油的识别展开研究,选择其中的 5 种主要脂肪酸(C16:0、C18:0、C18:1、C18:2、C18:3)作为识别指标,利用气相色谱技术结合化学计量学方法构建了 8 种微量植物油的识别模型。结果表明,建立的 Fisher 判别分析、卷积神经网络

和随机森林模型可以实现对低温(4℃)、常温(25℃)、高温(38℃)条件放置 1、3、7、14、30、45、60 d 的 8 种植物油的准确识别。不同识别模型构建的难易程度和识别准确率均有所不同,综合分析表明,随机森林算法具有更大的优势,识别准确率达 98.2%,模型参数设置简单,在实现 8 种微量植物油较准确识别的同时,还能够评估各脂肪酸对分类结果的重要性。研究所选前处理方法和检测方法高效、便捷,识别模型准确,有效解决了微量植物油种类识别困难的问题,为涉及微量植物油物证鉴定的相关案件提供技术支持和科学指南。后期,本研究还会进一步优化检验技术、拓宽样本研究范围、优化算法模型,以期对植物油种类识别和品质监测提供参考。

### 参 考 文 献:

- [1] SAAD B, LING C W, JAB M S, et al. Determination of free fatty acids in palm oil samples using non - aqueous flow injection titrimetric method [J]. Food Chem, 2007, 102(4): 1407 - 1414.
- [2] 韩洁, 杨清山, 王文昌, 等. GC - FID 内标法测定植物提取物中脂肪酸含量的方法研究 [J]. 中国食品添加剂, 2021, 32(6): 89 - 96.
- [3] 李刚, 康乐, 何建昇, 等. 基于核磁共振技术对亚麻籽油进行脂肪酸组成检测与品质鉴别 [J]. 中国油脂, 2022, 47(6): 148 - 152.
- [4] 林晨, 张方圆, 吴凌涛, 等. 气相色谱结合化学计量学分析 4 种食用植物油的指纹图谱 [J]. 分析测试学报, 2016, 35(4): 454 - 459.
- [5] ZHANG L, LI P, SUN X, et al. Classification and adulteration detection of vegetable oils based on fatty acid profiles [J]. J Agric Food Chem, 2014, 62 ( 34 ): 8745 - 8751.
- [6] 王岩, 黄浩博. 犯罪现场微量油脂物证酸败老化特征分析 [J]. 中国刑警学院学报, 2017(4): 78 - 83.
- [7] 潘斌, 韩强, 姚娅川. 基于卷积神经网络的白酒酒花分类研究 [J]. 食品与机械, 2021, 37(10): 30 - 37, 88.
- [8] 王升, 陈林, 周融融, 等. 基于气相色谱 - 质谱法和随机森林算法的不同产地香菇指纹图谱研究 [J]. 分析科学学报, 2019, 35(1): 85 - 89.