

# 基于近红外光谱与 KPCA-SVM 鉴别淡水鱼肉

周娇娇<sup>1,2</sup> 徐文杰<sup>1,2</sup> 许 竞<sup>3</sup> 尤 娟<sup>1,2</sup> 熊善柏<sup>1,2,4</sup>

1. 华中农业大学食品科学技术学院, 武汉 430070;

2. 国家大宗淡水鱼加工技术研发分中心(武汉), 武汉 430070;

3. 华中农业大学理学院, 武汉 430070; 4. 环境食品学教育部重点实验室, 武汉 430070

**摘要** 为实现淡水鱼品种的快速鉴别, 采用近红外光谱分析技术建立 7 种淡水鱼鲜肉的快速鉴别模型。试验采集了鲢、草鱼、乌鳢、鲫、鲤、青鱼、鳙 7 种淡水鱼共 772 个鲜鱼肉样品的近红外光谱数据, 分别考察标准正态变换 (standard normalized variate, SNV)、多元散射校正 (multiplicative signal correction, MSC) 的预处理方法及核主成分分析 (kernel principal component analysis, KPCA) 和主成分分析 (principal component analysis, PCA) 的特征提取方法对支持向量机 (support vector machine, SVM) 判别模型的影响。结果显示, 经 SNV 预处理和 KPCA 提取特征变量后, 对未知样品的整体正确判别率达到 92.68%。因此, 采用近红外光谱技术结合化学计量学方法所建 SVM 模型可以实现淡水鱼品种的快速鉴别。

**关键词** 近红外光谱; 特征提取方法; 淡水鱼肉; 水产品品质; 支持向量机; 鉴别; 正确判别率

**中图分类号** S 965.1 : O 657.33 **文献标识码** A **文章编号** 1000-2421(2019)05-0098-07

淡水鱼因其高蛋白、低脂肪、味道鲜美等优点, 已成为我国居民优质动物蛋白的重要来源<sup>[1]</sup>。目前, 国内外学者已对淡水鱼肌肉的营养特性<sup>[2-3]</sup>、凝胶特性<sup>[4-5]</sup>等方面进行了研究。对于淡水鱼鱼种的鉴别, 传统方法多利用机器视觉技术<sup>[6]</sup>以及肉眼判断规格均匀、体色正常、体表色泽、活动力强等特征<sup>[7]</sup>, 但是淡水鱼一经加工后, 难以对其种类进行准确鉴别, 为水产品品质控制带来困难<sup>[8]</sup>。因此, 探寻一种简便快速、客观准确的分析技术鉴别鲜鱼肉种类显得尤为重要。

近红外光谱 (near infrared spectroscopy, NIRS) 分析技术是利用物质对近红外光的选择性吸收及其吸收强度来预测其成分及含量, 主要用于有机物质定性和定量分析<sup>[9-10]</sup>。近红外谱区的信息主要是分子内部原子间振动的倍频与合频的信息, 几乎包括有机物中所有含氢基团 (如 C—H、O—H、N—H 和 C=O 等) 的信息, 信息量极为丰富<sup>[11-12]</sup>。目前, 采用近红外光谱技术分析鲜鱼肉的品种 (新鲜度、掺假等) 的报道较多<sup>[13-14]</sup>, 但关于鲜鱼品种鉴别的报道较少。Alamprese 等<sup>[15]</sup>采用红外光谱技术结合线性判别分析鉴别大西洋乌鱼和比目

鱼, 模型判别率达到了 100%。Reis 等<sup>[16]</sup>利用近红外光谱技术和主成分分析、偏最小二乘法结合线性判别分析建立了新鲜和冷冻-解冻金枪鱼鱼片无损鉴别模型, 结果表明, 对新鲜和冷冻-解冻金枪鱼鱼片的鉴别准确率分别达 92%、82%。Cozzolino 等<sup>[17]</sup>研究发现, 近红外光谱技术结合主成分分析和虚拟偏最小二乘法是快速、客观鉴别加工鱼粉用的鱼种的方法, 对未知样本的鉴别准确率达 80%。纵观现有研究, 运用近红外光谱技术并采用支持向量机来鉴别鲜鱼品种的较少。

支持向量机 (support vector machine, SVM) 由 Vapnik 等<sup>[18]</sup>提出, 是模式识别中广泛使用的强有力的分类工具。SVM 通过引入核函数将原始数据从输入空间映射到一个高维特征空间, 使线性不可分的问题变成普通的线性可分问题<sup>[19]</sup>。近年来, 其已成功用于肉类掺假<sup>[20]</sup>、苹果分类<sup>[21]</sup>、种子识别<sup>[22]</sup>等方面, 然而关于运用 SVM 方法鉴别不同种类的淡水鱼还未见报道。本研究以 7 种淡水鲜鱼肉为研究对象, 采用 SVM 结合不同近红外光谱预处理方法建立淡水鱼鱼种鉴别模型, 以为淡水鱼加工提供参考。

收稿日期: 2019-02-28

基金项目: 现代农业产业技术体系专项 (CARS-45-27); 湖北省技术创新专项重大项目 (2016ABA115)

周娇娇, 硕士研究生。研究方向: 水产品加工及贮藏工程。E-mail: 2236827292@qq.com

通信作者: 熊善柏, 教授。研究方向: 水产品加工及贮藏工程。E-mail: xionsb@mail.hzau.edu.cn

# 1 材料与amp;方法

## 1.1 试验材料

由于产地、季节和大小不同,鲜鱼肉样品的光谱曲线各异,为了扩大样品的采集范围,增加模型的适用性,分别从武汉、嘉鱼、黄冈等地点采集了不同季节、质量不等的鲜鱼肉样本共 772 个(鲢 120 个、草鱼 125 个、乌鳢 100 个、鲫 108 个、鲤 89 个、青鱼 130 个、鳙 100 个),以减少样本因素对试验结果的波动。

## 1.2 仪器与amp;设备

Supnir-2720 近红外光谱仪,杭州聚光科技股份有限公司;K600 食物调理机,德国博朗电器;BL220H 电子天平,日本岛津公司。

## 1.3 试验方法

1)光谱采集。将采集到的鲜活鱼样品于 10 ℃ 左右宰杀,取鱼背脊肉、绞碎、分装、随机编号后,置于半径为 3.5 cm、高度为 2.0 cm 的圆形黑色样品盒内,并用样品盒盖轻轻压平,排除样品盒底部的气泡。用 Supnir-2720 近红外光谱仪进行光谱采集,样品堆积厚度约为 2.0 cm。仪器检测参数为:扫描范围:1 000~1 799 nm;测量方式:漫反射;扫描次数:每个样本测量 3 次,再以 3 次的平均值作为最终的光谱值。

2)数据处理。数据预处理方法对建模精度影响较大,为确定最佳的光谱预处理方法,分别采用一阶导数(first derivate, 1st)、标准正态变换(standard normalized variate, SNV)、多元散射校正(multiplicative signal correction, MSC)对光谱进行预处理。对经预处理后的光谱曲线采用 SVM 方法建模,由于本研究是多分类问题,故直接使用 SVM 是不适合的。因此在设计 SVM 分类器时,为了得到较好的泛化能力,应使每个决策节点的类间隔尽可能地大,所以本研究基于类间可分性测度建立层次 SVM。即先将类间可分性测度较大的那类分离,再重复以上步骤直至分类完毕。这里分别采用最短距离、重心距离以及中间距离法来定义类与类之间的距离,建立基于重心距离、最短距离和中间距离的层次 SVM 判别模型<sup>[23]</sup>,最后通过外部验证考察模型的准确性和适应性。

本研究采用网格搜索法(grid searching technique)和类间距离可分性测度建立层次 SVM,即最短距离、重心距离以及中间距离法来定义类与类之间的距离。当进行 SVM 运算时,有 3 个关键问题需要解决:决定最优输入特征子集、核函数及其最佳

参数。径向基核函数能够减少训练过程中计算的复杂性。因此,取径向基函数作为核函数,其表达式为:

$$K(x, y) = \exp \{-\gamma \|x - y\|^2\} \quad (1)$$

以训练集五折交叉验证的平均正确率作为评价标准,使用网格搜索的方法寻找最优的参数值,确定径向基核函数参数,其表达式为:

$$\min f(X), X_s \leq X \leq X_f \quad (2)$$

式(2)中,  $X = (X_1, X_2, X_3, \dots, X_n)$  是  $n$  个自变量;  $X_s = (X_{1s}, X_{2s}, X_{3s}, \dots, X_{ns})$  是  $X$  的下限;  $X_f = (X_{1f}, X_{2f}, X_{3f}, \dots, X_{nf})$  是  $X$  的上限。

传统的多分类问题旨在追求高正确率,通常用错误率或正确率作为评价标准,经常会出现分类精度相同而分类结果却不同的问题,精度指标的缺陷在于不能区分分类错误所属的类别,这种评价标准在多分类问题上显然是不适用的。因此,笔者引入  $F_1$  评价模型<sup>[24]</sup>来研究多分类问题,其计算公式如下:

$$\text{recall} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{precision} = \frac{TP}{TP + FP} \quad (4)$$

$$F_1 = \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}} \quad (5)$$

式(3)~(5)中,TP 表示正类样本被正确预测为正类样本;FN 表示正类样本被错误预测为负类样本;FP 表示负类样本被错误预测为正类样本;查全率(recall)表示实际类别中被正确分类的实例个数;查准率(precision)表示预测类别中真正正确分类的实例个数; $F_1$  评价模型代表查准率与查全率有同样的重要性。本研究采用 20 次随机重复试验得到平均查准率、平均查全率,进而得到平均  $F_1$ 。

采用 Matlab 2016a (Mathworks Inc., Natick, MA) 软件进行数据分析。

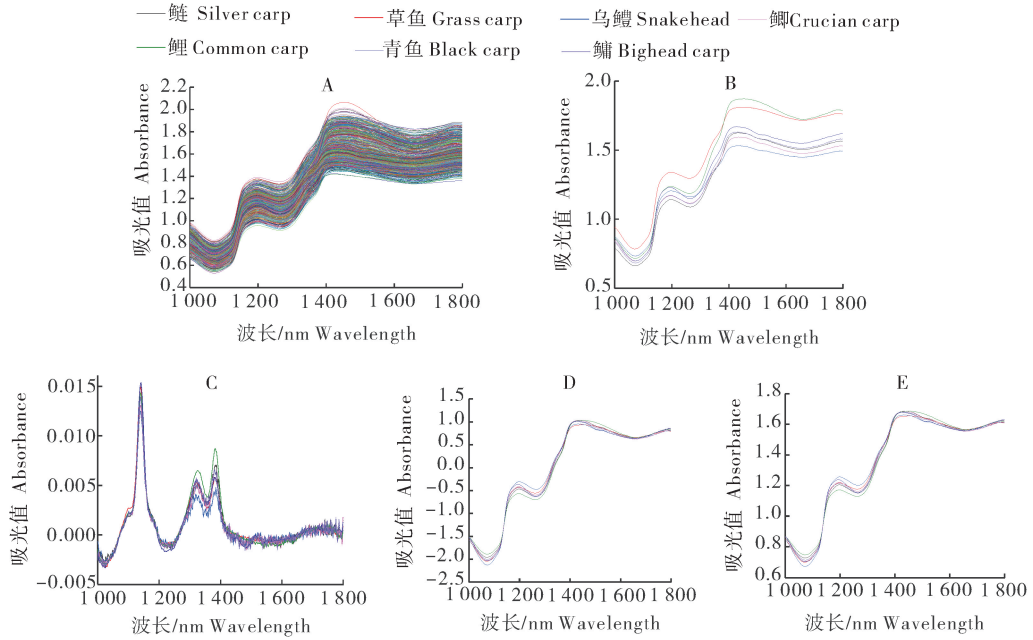
# 2 结果与分析

## 2.1 鲜鱼肉样品的光谱采集结果

采集不同规格、不同季节的 772 个鲜鱼肉样品的近红外光谱图及光谱的一阶导数图,见图 1。从图 1A 可以看出,不同品种的鱼肉样本的近红外光谱曲线变化趋势极其相似,只是强度有所差异,且光谱有交叉重合,存在比较明显的波谷和波峰,吸光度在 0.65~1.65,在样本较多时难以肉眼直接从光谱曲线上分辨鱼种类,因此,需要进行预处理。图 1B 是 7 种鱼的平均光谱图。图 1C 是 772 个样品原始光谱曲线经过一阶导数(1st)预处理后的光谱图。

从图 1C 可以看出,经 1st 预处理后,光谱基线漂移现象基本消除,局部干扰噪声也明显减少。1 130 nm 处的波峰与 1 210 nm 处的波谷,与 C—H 的伸缩振动有关,为样品中脂肪的吸收带;1 335 nm 和 1 379 nm 处较小的吸收带是由 C—H 的倍频和合频的伸缩振动及变形振动引起的;1 400 nm 左右处的波峰,与样品中 O—H 的一级伸缩振动有关,这

是因为鲜鱼肉中水分含量较大,所以 O—H 吸收峰很明显<sup>[14]</sup>。经 SNV 和 MSC 预处理后的光谱曲线如图 1D 和 1E 所示,光谱漂移现象基本消除,局部干扰噪声也明显减少,可见 2 种预处理方法均能有效消除基线漂移、光散射、样品分布不均等对光谱的干扰,滤掉原始光谱中无关信息,有效消除噪声、减少误差、提高光谱分辨率和灵敏度<sup>[25]</sup>。



A: 近红外光谱图 NIR spectra; B: 平均光谱图 Average spectra; C: 一阶导数图 First derivative spectra; D: 标准正态变换图 Standard normalized variate spectra; E: 多元散射校正图 Multiplicative signal correction spectra.

图 1 7 种淡水鱼鱼肉样品近红外光谱图、平均光谱图、一阶导数图、标准正态变换图及多元散射校正图

Fig.1 NIR spectra, average spectra, first derivative spectra, standard normalized variate spectra and multiplicative signal correction spectra of seven kinds of freshwater fish samples

## 2.2 判别模型的建立

1) 光谱预处理。本试验采集鲜鱼肉样品 772 个,近红外光谱采集了 800 个数据点,构成了  $772 \times 800$  的光谱矩阵。对样品进行 K-S 分类,80% 样本划分为定标集,其余则划分为验证集,结果如表 1 所示。

为了提高建模运算速度,减少运算量,利用核主成分分析(kernel principal component analysis, KPCA)和主成分分析(principal component analysis, PCA)对原始光谱输入变量进行降维压缩,提取原始信号的特征主成分。经不同光谱预处理后,所建 SVM 判别模型的预测效果如表 2 所示。由表 2 可知,不同的光谱预处理方法建立的 SVM 判别模型预测效果有所区别。PCA 提取特征向量后的判别正确率均低于 80%,KPCA 提取特征向量后的判

别正确率均高于 85%,主要原因是 PCA 对输入特征作线性变换,而 KPCA 通过核函数进行非线性变换<sup>[26]</sup>,本研究内容是淡水鱼鱼种鉴别,属于非线性判别,所以 KPCA 与 SVM 的结合能得到最优的特

表 1 校正集和验证集样本划分统计

Table 1 Division for calibration set and validation set of samples

鱼种 Fish species	校正集 Calibration set	验证集 Validation set	合计 Total
鲢 Silver carp	96	24	120
草鱼 Grass carp	100	25	125
乌鳢 Snakehead	80	20	100
鲫 Crucian carp	86	22	108
鲤 Common carp	71	18	89
青鱼 Black carp	104	26	130
鳊 Bighead carp	80	20	100

表 2 不同光谱预处理条件下的支持向量机的统计参数  
Table 2 Statistical parameters of SVM models under different spectral pretreatments

特征提取方法 Feature extraction methods	预处理方法 Preprocessing methods	校正集 Calibration						验证集 Validation					
		支持向量机 1 SVM1		支持向量机 2 SVM2		支持向量机 3 SVM3		支持向量机 1 SVM1		支持向量机 2 SVM2		支持向量机 3 SVM3	
		准确率/% Accuracy	标准差 Standard deviation	准确率/% Accuracy	标准差 Standard deviation	准确率/% Accuracy	标准差 Standard deviation	准确率/% Accuracy	标准差 Standard deviation	准确率/% Accuracy	标准差 Standard deviation	准确率/% Accuracy	标准差 Standard deviation
核主成分分析 KPCA	未预处理 Control	93.833	0.551	95.942	0.647	94.908	0.491	85.701	2.189	87.256	2.914	86.860	2.320
	1st	94.602	0.538	96.398	0.574	95.140	0.504	86.301	2.343	87.952	2.701	88.217	2.384
	SNV	99.092	0.327	97.892	0.376	99.383	0.306	91.463	1.835	91.098	2.664	91.860	1.961
	MSC	95.867	0.482	96.367	0.518	96.200	0.666	88.018	2.417	88.262	2.839	89.024	2.455
主成分分析 PCA	NONE	69.575	1.371	65.392	1.449	65.642	1.523	59.970	3.923	57.409	3.167	57.226	3.529
	1st	69.561	1.382	65.382	1.469	65.63	1.526	59.970	3.923	57.384	3.165	57.226	3.529
	SNV	72.692	1.675	67.233	1.365	68.075	1.450	62.317	3.736	57.134	3.654	57.348	4.667
	MSC	69.558	1.400	65.375	1.447	65.617	1.529	59.970	3.923	57.378	3.165	57.226	3.529

注：支持向量机 1：基于重心距离；支持向量机 2：基于最短距离；支持向量机 3：基于中间距离。Note: SVM1: Based on gravity distance; SVM2: Based on shortest distance; SVM3: Based on middle distance.



征提取结果。经 KPCA 提取特征向量后,建立的基于最短距离的层次 SVM 模型定标集和验证集均有最优的判别效果,对校正集正确判别率为 97.89%,对验证集正确判别率为 91.09%。因此,采用核主成分分析结合支持向量机技术建立的淡水鱼鱼种鉴别模型,其最佳的光谱预处理是 SNV。

2)SVM 判别模型的建立。基于  $F_1$  评价模型所

建 SVM 判别模型的判别效果如表 3 所示。由表 3 可知,无论是否经过光谱预处理,基于最短距离建立的模型的  $F_1$  值都是最大,这是因为基于最短距离建立的 SVM 模型使所有样本点到超平面的距离最小。经 SNV 预处理后,采用最短距离建立层次 SVM 判别模型的效果最好,相对于未经光谱预处理的建模结果, $F_1$  值提高了 11.33%。

表 3 不同光谱预处理条件下的支持向量机的  $F_1$  的评价指标的统计参数

Table 3 Statistical parameters of  $F_1$  evaluation indexes of SVM models under different spectral pretreatments

预处理方法 Preprocessing method	支持向量机 SVM	平均查全率/% Average recall	平均查准率/% Average precision	评价指标/% $F_1$
未预处理 Control	SVM1	86.314	86.748	86.530
	SVM2	86.327	88.468	87.384
	SVM3	86.327	86.327	86.327
一阶导数 1st	SVM1	88.413	89.016	88.713
	SVM2	88.436	89.180	88.806
	SVM3	88.436	88.436	88.436
标准正态变换 SNV	SVM1	90.429	90.754	90.591
	SVM2	90.448	92.145	91.289
	SVM3	90.448	90.448	90.448
多元散射校正 MSC	SVM1	87.916	88.888	88.399
	SVM2	89.040	90.544	89.786
	SVM3	89.040	89.040	89.040

注:平均查全率:20 次随机重复试验的平均查全率;平均查准率:20 次随机重复试验的查准率。Note: Average recall: Average recall of 20 randomized repetitive trials; Average precision: Average precision of 20 randomized repetitive trials.

### 2.3 判别模型的检验

为了检验所建立的最短距离层次 SVM 模型对外部未知样本的判别效果,选用 164 个未参与建模的样本,即验证集对所建 SVM 模型进行检验,结果如表 4 所示。由表 4 可知,经 SNV 预处理后所建的最短距离层次 SVM 判别模型,对鲢、草鱼和乌鳢的正确判别的  $F_1$  值均大于 95%,对鲫、鲤、青鱼和鳊的判别

率大于 85%,其中乌鳢的正确判别的  $F_1$  值为 100%,表明该模型对乌鳢的判别效果最好。每种鱼鉴别的正确个数之和除以总个数即为整体正确判别率,根据  $(21+31+21+18+17+23+21)/164$  得出该模型的整体正确判别率为 92.68%。因此,此模型可以将 7 种淡水鱼准确区分开来,外部验证结果证明 SVM 判别模型应用于淡水鱼品种的快速鉴别是可行的。

表 4 未知样品的鉴别结果

Table 4 Discrimination results of unknown samples

预测标签 Predicted lable	预测标签 Predicted lable							合计 Total	查全率/% Recall
	1	2	3	4	5	6	7		
1	21	0	0	0	0	0	0	21	100.0
2	0	31	0	1	0	0	0	32	96.9
3	0	0	21	0	0	0	0	21	100.0
4	0	1	0	18	0	0	0	19	94.7
5	1	0	0	0	17	1	1	20	88.5
6	0	0	0	0	3	23	0	26	88.5
7	0	0	0	0	0	4	21	25	84.0
合计 Total	22	22	21	19	20	28	22	164	
查准率/% Precision	95.5	96.9	100.0	94.7	85.0	82.1	95.5		
评价指标/% $F_1$	97.7	96.9	100.0	94.7	86.7	85.2	89.4		

注 Note: 1: 鲢 Silver carp; 2: 草鱼 Grass carp; 3: 乌鳢 Snakehead; 4: 鲫 Crucian; 5: 鲤 Common carp; 6: 青鱼 Black carp; 7: 鳊 Big-head carp.

## 3 讨论

结果表明,不同品种的鲜鱼肉样本的近红外光谱曲线变化趋势相似,且光谱有明显的重叠部分,单

从光谱图上难以直接鉴别出淡水鱼的种类。在样本数量较多以及温度、湿度等外界条件的影响下,光谱图会出现漂移现象,因此,需要对光谱进行预处理。1st 预处理后的光谱在 1 130 nm 处的波峰与 1 210

nm 处的波谷,与 C—H 的伸缩振动有关,是鲜鱼肉样品中脂肪的吸收带,在 1 335 nm 和 1 379 nm 处较小的吸收带是由 C—H 的倍频和合频的伸缩振动及变形振动引起的,在 1 400 nm 左右处的波峰,与样品中 O—H 的一级伸缩振动有关,这是因为鲜鱼肉中水分含量较大,所以 O—H 吸收峰很明显<sup>[14]</sup>。经 3 种预处理方法处理后,光谱基线漂移现象基本消除,局部干扰噪声也明显减少,有效消除仪器、样本背景、环境条件及其他因素的影响,滤掉原始光谱的无关信息变量,提高了模型的预测能力和稳定性。

SVM 具有泛化性能好、防止过拟合、适于非线性分析等特点,能够较好地解决小样本、高维数等实际问题,将其应用于近红外光谱分析中,建立近红外光谱预测模型具有重要意义。不同的特征提取方法和光谱预处理方法对建模都有影响。本研究结果表明,KPCA 提取特征向量的效果比 PCA 好,主要原因是 PCA 对输入特征作线性变换,而 KPCA 通过核函数进行非线性变换。本研究内容是淡水鱼鱼种鉴别,属于非线性判别,所以 KPCA 的提取效果较好。KPCA 结合基于最短距离的层次 SVM 建立的模型定标集和验证集均有最优的判别效果,对校正集正确判别率为 97.89%,对验证集正确判别率为 91.09%。Yu 等<sup>[25]</sup>运用近红外光谱技术结合 SVM 鉴别葡萄酒的产地时,发现 SNV 预处理能使鉴别的正确率达 98.36%。王文秀等<sup>[26]</sup>利用标准正态变量变换预处理的二维相关可见-近红外光谱,建立猪肉新鲜度的 SVM 判别模型,总体正确率为 98.28%,其结果进一步证明了近红外光谱技术结合 SNV 预处理方法建立 SVM 的判别模型的可行性。Alamprese 等<sup>[27]</sup>比较了不同光谱预处理方法(MSC、SNV、1st 以及二阶导数)对新鲜和冷冻的火鸡肉的偏小二乘判别模型的影响,发现 SNV 是最佳的预处理方法。Ma 等<sup>[28]</sup>研究了山药粉中的掺假玉米粉和小麦粉,发现最佳预处理方法是平滑、导数和 MSC 的结合。由此可见,不同的分析对象有不同的最佳光谱预处理方法。

查准率与查全率在评价模型时有着同样的重要性, $F_1$  值计算方法说明, $F_1$  值往往接近于二者中的较小值。当查全率和查准率都大时, $F_1$  值才会相应增大,故  $F_1$  值是分类问题中有效的评价准则。本研究建立的近红外光谱预测模型对外部样品鉴别的  $F_1$  值均达到了 85% 以上。同时该模型对未知样品的整体正确判别率达到 92.68%,具有很好的鉴别效

果,证实了将近红外光谱技术应用于淡水鱼品种鉴别的可行性。

## 参 考 文 献

- [1] 赵永峰,胡海彦,蒋高中,等. 我国大宗淡水鱼的发展现状及趋势研究[J]. 中国渔业经济,2012,30(5):91-99.
- [2] 姜启兴. 鳊鱼热加工特性及其机理研究[D]. 无锡:江南大学,2015.
- [3] 吕广英. 白鲢鱼骨酶解浓汤的制备及风味增强技术研究[D]. 武汉:华中农业大学,2012.
- [4] 关宏,丁玉琴,尤娟,等. pH 值对添加 L-精氨酸的草鱼糜凝胶特性的影响[J]. 华中农业大学学报,2018,37(5):110-116.
- [5] 李维,程荻,杨宏,等. 二次加热处理对鲢鱼糜制品品质的影响[J]. 华中农业大学学报,2019,38(2):94-101.
- [6] 张志强,牛智有,赵思明,等. 基于机器视觉技术的淡水鱼质量分级[J]. 农业工程学报,2011,27(2):350-354.
- [7] 何金宝. 如何鉴别鱼种质量优劣[J]. 渔业致富指南,2009(23):30-30.
- [8] LIU D, ZENG X A, SUN D W. NIR spectroscopy and imaging techniques for evaluation of fish quality: a review[J]. Applied spectroscopy reviews, 2013, 48(8):609-628.
- [9] 冯放. 现代近红外光谱分析技术及其应用[J]. 生命科学仪器, 2007, 5(10):9-13.
- [10] 吴习宇,赵国华,祝诗平. 近红外光谱分析技术在肉类产品检测中的应用研究进展[J]. 食品工业科技,2014,35(1):371-374.
- [11] LV H, XU W J, YOU J, et al. Classification of fresh water fish species by linear discriminant analysis based on near infrared reflectance spectroscopy[J]. Journal of near infrared spectroscopy, 2017, 25(1):54-62.
- [12] LIU D, SUN D W, ZENG X A. Recent advances in wavelength selection techniques for hyperspectral image processing in the food industry[J]. Food & bioprocess technology, 2014, 7(2):307-323.
- [13] KIMIYA T, SIVERTSEN A H, HEIA K. VIS/NIR spectroscopy for non-destructive freshness assessment of Atlantic salmon (*Salmo salar* L.) fillets[J]. Journal of food engineering, 2013, 116(3):758-764.
- [14] ZHOU J J, WU X Y, CHEN Z, et al. Evaluation of freshness in fresh water fish based on near infrared reflectance spectroscopy and chemometrics [J]. LWT-food science and technology, 2019, 106:145-150.
- [15] ALAMPRESE C, CASIRAGHI E. Application of FT-NIR and FT-IR spectroscopy to fish fillet authentication[J]. LWT-food science and technology, 2015, 63(1):720-725.
- [16] REIS M M, MARTINEZ E, SAITUA E, et al. Non-invasive differentiation between fresh and frozen/thawed tuna fillets using near infrared spectroscopy (Vis-NIRS)[J]. LWT-food science and technology, 2017, 78:129-137.
- [17] COZZOLINO D, CHREE A, SCAIFE J R, et al. Usefulness of near-infrared reflectance (NIR) spectroscopy and chemometrics to discriminate fishmeal batches made with different fish spe-

- cies[J]. *Journal of agricultural and food chemistry*, 2005, 53: 4459-4463.
- [18] VAPNIK, VLADIMIR N. The nature of statistical learning theory[J]. *IEEE transactions on neural networks*, 1997, 38(4): 409-409.
- [19] ANGUIA D, RIDELLA S, RIVIECCIO F, et al. Hyperparameter design criteria for support vector classifiers[J]. *Neurocomputing*, 2003, 55(1): 109-134.
- [20] 张丽华, 郝莉花, 李顺峰, 等. 基于支持向量机的近红外光谱技术快速鉴别掺假羊肉[J]. *食品工业科技*, 2015, 36(23): 289-293.
- [21] 赵杰文, 呼怀平, 邹小波. 支持向量机在苹果分类的近红外光谱模型中的应用[J]. *农业工程学报*, 2007, 23(4): 149-152.
- [22] 彭彦昆, 赵芳, 李龙, 等. 利用近红外光谱与 PCA-SVM 识别热损伤番茄种子[J]. *农业工程学报*, 2018, 34(5): 159-165.
- [23] TAKAHASHI F, ABE S. Decision-tree-based multiclass support vector machines[C]. *Proceedings of the 9th international conference on neural information processing*, 2002, 3(3): 1418-1422.
- [24] GASPARINI F, CORCHS S, SCHETTINI R. Recall or precision-oriented strategies for binary classification of skin pixels [J/OL]. *Journal of electronic imaging*, 2008, 17(2): 023017 [2019-02-28]. <https://doi.org/10.1117/1.2916715>.
- [25] YU J, ZHAN J, HUANG W. Identification of wine according to grape variety using near-infrared spectroscopy based on radial basis function neural networks and least-squares support vector machines[J]. *Food analytical methods*, 2017, 10(10): 3306-3311.
- [26] 王文秀, 彭彦昆, 孙宏伟, 等. 二维相关可见/近红外光谱结合支持向量机评价猪肉新鲜度[J]. *食品科学*, 2018, 39(18): 273-279.
- [27] ALAMPRESE C, AMIGO J M, CASIRAGHI E, et al. Identification and quantification of turkey meat adulteration in fresh, frozen-thawed and cooked minced beef by FT-NIR spectroscopy and chemometrics[J]. *Meat science*, 2016, 121: 175-181.
- [28] MA H L, WANG J W, CHEN Y J, et al. Rapid authentication of starch adulterations in ultrafine granular powder of Shanyao by near-infrared spectroscopy coupled with chemometric methods[J]. *Food chemistry*, 2017, 215: 108-115.

## Identification of freshwater fish species based on near infrared spectroscopy and KPCA-SVM method

ZHOU Jiaojiao<sup>1,2</sup> XU Wenjie<sup>1,2</sup> XU Jing<sup>3</sup> YOU Juan<sup>1,2</sup> XIONG Shanbai<sup>1,2,4</sup>

1. *College of Food Science and Technology, Huazhong Agricultural University, Wuhan 430070, China;*

2. *National R&D Branch Center for Conventional Freshwater Fish*

*Processing (Wuhan), Wuhan 430070, China;*

3. *College of Science, Huazhong Agricultural University, Wuhan 430070, China;*

4. *Environmental Food Science Key Laboratory of Ministry of Education, Wuhan 430070, China*

**Abstract** To realize the rapid identification of freshwater fish species, near infrared reflectance spectroscopy was employed to establish the identification models of fish species. 772 samples of 7 freshwater fish species (silver carp, grass carp, snakehead, crucian carp, common carp, black carp, bighead carp) were prepared to collect near infrared spectra data. The effects of preprocessing methods including standard normalized variate (SNV), multiple scattering correction (MSC) and the feature extraction methods including kernel principal component analysis (KPCA) and principal component analysis (PCA) on the discrimination models of support vector machine (SVM) were investigated, respectively. The results showed that the overall accuracy rate was 92.68% for the unknown sample after the SNV preprocessing and KPCA extraction of characteristic variables. Therefore, the SVM model constructed by near infrared spectroscopy combined with chemometric methods is feasible for rapid identification of freshwater fish species.

**Keywords** near infrared spectrum; feature extraction method; freshwater fish; aquatic product quality; support vector machine; identification; accuracy rate

(责任编辑:赵琳琳)