

文章编号: 2095-4980(2024)04-0385-09

太赫兹光谱结合机器学习的甜味剂检测

钟芸襄^a, 张 然^a, 熊子仪^a, 邹 斌^{a,b}, 杨玉平^{*a,b}

(中央民族大学 a.理学院; b.光子系统工程软件教育部工程研究中心, 北京 100081)

摘 要: 以三氯蔗糖、赤藓糖醇、木糖醇 3 种人工甜味剂为研究对象, 采用太赫兹时域光谱技术, 结合多种机器学习和优化算法对甜味剂与面粉混合物的光谱数据进行系统的分类识别和定量回归研究。结果表明, 麻雀搜索算法-支持向量机模型/支持向量回归模型(SSA-SVM/SVR)对混合物的定性及定量分析结果均达到最优, 分类预测的准确率达到 95.56%, 定量回归预测的最佳回归系数 R^2 为 0.999 8, 实现了 3 种甜味剂和面粉混合物的高精确度分类和定量分析, 为人工甜味剂的快速检测提供了一种有效可靠的新思路。

关键词: 太赫兹; 甜味剂; 机器学习; 优化算法

中图分类号: O433

文献标志码: A

doi: 10.11805/TKYDA2023284

THz spectroscopic detection of sweeteners based on machine learning algorithms

ZHONG Yunxiang^a, ZHANG Ran^a, XIONG Ziyi^a, ZOU Bin^{a,b}, YANG Yuping^{*a,b}

(a.School of Science; b.Engineering Research Center of Photonic Design Soft, Minzu University of China, Beijing 100081, China)

Abstract: Three artificial sweeteners, sucralose, erythritol and xylitol, are qualitatively and quantitatively studied based on Terahertz time-domain spectroscopy combined with machine learning algorithms and optimization algorithms. The results show that the Sparrow Search Algorithm-Support Vector Machines/Support Vector Regression(SSA-SVM/SVR) model is optimal for qualitative and quantitative analysis of the mixture. The accuracy of classification prediction is up to 95.56%, and the optimal regression coefficient for quantitative regression prediction is 0.999 8, so that a high-precision classification and quantitative analysis of three sweetener-flour mixtures is achieved. This provides an effective and reliable method for the rapid detection of artificial sweeteners.

Keywords: terahertz; sweeteners; machine learning; optimization algorithms

食品添加剂包括天然食品添加剂和人工合成添加剂, 可用于改善食品的味道、增添营养或延长食品的保质期。其中三氯蔗糖、赤藓糖醇、木糖醇为 3 种人工合成的甜味剂, 它们进入人体后基本不参与糖代谢, 也不会迅速升高体内的血糖, 较为适宜糖尿病人食用; 并且由于其热量极低, 也广泛用于代糖产品中。研究表明, 人工甜味剂可通过污水灌溉、污泥施肥、粪肥返田等农业活动进入土壤^[1], 其中三氯蔗糖因其在环境中表现出的持久性和稳定性, 被视为一种新兴的持久性环境污染物。食品添加剂的传统检测方法主要有色谱技术检测法^[2]、分子光谱技术检测法^[3]和 X 射线粉末衍射法^[4], 但这些检测技术都存在检测过程繁琐、时间冗长等问题, 因此迫切需要找到一种快速有效的检测食品添加剂的方法。

太赫兹(THz)波的能量在 meV 量级, 对待测物质无光损伤和光化电离, 可以保留被测物的原始结构^[5]; THz 波穿透力强, 可以对物质内部的化学键进行振动来源分析^[6]; THz 波光谱范围广, 能涵盖大多数物质的旋转振荡频率, 在太赫兹波段形成特有的“指纹谱”。因此将太赫兹时域光谱技术用于物质鉴别已成为一种新兴的检测手段, 并在食品安全检测领域显示出极大的应用前景, 从新的视角为进一步快速检测糖类等食品添加剂提供了新

收稿日期: 2023-09-27; 修回日期: 2023-11-22

基金项目: 国家自然科学基金资助项目(62075248); 国家重点基础研究发展计划资助项目(2020YFB2009303; 2017YFB00405402); 国家外国专家资助项目(G2022184001L)

*通信作者: 杨玉平 email: ypyang@muc.edu.cn

的有利手段^[7-10]。但实际测量过程中,由于被测样本中存在多种组分,对太赫兹波有很强的吸收,会造成光谱测量结果的偏差。因此,将太赫兹时域光谱技术与机器学习算法模型、数据处理算法相结合,以提升分析的准确率,实现对物质的分类和定量分析。2017年,LI等^[11]将遗传算法(Genetic Algorithm, GA)与支持向量机(SVM)结合,对4种不同产地的绿茶进行识别,准确率可达96.25%,证明了机器学习可以对绿茶产地来源进行快速识别。2018年,李鹏鹏等^[12]利用一元线性回归(Simple Linear Regression, SLR)、偏最小二乘回归(Partial Least Squares Regression, PLSR)对麦芽糖和三聚乙炔的混合物进行THz光谱研究,通过结合小波变换和线性回归模型,实现了对麦芽糖的定量分析。LONG等^[13]利用反向传播神经网络(Back Propagation Neural Network, BPNN)对饲料中3种氟喹诺酮类抗生素的THz光谱进行分类识别,检测精确度可达80.56%;并对这3种抗生素进行了定量分析,建立了多元线性回归模型,其定量回归相关系数分别为0.867、0.828和0.964。2019年,SUN等^[14]以小麦粉中的苯甲酸为研究对象,通过LS(Least Square)-SVM方法进行定量分析,预测集相关系数高达0.994,而均方根误差(Root Mean Square Error, RMSE)仅为0.12%。2020年,LIAN等^[15]研究了奶粉中山梨酸钾含量的太赫兹光谱,并结合PLS实现单种物质成分的定量检测。胡军等^[16-17]通过比较PLS、SVM、BPNN和广义回归神经网络(Generalized Regression Neural Network, GRNN)4种定量检测算法在奶粉中三聚氰胺及面粉中苯甲酸的定量分析结果,BPNN精确度最高,且相关系数均达到了0.99以上。

为提高模型的准确度,研究者进一步将优化算法与分类算法模型相结合,并比较多种算法得出一个最优模型。2021年,XIAO等^[18]比较了PLS、BPNN、GA-SVR、改进灰狼算法(Grey Wolf Optimizer, GWO)-支持向量回归(SVR)4种定量回归模型在大豆蛋白质定量检测中的表现,发现BPNN结合相关预处理可得到最理想的检测精确度,检测相关系数高达0.9677。同年,CHEN等^[19]采用主成分分析(Principal Component Analysis, PCA)与t-SNE(t-distributed Stochastic Neighbor Embedding)方法相结合提取光谱特征,结合流形学习算法和改进SVM的太赫兹光谱法,识别香豆素类食品添加剂。其中基于差分进化(Differential Evolution, DE)和GWO的混合优化算法(DEGWO)模型的预测集精确度达到98.61%,对建立高效、准确的食物添加剂鉴别模型有很好的参考价值。2022年,PAN等^[20]利用太赫兹时域光谱对6级成虫浓度的红参进行实验,通过PCA和蒙特卡罗非信息变量消除(Monte Carlo Non-information Variable Elimination, MCNVE)分别提取太赫兹光谱数据的信息,然后通过PLSR和SVR分别进行定量分析。由于考虑了红参样品太赫兹光谱曲线中的非线性因子,结果表明MCNVE-SVR具有0.99的高相关系数。大量研究结果表明,基于太赫兹光谱技术,选用合适机器学习算法和优化算法可准确实现食品中添加剂的定量检测。

本文以3种适合糖尿病人食用的甜味剂(三氯蔗糖、木糖醇、赤藓糖醇)作为研究对象,采用太赫兹时域光谱技术和有效介质理论,对3种甜味剂进行指纹谱测试和光学参数提取,得到3种甜味剂的吸收系数和折射率;并进一步对甜味剂与面粉混合的太赫兹光谱图进行研究,结合优化算法和分类/回归算法模型对混合品进行分析、建模,实现混合物的定性识别和定量分析,为食品安全检测手段提供新的参考。

1 实验与分析方法

1.1 样品介绍

实验所用的3种甜味剂样本(三氯蔗糖、赤藓糖醇、木糖醇)均来自于阿拉丁试剂(上海)有限公司;面粉购自超市。其中,三氯蔗糖(俗称蔗糖素)是一种高倍甜味剂,稳定性高,极易溶于水、甲醇和乙醇,通常为淡黄色粉末;木糖醇是从植物原料中提取出的一种天然甜味剂,极易溶于水,微溶于乙醇与甲醇,通常为白色晶体或结晶粉末;藓糖醇是一种填充型甜味剂,极易溶于水,微溶于醇,通常为白色结晶粉末。高密度聚乙烯(PE)在太赫兹波段基本上是透明的,故采用聚乙烯作为填充材料,在一定程度上起到拓宽实验光谱测量范围的作用。

1.2 压片制样

3种粉末状样品与PE搅拌均匀后,用压强为10 MPa的压片机将其压成直径为13 mm、厚度为1.2 mm的平整圆形压片。因面粉颗粒较大且不均匀,需经过300目的网筛并在40℃下进行烘干以除去水分。为进一步探究3种甜味剂与面粉混合物的太赫兹光谱,将3种甜味剂与面粉以100%、30%、20%、15%、10%、8%、5%、3%、1%、0%的比例进行混和,并搅拌均匀,然后用与上述同样的方法进行压片。将压好的圆片样品放入太赫兹时域光谱系统的样品室内,充入氮气,使样品室内的湿度降到10%以下再进行测量。

1.3 光学参数提取

使用透射式太赫兹时域光谱系统检测样品时,采集到的信号均为样品的时域数据。待测样品和参考样品的

时域脉冲波形分别为 $E_S(t)$ 和 $E_R(t)$ ，对其进行快速傅里叶变换后，得到对应频域信号 $E_S(\omega)$ 和 $E_R(\omega)$ 。当太赫兹波垂直入射到样品表面时，其透射函数为 $T(\omega) = E_S(\omega)/E_R(\omega)$ ，样品的复折射率可表示为 $\mathbf{n}(\omega) = n(\omega) - ik(\omega)$ ，其中 $\mathbf{n}(\omega)$ 为复折射率、 $n(\omega)$ 为折射率、 $k(\omega)$ 为样品的消光系数。根据 Duvillaret 和 Dorney 等提出的太赫兹光学参数提取模型，获得所测混合压片样品的光学参数， $k(\omega)$ 可表示为^[21-22]：

$$k(\omega) = \frac{c}{\omega d} \ln \left\{ \frac{4n(\omega)}{\rho(\omega)[n(\omega) + 1]^2} \right\} \quad (1)$$

式中： ω 为太赫兹波的角频率； c 为真空中光速； $\rho(\omega)$ 为样品信号和参考信号比值的模； d 为样品厚度。

吸收系数 $\alpha(\omega)$ 可表示为：

$$\alpha(\omega) = \frac{2\omega}{c} k(\omega) = \frac{2}{d} \ln \left\{ \frac{4n(\omega)}{\rho(\omega)[n(\omega) + 1]^2} \right\} \quad (2)$$

样品折射率 $n(\omega)$ 可表示为：

$$n(\omega) = \varphi(\omega) \frac{c}{\omega d} + 1 \quad (3)$$

式中 $\varphi(\omega)$ 为样品信号和参考信号比值的相位。

进一步采用 Maxwell-Garnett(MG) 提出的有效介质理论，提取 3 种甜味剂纯品的光学常数^[23]。根据 MG 有效介质理论，假设复合材料是对称的立方点阵结构，将立方点阵结构的样品看作阵点物质与空气的复合体。在计算整个复合体的介电函数时，假设颗粒是球形的，且仅考虑颗粒间相互耦合作用。利用文献[24]中的分析方法，可以计算出被分析物本身的介电常数，进而获取折射率、消光系数和吸收系数等光学参数。

1.4 机器学习模型与优化算法

运用 SVM 分类算法和 SVR 回归算法对 3 种甜味剂进行分类和定量研究，再利用 GA、网格搜索算法(Grid Search, GS)和 SSA 算法对 SVM 进行参数寻优，对比得出一个最优的分析模型，实现 3 种甜味剂与面粉混合物的定性识别。SVM 是 20 世纪 90 年代初 Vapnik 等^[25]根据统计学理论提出的一种二分类模型机器学习方法，适用于解决小样本、非线性及高维度问题，SVM 的惩罚参数 c 和核参数 g 对 3 种甜味剂与面粉混合物的分类结果有很大影响。本文采用的评价模型方法主要基于模型分类的准确率，即训练集和测试集的正确分类样本数和预测样本数的比值。

SVR 模型与 SVM 分类方法不同，其在构造最优超平面的基础上额外加入一个变量，形成与超平面平行的上下平面。2 个平面将采样点夹在中间，同时最小化两者之间的距离。由于其强大的线性和非线性回归分析能力，SVR 回归方法已成功用于物质的定量分析。本文主要选取决定系数 R^2 和均方根误差(RMSE)作为定量模型预测性能的评价指标： R^2 表示预测值与真实值总变化的百分比，衡量模型的整体拟合程度， R^2 值越接近于 1，表面模型的拟合程度越高；RMSE 用于评估预测值与真实值之间的误差，RMSE 值越接近于 0，表面模型的预测偏差越小。

2 3 种甜味剂的 THz 光谱测试与分析

2.1 3 种甜味剂压片的 THz 光谱

3 种甜味剂的太赫兹时域波形和快速傅里叶变换频谱如图 1 所示。在 0.50~2.50 THz 范围内，三氯蔗糖的频谱强度最大，吸收最小，频谱宽度最大，在 1.00~2.50 THz 范围内有 3 个明显的、较窄的吸收谷出现；木糖醇的吸收最大，频谱宽度最小，在 1.50 THz 之后几乎被全部吸收；赤藓糖醇的频谱强度介于二者之间，在 1.50~2.50 THz 范围内有 2 个较宽的强吸收带出现。由于本实验采用的测试系统频谱范围较宽，在 0~3 THz 范围内，基于以上特征信息可以对 3 种甜味剂进行初步的定性鉴别。

为更加高效、准确地进行鉴别，利用 MG 有效介质理论，进一步提取混合压片中 3 种甜味剂纯品的吸收系数和折射率，如图 2 所示。在 0.50~2.50 THz 范围内，三氯蔗糖在 0.98 THz、1.17 THz、1.54 THz 和 2.21 THz 处有 4 个明显的吸收峰；赤藓糖醇在 1.77 THz 和 1.99 THz 附近存在 2 个较强、较宽的吸收峰，且两者叠加，导致在 1.50~2.30 THz 范围内存在强吸收带；木糖醇的吸收最大，且无特征峰。三者的折射率中，三氯蔗糖的折射率最大，且在 4 个吸收峰处均存在较强的色散效应；木糖醇的折射率曲线较为平坦，平均值约为 1.4；赤藓糖醇的折射率曲线波动较大，在低频处随着频率的增加而增大，在 1.00~1.70 THz 范围内，赤藓糖醇的折射率大于木糖醇；

由于 2 个强吸收带的共振, 导致 1.60~2.30 THz 范围内的折射率存在较强的色散效应。结合数据处理得到的吸收系数和折射率谱, 可以更加直观、显著地实现 3 种甜味剂的定性鉴别。

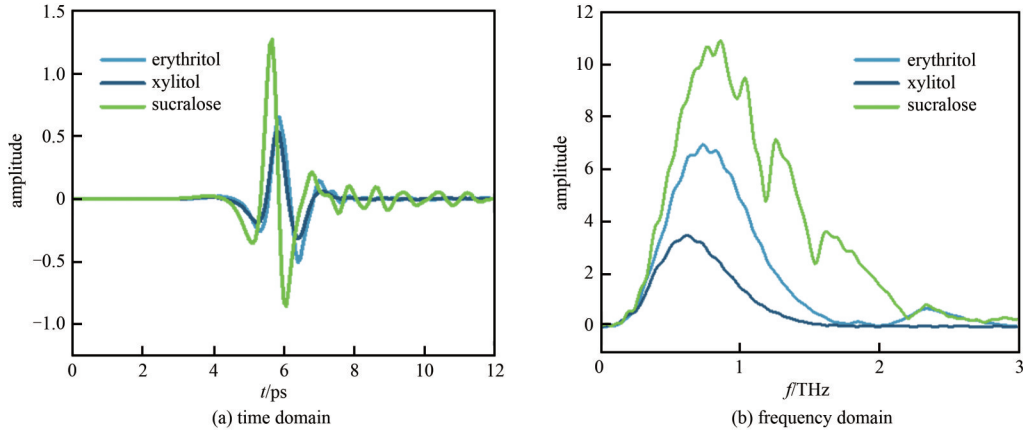


Fig.1 THz spectra of sucralose, erythritol and xylitol
图 1 三氯蔗糖、赤藓糖醇、木糖醇压片的太赫兹光谱图

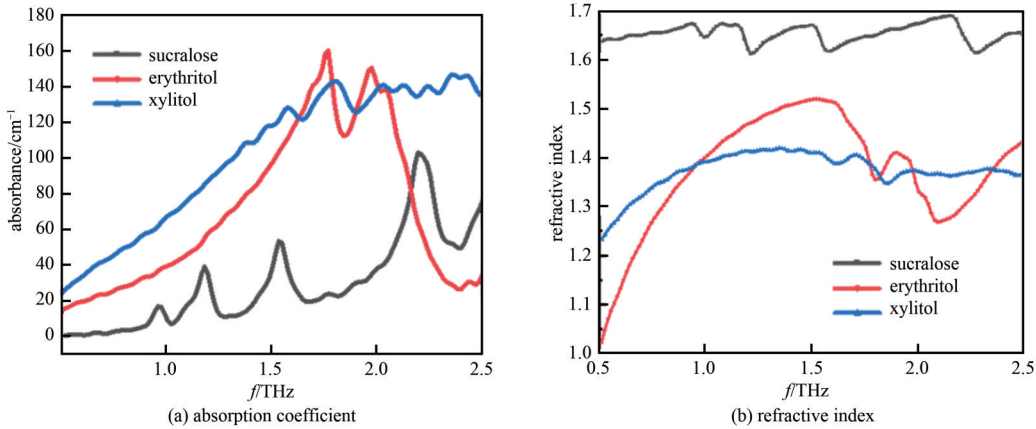


Fig.2 Optical parameters of three sweeteners in the range of 0.50~2.50 THz
图 2 0.50~2.50 THz 范围内 3 种甜味剂纯品的光学参数

2.2 3 种甜味剂与面粉混合物的 THz 光谱

不同比例混合的 3 种甜味剂压片在 0.50~2.50 THz 的吸收谱如图 3 所示, 可以看到, 面粉无吸收峰, 其吸收强度随着频率增大而增强; 三氯蔗糖在高浓度(>30%)下可观察到 3 个吸收峰, 分别位于 1.17 THz、1.54 THz 及 2.21 THz 处; 另外 2 种甜味剂与面粉的混合样品则没有明显吸收峰, 其吸收强度随频率增加而增强。由于面粉颗粒散射大的原因, 低浓度的样品在高频(>1.50 THz)的噪声较大。

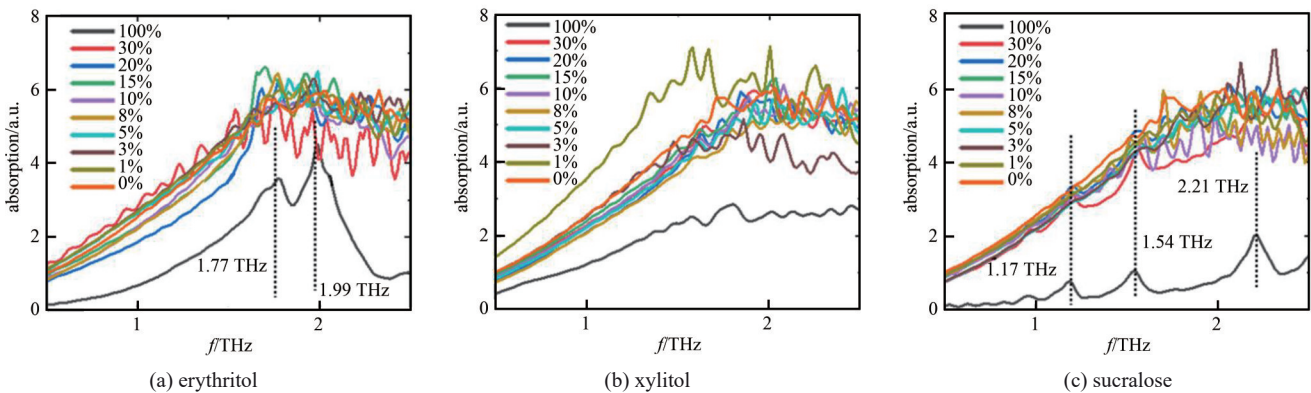


Fig.3 Absorption spectra of three sweeteners mixed with flour in different proportions
图 3 3 种甜味剂与面粉以不同比例混合的吸收谱

3 结合机器学习算法的定性识别和定量分析

3.1 混合物的定性分析

3 种甜味剂与面粉混合的数据分别有 49 组，共计 147 组数据。由于 0~0.50 THz 以及 1.50~2.50 THz 范围的光谱信号噪声较大，仅保留 0.50~1.50 THz 的光谱数据。先采用 S-G 滤波(Savitzky-Golay filter)对光谱数据进行平滑处理，再将 147 组光谱数据分为训练集(102 组)和测试集(45 组)，导入 Matlab 中分别采用 SVM、GA-SVM、GS-SVM 和 SSA-SVM 算法进行分类，分类结果如表 1 所示。SVM 模型在未采用优化算法的情况下，训练集和测试集准确度分别为 81.37% 和 73.33%；采用优化算法后，训练集和测试集的准确度都有较大提升。其中 GS-SVM 和 SSA-SVM 模型的分类效果更好，二者的训练集准确度均为 95.10%，提高了 13.73%；测试集分类准确度均为 95.56%，提高了 22.23%。

表 1 3 种甜味剂与面粉混合物的分类结果

Table 1 Results of classification of three sweetener-flour mixtures

model	train set correct quantity/	accuracy of the	test set correct quantity/	accuracy of the	optimal penalty	optimal kernel	training time/s
	totality	train set/%	totality	test set/%	factor <i>c</i>	parameter <i>g</i>	
SVM	83/102	81.37	33/45	73.33	-	-	5.02
GA-SVM	95/102	93.14	40/45	88.89	74.64	0.34	17.82
GS-SVM	97/102	95.10	43/45	95.56	128	0.25	33.86
SSA-SVM	97/102	95.10	43/45	95.56	2.98	15.90	13.29

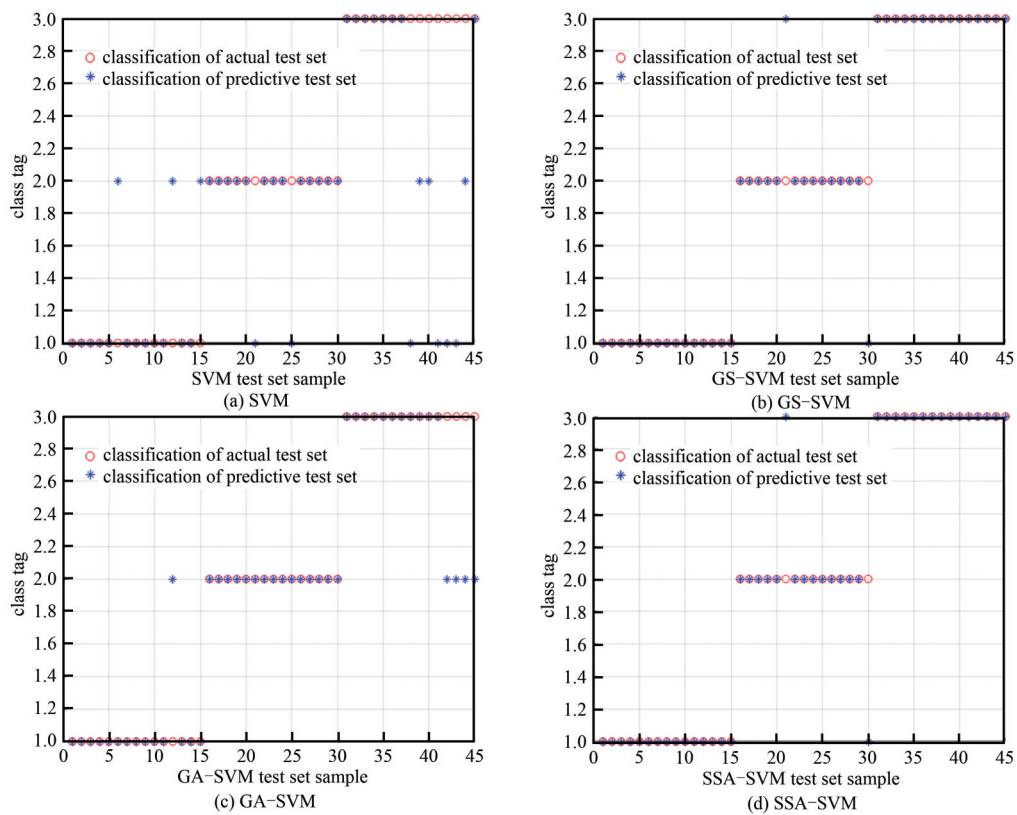


Fig.4 Actual classification and predictive classification of the test set

图 4 测试集的实际分类和预测分类图

4 种算法模型对测试集的实际分类和预测分类比较结果如图 4 所示，其中类别标签 1、2、3 分别代表三氯蔗糖、赤藓糖醇和木糖醇样本。可以看出，只用 SVM 模型进行分类时，三氯蔗糖样本被错误地分类到赤藓糖醇样本，赤藓糖醇样本也被错误分类到三氯蔗糖样本，而木糖醇样本被错误地分类到其他 2 种样本。比较 GS-SVM、GA-SVM 和 SSA-SVM 模型可以看出，错误分类样本明显减少。当采用 GA-SVM 模型时，三氯蔗糖有 1 个样本被错误分类成赤藓糖醇，赤藓糖醇样本没有被错误分类的样本，木糖醇有 4 个样本被错误地分类成赤藓糖醇；当采用 GS-SVM 模型时，三氯蔗糖和木糖醇没有被错误分类的样本，赤藓糖醇有 2 个样本分别被错误地分类成木糖醇和三氯蔗糖；当采用 SSA-SVM 模型时，三氯蔗糖和木糖醇没有被错误分类的样本，赤藓糖醇有 2 个样本分别

被错误地分类成木糖醇和三氯蔗糖。比较 3 种模型可以看出, GS-SVM 和 SSA-SVM 分类模型的准确率较高, 但由于 GS-SVM 模型在寻优时参数组合非常多, 导致训练时间较长; SSA-SVM 模型的训练时间较短, 效率更优。说明采用优化算法结合 SVM 可以对 3 种甜味剂与面粉混合物的样本很好地分类, 其中 SSA-SVM 模型可作为识别 3 种甜味剂和面粉混合物的最优模型。

3.2 混合物的定量分析

采用 GS-SVR、GA-SVR 和 SSA-SVR 三种算法模型对 3 种甜味剂与面粉混合物进行定量回归分析。每种甜味剂与面粉混合物的数据均有 49 组, 其中训练集数据 29 组, 测试集数据 20 组, 定量分析结果如表 2 所示。可以看到, 对赤藓糖醇采用 SSA-SVR 算法时, 测试集 RMSE 远小于 GS-SVR 和 GA-SVR 模型, 低至 0.009%; 拟合程度 R^2 最高, 为 0.999 8。说明 SSA-SVR 预测偏差远小于另外 2 种模型, 因此定量分析的最佳模型为 SSA-SVR。对木糖醇和三氯蔗糖与面粉混合物的最佳定量预测模型也是 SSA-SVR, 其测试集 RMSE 分别为 0.09%、0.28%, R^2 分别为 0.999 8、0.991 7, 两项指标均优于 GS-SVR 模型和 GA-SVR 模型。说明运用 SSA-SVR 模型可以很好地对木糖醇和三氯蔗糖在面粉中的浓度进行定量回归分析, 且拥有更精准的预测能力和更好的拟合关系。

表 2 木糖醇、赤藓糖醇和三氯蔗糖与面粉混合物的定量分析结果
Table 2 Results of quantitative analysis of xylitol, erythritol and sucralose mixed with flour

sweetener	model	train set RMSE/%	train set R^2	test set RMSE/%	test set R^2	optimal penalty factor c	optimal kernel parameter g
xylitol	GS-SVR	2.95	0.932 1	2.53	0.922 9	55.71	0.02
	GA-SVR	0.09	0.999 8	0.68	0.984 3	6.12	1.17
	SSA-SVR	0.09	0.999 8	0.09	0.999 8	40.25	0.23
erythritol	GS-SVR	3.080	0.922 6	5.440	0.875 8	1.74	1.15
	GA-SVR	2.580	0.924 6	5.280	0.861 5	1.88	1.26
	SSA-SVR	0.009	0.999 8	0.009	0.999 8	100.00	1.56
sucralose	GS-SVR	1.69	0.950 4	1.63	0.951 1	256.00	0.03
	GA-SVR	2.12	0.942 4	1.41	0.960 5	82.63	0.08
	SSA-SVR	0.68	0.991 2	0.28	0.991 7	100.00	0.23

3 种模型用于三氯蔗糖与面粉混合物定量分析的测试集数据真实值与预测值拟合关系和适应度曲线如图 5 所示。其中, GS 算法寻优后得到的最佳惩罚因子 c 为 256, 最佳核函数参数 g 为 0.03。采用 GA 算法寻优时, 将最大迭代次数设为 100, 种群数设为 30, 可以看到 GA-SVR 算法一共迭代了 100 次, 从第 40 代开始, 其平均适应度趋于最佳适应度, 比较稳定; 采用 SSA 算法寻优时, 将最大迭代次数设为 100, SSA-SVR 算法一共迭代了 100 次, 前期收敛速度较慢, 从第 80 代开始趋于稳定。从以上分析可以看出, GS-SVR、GA-SVR 以及 SSA-SVR 模型都可以对混合物进行一个相对较好的定量分析, 其中 SSA-SVR 的分析效果最佳。

基于 3 种甜味剂与面粉混合物的太赫兹光谱数据, 分别构建了 GA-SVM、GS-SVM、SSA-SVM 模型对混合物进行了分类研究, GA-SVR、GS-SVR、SSA-SVR 模型对混合物进行了定性分析。对于分类模型, SSA-SVM 模型效果最佳, 其训练集和测试集准确度相较于未结合优化参数选取算法的 SVM 模型有明显的提升。在 3 种甜味剂与面粉的定量分析研究中, 采用 SSA-SVR 算法最佳, 其 R^2 均在 0.99 以上, 说明它可以很好地对 3 种甜味剂与面粉的混合物浓度进行定量回归分析。

4 结论

本文以适用于糖尿病人食用的 3 种甜味剂为研究对象, 结合太赫兹时域光谱技术和机器学习算法对样品进行了定性识别和定量分析。首先通过从太赫兹光谱中提取样品的吸收系数和折射率, 实现了 3 种甜味剂的定性识别; 然后, 结合机器学习算法进一步对混合物进行了分类和定量分析, 其中 SSA-SVM 和 SSA-SVR 模型分类预测最高准确率达 95.56%, 定量回归预测的最佳回归系数 R^2 为 0.999 8, 实现了良好的分类和定量分析效果。本文提出的 3 种甜味剂的快速检测结合了太赫兹时域光谱技术和优化算法、机器学习算法, 算法成熟, 技术先进, 确保了本研究能有效地参与市场质量监管工作。实际应用中, 需根据测试样品的状态、测试系统、测试环境等测试条件, 选择合适的优化算法和机器学习算法, 进一步提高市场应用中的鉴别准确率。本研究对人们的日常饮食健康及环境的污染防控等都有着极大的价值与意义。

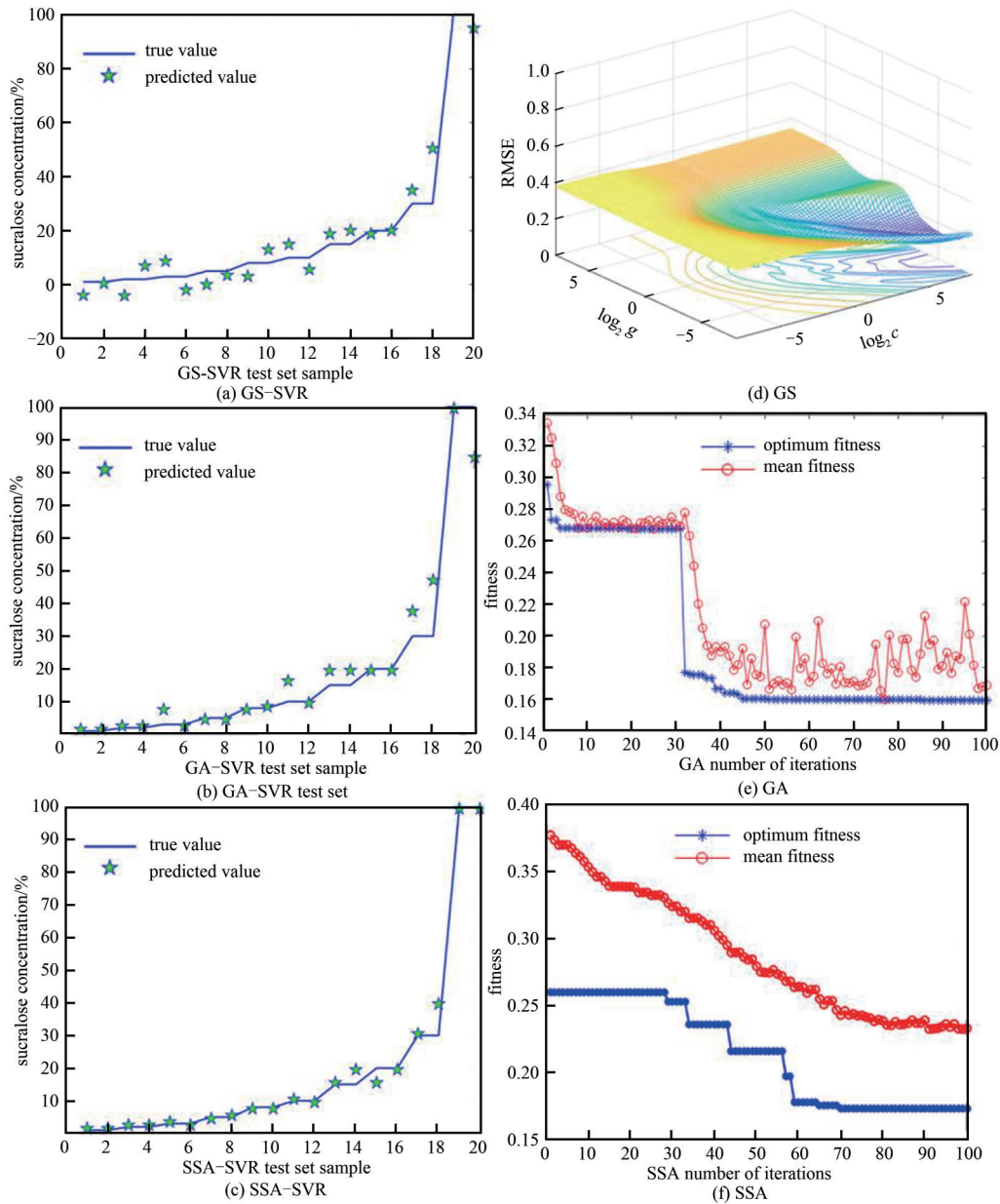


Fig.5 Test set results of sucralose-flour mixtures for three models

图 5 三氯蔗糖与面粉混合物的 3 种模型的测试集结果

参考文献：

[1] BUERGE I J,KELLER M,BUSER H R,et al. Saccharin and other artificial sweeteners in soils: estimated inputs from agriculture and households,degradation,and leaching to groundwater[J]. Environmental Science and Technology, 2011,45(2):615-621. doi: 10.1021/es1031272.

[2] BIEDERMANN M,MUNOZ C,GROB K. Update of on-line coupled liquid chromatography-gas chromatography for the analysis of mineral oil hydrocarbons in foods and cosmetics[J]. Journal of Chromatography A, 2017(1521):140-149.

[3] MOHAMMADZADEH-AGHDASH H, EZZATI-NAZHAD-DOLATABADI J, DEHGHAN P, et al. Multi-spectroscopic and molecular modeling studies of bovine serum albumin interaction with sodium acetate food additive[J]. Food Chemistry, 2017 (228):265-269. doi:10.1016/j.foodchem.2017.01.149.

[4] GOWEN A A,O'SULLIVAN C,O'DONNELL C P. Terahertz time domain spectroscopy and imaging:emerging techniques for food process monitoring and quality control[J]. Trends in Food Science & Technology, 2012,25(1):40-46. doi:10.1016/j.tifs.2011. 12.006.

- [5] 李明亮. 基于太赫兹时域光谱技术的生物分子鉴别及相互作用研究[D]. 长春:吉林大学, 2018. (LI Mingliang. Research on biomolecules identification and their interactions based on terahertz time-domain spectroscopy technology[D]. Changchun, Jilin, China: Jilin University, 2018.)
- [6] LU Shaohua, ZHANG Xin, ZHANG Zhuoyong, et al. Quantitative measurements of binary amino acids mixtures in yellow foxtail millet by terahertz time domain spectroscopy[J]. Food Chemistry, 2016(211):494–501. doi:10.1016/j.foodchem.2016.05.079.
- [7] LIU H B, ZHANG X C. Dehydration kinetics of D-glucose monohydrate studied using THz time-domain spectroscopy[J]. Chemical Physics Letters, 2006,429(1/3):229–233. doi:10.1016/j.cplett.2006.07.100.
- [8] 李斌, 龙园, 刘海顺, 等. 基于太赫兹光谱技术的 D-无水葡萄糖定性定量分析研究[J]. 光谱学与光谱分析, 2017,37(7):2165–2170. (LI Bin, LONG Yuan, LIU Haishun, et al. The determination of glucose based on terahertz spectroscopy[J]. Spectroscopy and Spectral Analysis, 2017,37(7):2165–2170.) doi:10.3964/j.issn.1000-0593(2017)07-2165-06.
- [9] MASSAOUTI M, DASKALAKI C, GORODETSKY A, et al. Detection of harmful residues in honey using terahertz time-domain spectroscopy[J]. Applied Spectroscopy, 2013,67(11):1264–1269. doi:10.1366/13-07111.
- [10] FENG Lu, CAI Jinhui, ZHANG Guangxin, et al. Research of benzoyl peroxide residues in flour by terahertz spectroscopy technology[J]. Advanced Materials Research, 2013(712/715):2039–2042.
- [11] LI Mingliang, DAI Guangbin, CHANG Tianying, et al. Accurate determination of geographical origin of tea based on terahertz spectroscopy[J]. Applied Sciences, 2017,7(2):172.
- [12] 李鹏鹏. 基于小波分析的太赫兹光谱信号分析研究[D]. 郑州:河南工业大学, 2018. (LI Pengpeng. Research on terahertz spectrum signal analysis based on wavelet analysis[D]. Zhengzhou, Henan, China: Henan University of Technology, 2018.)
- [13] LONG Yuan, LI Bin, LIU Huan. Analysis of fluoroquinolones antibiotic residue in feed matrices using terahertz spectroscopy[J]. Applied Optics, 2018,57(3):544–550. doi:10.1364/AO.57.000544.
- [14] SUN Xudong, ZHU Ke, LIU Junbin, et al. Terahertz spectroscopy determination of benzoic acid additive in wheat flour by machine learning[J]. Journal of Infrared, Millimeter and Terahertz Waves, 2019,40(4):466–475. doi:10.1007/s10762-019-00579-z.
- [15] LIAN F Y, FU M X, XU D G. Qualitative and quantitative analyses of potassium sorbate in milk powder using terahertz spectra[J]. Journal of Applied Spectroscopy, 2020,87(4):764–772. doi:10.1007/s10812-020-01067-2.
- [16] HU Jun, LIU Yande, SUN Xudong, et al. Quantitative determination of benzoic acid in flour based on terahertz time-domain spectroscopy and BPNN model[J]. Laser & Optoelectronics Progress, 2020,57(7):073002. doi:10.3788/LOP57.073002.
- [17] HU Jun, XU Zhen, LI Maopeng, et al. Determination of melamine content in milk powder based on neural network algorithm and terahertz spectrum detection[J]. Laser & Optoelectronics Progress, 2020,57(22):223001. doi:10.3788/LOP57.223001.
- [18] WEI Xiao, LI Song, ZHU Shiping, et al. Terahertz spectroscopy combined with data dimensionality reduction algorithms for quantitative analysis of protein content in soybeans[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2021(253):119571. doi:10.1016/j.saa.2021.119571.
- [19] CHEN Tao, MA Lingjie, TANG Zongqing, et al. Identification of coumarin-based food additives using terahertz spectroscopy combined with manifold learning and improved support vector machine[J]. Journal of Food Science, 2022,87(3):1108–1118. doi:10.1111/1750-3841.16064.
- [20] PAN Shubao, ZHANG Huo, LI Zhi, et al. Quantitative determination of sucrose adulterated in red ginseng by Terahertz Time-Domain Spectroscopy (THz-TDS) with Monte Carlo Uninformative Variable Elimination (MCUVE) and Support Vector Regression (SVR)[J]. Journal of Spectroscopy, 2022(1):1–10. doi:10.1155/2022/5847819.
- [21] DUVILLARET L, GARET F, COUTAZ J L. A reliable method for extraction of material parameters in terahertz time-domain spectroscopy[J]. IEEE Journal of Selected Topics in Quantum Electronics, 1996,2(3):739–746. doi:10.1109/2944.571775.
- [22] DORNEY T D, BARANIUK R G, MITTLEMAN D M. Material parameter estimation with terahertz time-domain spectroscopy[J]. Journal of the Optical Society of America A Optics Image Science and Vision, 2001,18(7):1562–1571. doi:10.1364/josaa.18.001562.
- [23] PARROTT E P, ZEITLER J A, GLADDEN L F. Accurate determination of optical coefficients from chemical samples using terahertz time-domain spectroscopy and effective medium theory[J]. Optics Letters, 2009,34(23):3722–3724. doi:10.1364/OL.34.003722.
- [24] ZHAO Rong, ZOU Bin, ZHANG Guling, et al. High-sensitivity identification of aflatoxin B1 and B2 using terahertz time-domain spectroscopy and metamaterial-based terahertz biosensor[J]. Journal of Physics D: Applied Physics, 2020,53(19):195401.
- [25] CORTES C, VAPNIK V. Support-vector networks[J]. Machine Learning, 1995,20(3):273–297. doi:10.1007/BF00994018.

作者简介：

钟芸襄(1998—)，女，在读硕士研究生，主要研究方向为太赫兹超材料传感器的设计和优化。email: 21302271@muc.edu.cn.

张然(1997—)，女，在读硕士研究生，主要研究方向为太赫兹光谱数据分析及超材料结构设计。

熊子仪(2000—)，女，在读硕士研究生，主要研究方向为太赫兹光谱分析与传感性能优化。

邹斌(1980—)，男，博士，副教授，硕士生导师，主要研究方向为表面增强拉曼散射机理与实验研究。

杨玉平(1976—)，女，博士，教授，硕士生导师，主要研究方向为太赫兹光谱表征。

(上接第377页)

- [6] YANG Z Y, ISHII S, YOKOYAMA T, et al. Narrowband wavelength selective thermal emitters by confined Tamm plasmon polaritons[J]. ACS Photonics, 2017,4(9):2212–2219. doi:10.1021/acsp Photonics.7b00408.
- [7] LEE K J, WU J W, KIM K. Enhanced nonlinear optical effects due to the excitation of optical Tamm plasmon polaritons in one-dimensional photonic crystal structures[J]. Optics Express, 2013,21(23):28817–28823. doi:10.1364/OE.21.028817.
- [8] XU Wenhui, CHOU Y H, YANG Z Y, et al. Tamm Plasmon–Polariton ultraviolet lasers[J]. Advanced Photonics Research, 2022, 3(1):2100120. doi:10.1002/adpr.202100120.
- [9] HU Jigang, YAO Enxu, XIE Weiqiang, et al. Strong longitudinal coupling of Tamm plasmon polaritons in graphene/DBR/Ag hybrid structure[J]. Optics Express, 2019,27(13):18642–18652. doi:10.1364/OE.27.018642.
- [10] ZHANG Kun, LIU Yan, XIA Feng, et al. Tuning of the polariton modes induced by longitudinal strong coupling in the graphene hybridized DBR cavity[J]. Optics Letters, 2020,45(13):3669–3672. doi:10.1364/OL.397342.
- [11] JENA S, TOKAS R B, THAKUR S, et al. Rabi-like splitting and refractive index sensing with hybrid Tamm plasmon–cavity modes[J]. Journal of Physics D: Applied Physics, 2022,55(17):175104. doi:10.1088/1361-6463/ac49b3.
- [12] 张葆青,冯明明,张翼飞,等. 石墨烯动态调控太赫兹表面等离子激元[J]. 太赫兹科学与电子信息学报, 2023,21(8):965–970, 976. (ZHANG Baoqing, FENG Mingming, ZHANG Yifei, et al. Active modulation of terahertz surface plasmons polaritons with graphene[J]. Journal of Terahertz Science and Electronic Information Technology, 2023, 21(8): 965–970, 976.) doi:10.11805/TKYDA2022163.
- [13] WANG Xi, JIANG Xing, YOU Qi, et al. Tunable and multichannel terahertz perfect absorber due to Tamm surface plasmons with graphene[J]. Photonics Research, 2017,5(6):536–542. doi:10.1364/PRJ.5.000536.
- [14] QING Yeming, REN Yongze, LEI Danguan, et al. Strong coupling in two-dimensional materials-based nanostructures: a review[J]. Journal of Optics, 2022,24(2):024009. doi:10.1088/2040-8986/ac47b3.
- [15] QIN Jian, CHEN Yuhui, ZHANG Zhepeng, et al. Revealing strong plasmon–exciton coupling between nanogap resonators and two-dimensional semiconductors at ambient conditions[J]. Physical Review Letters, 2020, 124(6): 063902. doi: 10.1103/PhysRevLett.124.063902.

作者简介：

谢金霖(1998–)，男，在读硕士研究生，主要研究方向为二维材料中近场极化激元研究与应用。email: 1648270540@qq.com.

张天宇(1996–)，博士，主要研究方向为新型材料与太赫兹波相互作用研究与应用。

胡旻(1981–)，男，博士，教授，博士生导师，主要研究方向为新型太赫兹辐射源和太赫兹科学技术应用等。