

基于电子舌和电子鼻结合 CNN-Transformer 模型的绿茶种类识别

Green tea species recognition based on electronic tongue and electronic nose combined with CNN-Transformer model

刘川正 马景余 白雪瑞 曾琬晴 王志强

LIU Chuangzheng MA Jingyu BAI Xuerui ZENG Wanqing WANG Zhiqiang

(山东理工大学计算机科学与技术学院, 山东 淄博 255049)

(School of Computer Science and Technology, Shandong University of Science and Technology, Zibo, Shandong 255049, China)

摘要:[目的]实现绿茶种类的快速识别。[方法]提出一种基于电子舌和电子鼻结合 CNN-Transformer 组合模型实现绿茶种类辨识的快速检测方法。分别使用电子舌、电子鼻对 5 种不同种类绿茶采集味觉、嗅觉的指纹信息,利用短时傅里叶变换(short-time fourier transform, STFT)将一维电子舌和电子鼻信号转换为二维时频图,充分揭示信号能量在时频域的分布特性;提出一种 CNN-Transformer 组合模型实现电子舌和电子鼻的信息融合和模式识别。该模型引入选择性核卷积和归一化注意力设计卷积模块来替换传统的 CNN 卷积层,以实现信号时频图的局部特征动态提取;采用 Transformer 编码器中的多头自注意力机制提取电子舌和电子鼻特征的全局时序信息,并实现其特征的加权融合;最后,通过全连接层进行分类识别。[结果]基于电子舌和电子鼻的信息融合方法能够有效提取绿茶样本的味觉和嗅觉信号深层特征,并为模型提供更丰富的融合特征表征,以实现对不同绿茶种类的高准确识别,其测试集准确率、精确率、召回率和 F_1 -Score 分别达到 99.00%, 99.05%, 99.00%, 99.00%。[结论]试验方法具有成本低、快速、高效等特点。

关键词:绿茶;种类识别;电子鼻;电子舌;Transformer;信息融合

Abstract: [Objective] To realize rapid detection of green tea species identification. [Methods] A rapid detection method based

基金项目:山东省自然科学基金项目(编号:ZR2022MF330);教育部科技发展中心产学研创新基金项目(编号:2018A02010)

作者简介:刘川正,男,山东理工大学在读硕士研究生。

通信作者:王志强(1977—),男,山东理工大学教授,博士。

E-mail: wzq@sdu.edu.cn

收稿日期:2023-11-01 **改回日期:**2024-01-04

on the combination of electronic tongue and electronic nose combined with CNN-Transformer composite model was proposed. The electronic tongue and electronic nose were used to collect the fingerprint information of taste and smell for five different kinds of green tea. The one-dimensional electronic tongue and electronic nose signals were transformed into two-dimensional time-frequency maps using the short-time Fourier transform (STFT), which fully revealed the distribution characteristics of the signal energy in the time-frequency domain. A CNN-Transformer combination model was proposed to realize the fusion of the electronic tongue and the electronic nose information and pattern recognition. The model adopted selective kernel convolution and normalized attention in designing convolution module to replace the convolution layer of the traditional CNN to achieve the dynamic extraction of local features from the time-frequency map of the signal. The multi-head self-attention mechanism in the Transformer encoder was used to extract the global temporal information in the features of the electronic tongue and the electronic nose and achieve the weighted fusion of their features. Finally, classification recognition was carried out by the fully connected layer. [Results] The information fusion method based on electronic tongue and electronic nose could effectively extract the deep features of the taste and smell signals from green tea samples and provide richer fused feature representations for the model to achieve highly accurate recognition of different species, with a test set accuracy, precision, recall and F_1 -Score of 99.00%, 99.05%, 99.00%, and 99.00%, respectively. [Conclusion] This study provides a low-cost, fast and efficient detection method for green tea species recognition.

Keywords: green tea; species identification; electronic nose;

electronic tongue; transformer; information fusion

绿茶深受消费者的喜爱,但因产地、工艺等影响,其品质价格差异较大,导致部分商家冒用种类,以次充好,严重损害了茶叶市场的健康秩序和消费者利益^[1]。绿茶种类识别可保证茶叶品质与商业价值,促进品牌培育与保护。传统的绿茶种类识别主要基于感官评定法和仪器分析法等。感官评定法主要是基于评审人员通过视觉、嗅觉、味觉等感官进行评定,存在主观性强、难以量化、易受环境干扰、稳定性差等问题^[2-3]。仪器分析法主要有高效液相色谱法^[4]、气相色谱法^[5]等,但此类方法存在仪器成本高、体积大、分析流程长,需要专业操作人员等问题^[6],难以应用于现场快速检测。

电子舌、电子鼻是一类新型仿生学仪器,通过利用电化学传感阵列结合模式识别技术来分析液态、气态样本的味觉和嗅觉指纹图谱,具有检测速度快、成本低、操作简单、客观性强等^[7-8]优点,已被广泛应用于茶叶识别领域。Kumar 等^[9]采用电子舌结合线性模型对红茶中的总茶黄素(TF)含量进行了预测;Chen 等^[10]采用电子鼻结合多变量统计分析对滇红茶的香气质量进行了评价。但绿茶中含有丰富的化学成分,其品质特征在味觉和嗅觉上都有体现,单一采用电子舌或电子鼻均无法获取完整的样本特征信息,影响分类效果。Lu 等^[11]采用电子舌和电子鼻结合局部线性嵌入—极限学习机(LLE-SVM)模型对传统和杂交大米进行了区分。Banerjee 等^[12]采用电子舌和电子鼻结合 K-最近邻(KNN)分类算法对不同等级的印度红茶进行了分类。上述研究均采用传统的机器学习算法,高度依赖于特征的手动提取且泛化能力有限^[13]。

深度学习能够基于大数据自动提取特征并分类,从而实现端到端的模式识别过程。卷积神经网络(convolutional neural network, CNN)作为一种常用深度学习模型在处理时频域图像时表现出优异的性能^[14]。但传统 CNN 的固定卷积核难以适应具有不同方向、尺度特征的数据。同时,由于 CNN 的卷积操作主要关注局部特征,在处理长距离依赖性或者全局信息时存在一定局限^[15]。Transformer 是近年来提出的一种新型深度学习模型,其通过多头自注意力机制计算不同信息的注意力权重,从而提高模型处理多模态数据时的信息互补能力^[16]。同时,多头自注意力机制能够帮助 Transformer 编码器有效地捕获全局信息,在处理序列数据时展现出显著的优势^[17],但是 Transformer 由于缺乏卷积结构,可能导致忽视数据中的局部结构信息^[18]。近期有学者^[19-20]提出将 CNN 和 Transformer 模型相结合,利用卷积和多头自注意力机制的互补优势,实现对时频域信息的深度特征提取。但目前将其应用于电子舌、电子鼻的模式识别中的研究尚未见报道。

研究拟提出一种基于电子舌和电子鼻结合 CNN-Transformer 组合模型的绿茶种类识别方法。分别使用电子舌、电子鼻捕获绿茶样本的指纹信息,利用短时傅里叶变换(short-time fourier transform, STFT)将电子舌、电子鼻信号转换为二维时频图;提出一种 CNN-Transformer 组合模型实现电子舌、电子鼻信号的特征提取、信息融合与分类识别,并引入选择性核卷积和归一化注意力设计卷积模块,利用 Transformer 的多头自注意力机制关注信号时频图的全局信息,并基于多头自注意力机制实现特征信息的加权融合,从而为实现对绿茶种类的快速识别检测提供依据。

1 材料与仪器

1.1 试验材料

分别选取西湖龙井、黄山毛峰、信阳毛尖、崂山绿茶、日照绿茶 5 种不同种类的绿茶样本。每种绿茶准确称取 5 g,用 200 mL 沸腾蒸馏水冲泡 5 min,过滤,冷却至室温。采用电子舌采集信号时,每种绿茶采集 200 个样本,共采集 1 000 个。采用电子鼻采集信号时,使用微型气泵抽取新鲜空气对传感器进行清洗,使传感器阵列达到稳定状态再对密闭容器中的绿茶溶液进行信号采集,每种绿茶采集 200 个样本,共采集 1 000 个。从采集到的电子舌、电子鼻信号中随机抽取 900 个样本数据按 8:2 的比例划分为训练集与验证集,剩余 100 个样本数据作为测试集。

1.2 电子舌与电子鼻系统

电子舌和电子鼻系统由实验室自行研发。电子舌系统(图 1)主要由 4 个部分组成:传感器阵列、信号调理电路、数据采集设备以及 LabVIEW 上位机软件。传感器阵列包括 1 个铂辅助电极、8 个不同的工作电极(玻碳、银、金、钛、铂、钨、镍、钨)以及 1 个银/氯化银参比电极。检测流程:上位机软件控制数据采集卡产生激励信号,该信号经过信号调理后,被加载到传感器阵列上。在激励信号的作用下,样本溶液在不同工作电极表面发生氧化还原反应,产生微弱的电流信号。电流信号经放大和滤波处理后,被送到数据采集卡中进行数据采集。最后,由上位机软件进行模式识别处理。

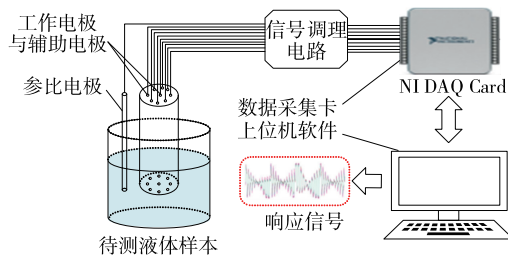


图 1 电子舌系统结构图

Figure 1 Structure of the electronic tongue system

电子鼻系统(图 2)主要由传感器阵列、检测腔及气路、信号调理电路、数据采集装置以及 LabVIEW 上位机软件组成。传感器阵列由 6 种金属氧化物半导体气敏传感器组成,分别为 TGS2600、TGS2602、TGS2603、TGS2610、TGS2611、TGS2620。电子鼻检测流程:首先,利用采样泵将待测气体样本引入检测室,确保气体持续流动并与传感器充分接触。当气体中的活性成分与传感器表面接触时,发生化学反应,传感器的电导率变化并产生相应的电信号,该信号经放大、滤波后,被数据采集卡捕获,最后送至上位机软件进行处理。

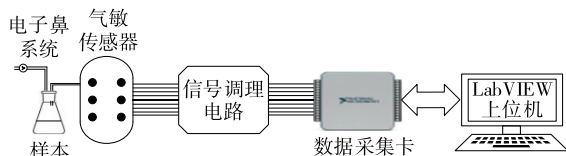


图 2 电子鼻系统结构图

Figure 2 Structure of the electronic nose system

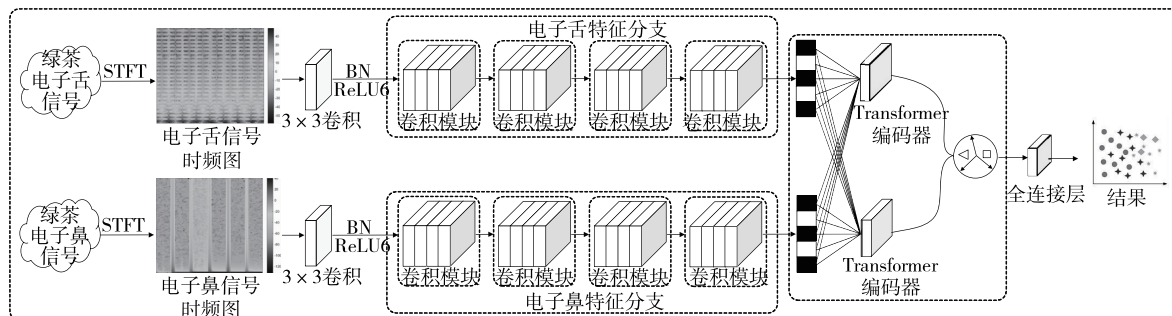


图 3 CNN-Transformer 模型结构

Figure 3 CNN-Transformer model structure diagram

2.2 短时傅里叶变换(STFT)

短时傅里叶变换可将信号通过窗函数分割成多个短时的片段,并对每个片段都进行傅里叶变换以获得各时刻的频域特性。STFT 通过滑动窗函数重复此变换过程,得到一系列的频谱函数,将这些结果依次拼接可以得到完整的二维时变频谱图,其计算过程为:

$$\text{STFT}\{x(t)\}(\tau, \omega) = \int_{-\infty}^{+\infty} x(t) \omega(t - \tau) e^{-i\omega t} dt, \quad (1)$$

式中:

$x(t)$ —— 原始信号;

$\omega(t - \tau)$ —— 窗函数;

τ —— 窗函数的中心位置;

ω —— 角频率, s^{-1} ;

$e^{-i\omega t}$ —— 复指数函数。

STFT 中不同的窗函数大小和步长设定会导致时频图展现不同方向、尺度的特征^[21]。方向特征反映信号的频率变化趋势,而尺度特征反映信号的频率和时间分辨率。通过不同尺度和方向的特征进行交互,可以揭示信

2 模式识别与信息融合方法

2.1 CNN-Transformer 组合模型结构

CNN-Transformer 模型结构如图 3 所示。采集不同种类绿茶样本的电子舌、电子鼻信号,并采用短时傅里叶变换将其分别转换为二维时频图。时频图分别通过初始卷积层(3×3)进行维度调整,确保网络中后续层的输入维度与预期相匹配。然后执行批量归一化(BN)和 ReLU6 激活函数,以增强模型的鲁棒性和非线性表达能力。针对信号时频图存在不同方向、尺度特征的问题,引入选择性核卷积和归一化注意力设计 4 个卷积模块(CNNBlock),实现对信号时频图中的局部特征的动态提取。将特征向量送入 Transformer 编码器,利用多头自注意力机制捕获特征中的全局时序信息,并利用编码器不同头生成的注意力权重向量之间的相似度,通过特征加权方式实现特征融合。最后,通过全连接层(FC)对特征进行分类,实现绿茶种类识别。

号复杂的时频结构信息。

2.3 基于改进 CNN 的局部特征提取方法

针对传统卷积层在处理信号时频图时固定卷积核的限制,提出一种基于改进 CNN 的局部特征提取方法。首先,通过选择性核卷积(selective kernel convolution, SKConv)动态调整卷积核的尺度和方向,以适应不同的特征表征,有效捕捉信号时频图中的多尺度和多方向特征。同时,利用归一化注意力模块(normalization-based attention module, NAM)集中关注重要的局部特征,优化特征提取过程。CNNBlock 局部特征提取模块如图 4 所示。在 CNNBlock 中,输入特征首先通过一个 1×1 的卷积层调整特征通道。随后,通过选择性核卷积调整对特征的关注度,经批量归一化和 ReLU6 激活函数处理后,再次通过一个 1×1 的卷积层调整特征通道。最后,特征进入归一化注意力模块进行局部特征强化并降低噪声影响。

选择性核卷积使用两种不同大小的卷积核动态调整对特征的关注度,其结构如图 5 所示。其两个卷积核分

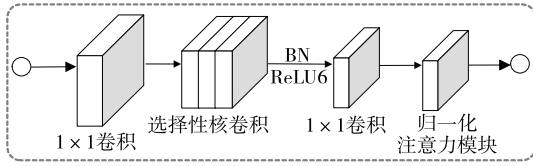


图 4 CNNBlock 结构图

Figure 4 CNNBlock structure diagram

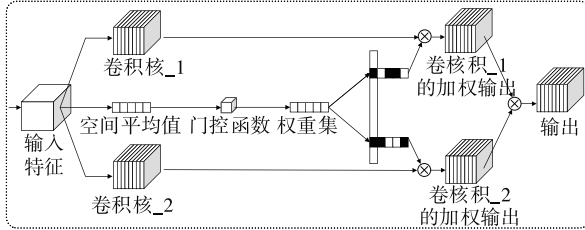


图 5 SKConv 结构图

Figure 5 Selective kernel convolution structure diagram

别为 Conv_1 和 Conv_2, 对输入特征 x , 其输出分别记为 Conv1_out 和 Conv2_out。同时, 计算输入特征 x 的空间平均值为 x_{mean} 。通过门控函数 g 计算门控权重集 gate_weights。最终, 得到 SKConv 的输出为权重集与对应卷积输出的线性组合, 其计算过程为:

$$\begin{cases} \text{gate_weights} = \text{softmax}[g(x_{\text{mean}})], \\ \text{SKConv_out} = \alpha \times \text{Conv1_out} + \beta \times \text{Conv2_out} \end{cases}, \quad (2)$$

式中:

α, β ——来自权重集 gate_weights 的权重 (对应于两种不同的卷积核输出)。

归一化注意力模块能够抑制无关特征对模型的影响, 其结构如图 6 所示, NAM 使用批归一化 (batch normalization, BN) 的缩放因子 γ_i 和 λ_i 量化每个通道和

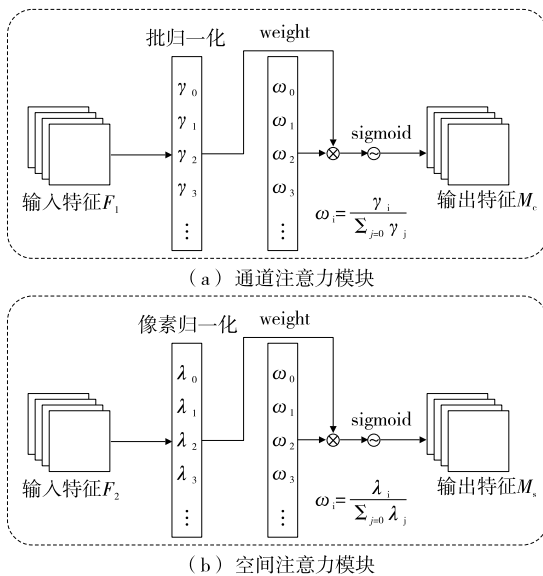


图 6 归一化注意力模块图

Figure 6 Diagram of the normalized attention module

空间的重要性, 其计算过程为:

$$\begin{cases} B_{\text{out}} = \text{BN}(B_{\text{in}}) = \gamma \frac{B_{\text{in}} - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} + \beta, \\ M_c = \text{sigmoid}\{\omega_\gamma [\text{BN}(F_1)]\}, \\ M_s = \text{sigmoid}\{\omega_\lambda [\text{BN}(F_2)]\} \end{cases}, \quad (3)$$

式中:

μ_B, σ_B ——小批量 Batch 的均值和方差;

γ, β ——可训练的参数;

M_c, M_s ——通道、空间注意力模块输出特征;

$\omega_\gamma, \omega_\lambda$ ——每个通道、空间像素的权重;

F_1, F_2 ——通道、空间注意力机制的输入特征。

2.4 基于 Transformer 的全局特征提取方法

Transformer 编码器负责捕获电子舌、电子鼻的全局特征以及长距离依赖关系, 其结构如图 7 所示。编码器主要由两个部分组成: 多头自注意力机制 (multihead selfattention, MHSA) 和前馈神经网络。其中, 多头自注意力机制的计算式为:

$$\begin{cases} \text{MHSA}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \\ \text{head}_h) \mathbf{W}_O, \\ \text{head}_i = \text{Attention}(\mathbf{Q} \mathbf{W}_{Qi}, \mathbf{K} \mathbf{W}_{Ki}, \mathbf{V} \mathbf{W}_{Vi}), \\ \text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q} \mathbf{K}^T}{\sqrt{d_k}}\right) \mathbf{V} \end{cases}, \quad (4)$$

式中:

$\mathbf{Q}, \mathbf{K}, \mathbf{V}$ ——查询 (query)、键 (key) 和值 (value) 矩阵;

d_k ——键和查询的维度;

$\mathbf{W}_{Qi}, \mathbf{W}_{Ki}, \mathbf{W}_{Vi}, \mathbf{W}_O$ ——模型需要学习的参数矩阵;

h ——头的数量。

将电子舌和电子鼻信号时频图特征分别作为多头自注意力机制的输入, 得到 $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ 矩阵。多头自注意力机制通过计算 $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ 之间的相互关系, 为每个特征计算一

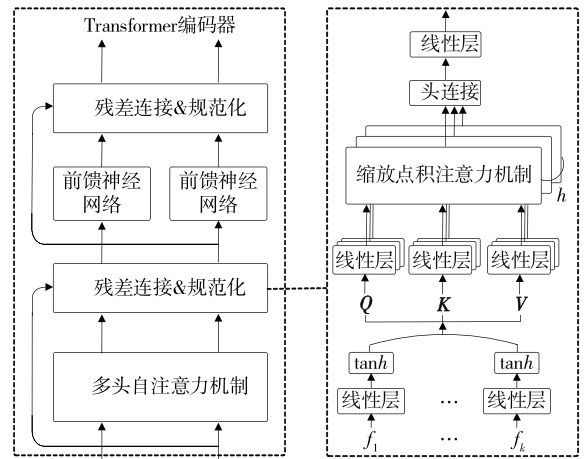


图 7 Transformer 编码器结构

Figure 7 Structure of transformer encoder

个注意力分数 QK^T , 促使模型关注关键区域。每个头在不同的表示子空间中执行此操作, 允许模型并行捕获长距离依赖关系, 得到全局时序特征表示。

2.5 基于多头自注意力机制的特征融合策略

在多传感器系统中, 将特征直接拼接是一种常见策略。但是, 该方法无法充分挖掘特征间的相关性和互补性, 限制融合效果。试验基于 Transformer 的多头自注意力机制实现电子舌和电子鼻特征信息的融合, 过程如图 8 所示。

电子舌特征 X_{tongue} 由 K_1 个特征组成, 表示为 $\{f_{t,1(x)}, \dots, f_{t,K_1(x)}\}$, 电子鼻特征 X_{nose} 由 K_2 个特征组成, 表示为 $\{f_{n,1(y)}, \dots, f_{n,K_2(y)}\}$ 。构建两个 Transformer 编码器, 分别用于电子舌、电子鼻特征融合, 对于每一个电子舌或电子鼻输入特征, 通过多头自注意力机制分别计算其与其他所有特征的注意力权重, 得到注意力权重向量 $r_i(x)$ 或 $r_i(y)$ 。利用余弦相似性函数衡量两个 Transformer 编码器不同头生成的注意力权重向量之间的相似度, 即通过计算每个头生成的注意力权重向量 $r_i(x)$ 与 $r_i(y)$ 间的余弦相似度, 并求平均值, 得到平均相

似度 $s(x, y)$, 得到一个全面的特征相似度度量。最后, 对应位置的电子舌特征 X_{tongue} 和电子鼻特征 X_{nose} 分别与平均相似度 $s(x, y)$ 进行加权, 得到最终的融合特征 X_{fusion} 。该过程可以表示为:

$$\begin{cases} r(x) = \text{MHSA}_t(f_{t,1(x)}, \dots, f_{t,K_1(x)}), \\ r(y) = \text{MHSA}_n(f_{n,1(y)}, \dots, f_{n,K_2(y)}), \\ \text{similarity}[r(x), r(y)] = \frac{r(x) \cdot r(y)}{\|r(x)\| \cdot \|r(y)\|}, \\ s(x, y) = \frac{1}{h} \sum_{i=1}^h \text{similarity}[r_i(x), r_i(y)], \\ X_{\text{fusion}} = \text{Fusion}[s(x, y) \cdot X_{\text{tongue}}, s(x, y) \cdot X_{\text{nose}}] \end{cases} \quad (5)$$

式中:

MHSA——多头自注意力机制;

similarity——余弦相似度函数;

h ——多头自注意力机制中头的数量;

X_{fusion} ——电子舌和电子鼻的融合特征。

3 试验结果与分析

3.1 电子舌和电子鼻信号

由图 9 可知, 绿茶样本溶液经检测后, 不同电极区域显示不同种类样本的指纹信息。电子舌可以采集到 8 000 个数据点 ($1\,000 \times 8$ 个电极); 检测开始后, 电子鼻传感器两端的电压迅速上升, 到达峰值后又逐步趋于稳定。对于每个绿茶样本, 电子鼻传感器可以采集到 6 000 个数据点 ($1\,000 \times 6$ 个气敏传感器)。

3.2 STFT 变换

对电子舌、电子鼻信号分别进行 STFT 变换, 得到如图 10 所示的时频图, 其分辨率为 224×224 。由图 10(a) 可知, 电子舌信号在低频区域显示出高能量分布, 可能是电子舌对绿茶样本溶液中的特定化学成分增强响应, 而低频信号与这些化学成分的浓度有关。由图 10(b) 可知, 电子鼻信号能量在时域内分布变化较大, 可能是由于不同的传感器对绿茶样本中的化学成分的敏感度不同, 导致输出存在较大差异。

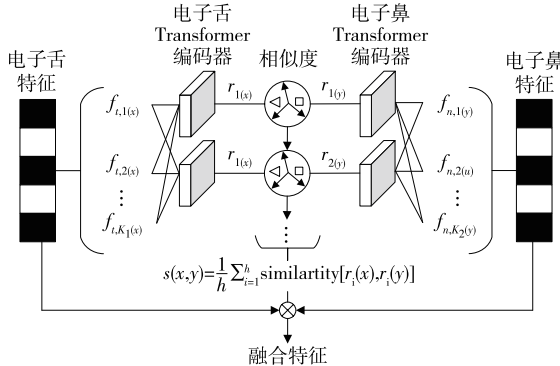


图 8 Transformer 编码器实现电子舌、电子鼻特征融合过程图

Figure 8 Transformer encoder to realize the fusion process of electronic tongue and electronic nose features

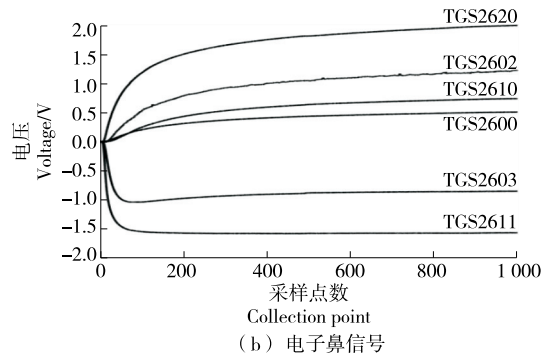
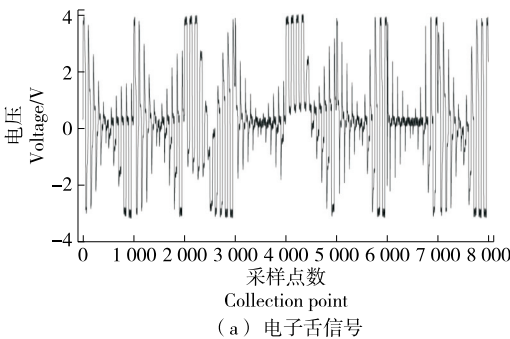
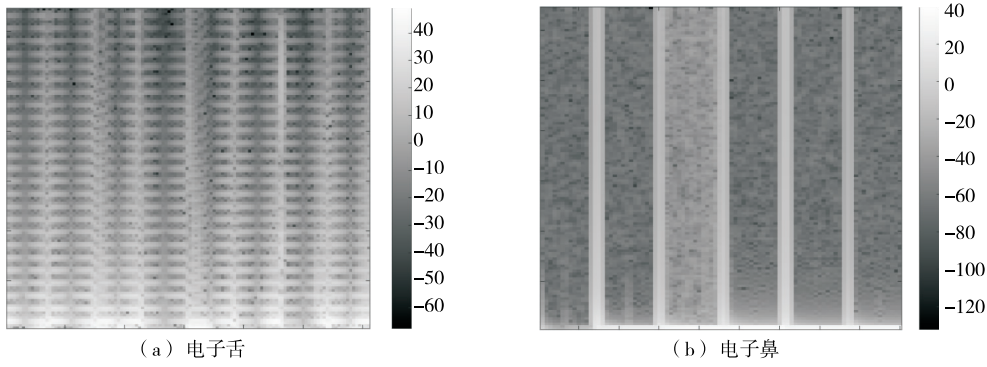


图 9 电子舌和电子鼻信号

Figure 9 Electronic tongue and nose signals



颜色深浅反映了信号能量在时频域的分布特征,深色表示能量较高,浅色表示能量较低

图 10 电子舌和电子鼻信号时频图

Figure 10 Time-frequency diagram of electronic tongue and electronic nose signals

3.3 模型结构和参数优化

为了对模型的超参数、网络架构进行选择和优化,分别采用模型在测试集上的准确率(Accuracy)、精确率(Precision)、召回率(Recall)、F₁分数(F₁-Score)作为模型性能的评价指标^[22]。

3.3.1 CNN 结构优化 为了确定性能最佳的 CNNBlock 数量,设置 CNNBlock 数量范围为[1,5],步长为 1。采用不同数量 CNNBlock 模型训练 5 次,验证集和测试集分类准确率如图 11 所示,试验表明,当 CNNBlock 数量为 4 时,分类效果最优。

CNNBlock 的选择性核卷积大小影响模型的局部特征提取能力,较大的卷积核能捕获更广泛的空间信息,但可能导致局部特征失真;较小的卷积核能精准抓取局部信息,但可能限制信息融合效果。为筛选最优的 SKConv 卷积核,设计 Conv1×Conv2 大小组合分别为 3×3 和 3×3、3×3 和 5×5、5×5 和 5×5、7×7 和 7×7 的模型训练 5 次,分析其训练过程中验证集准确率的变化,结果如图 12 所示。由图 12 可知,当卷积核的大小组合为 3×3 和 5×5 时,模型验证集的准确率达到峰值且更加稳定,

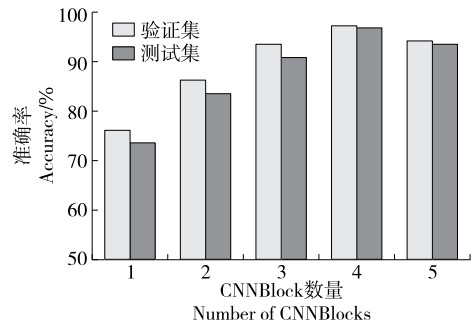
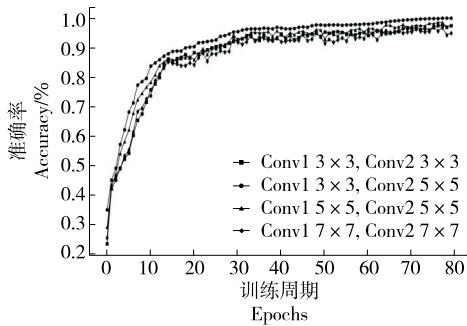


图 11 不同 CNNBlock 数量的模型的验证集和测试集准确率

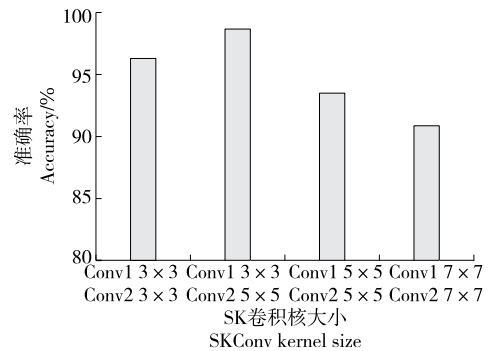
Figure 11 Accuracy of validation and test sets for models with different number of CNNBlocks

其测试集平均准确率也优于其他模型。因此,选择 3×3 和 5×5 为 SKConv 卷积核的最优组合。

3.3.2 模型超参数优化 采用单变量优化法对 CNN-Transformer 模型的超参数设置进行调优,超参数选择范围见表 1,模型训练 5 次,优化结果如图 13 所示。通过对比测试集上各种超参数的分类性能,确定最优的模型超



(a) 验证集



(b) 测试集

图 12 不同选择性核卷积大小组合模型的验证集和测试集准确率

Figure 12 Accuracy of validation and test sets for models with different combinations of selective kernel convolution sizes

参数,其中学习率为 0.001、训练周期为 80、批处理大小为 20、激活函数为 ReLU6。

3.4 消融试验

为了验证 SKConv 选择性核卷积、NAM 注意力模块等改进方式对 CNN-Transformer 模型性能的提升效果,引入不同的增强模块对基础的 CNN-Transformer 模型进行改进,并通过消融试验对其有效性进行验证,结果见表 2。在控制变量法的基础上,不同增强模块采用相同的

表 1 CNN-Transformer 模型超参数选择范围
Table 1 CNN-Transformer model hyperparameter selection range

超参数	选择范围
学习率	[0.005,0.001,0.000 5,0.000 1]
批处理大小	[18,20,24,32]
训练周期	[40,60,80,100]
激活函数	[ReLU,LeakyReLU,ReLU6,ELU]

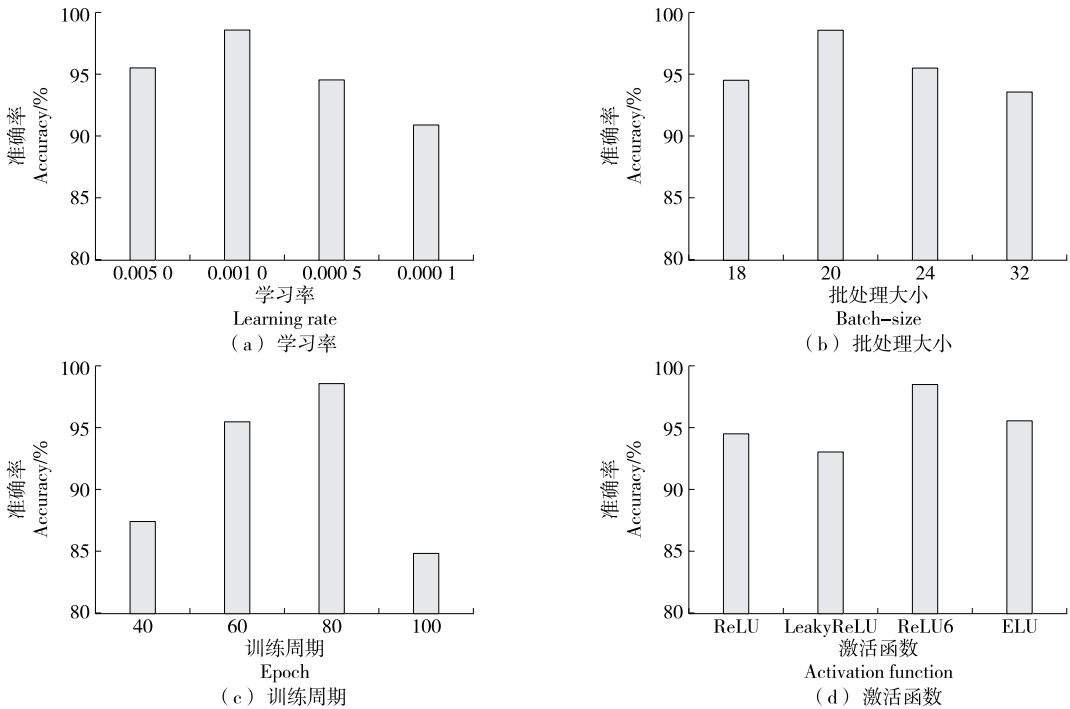


图 13 CNN-Transformer 模型在不同超参数下测试集的准确率

Figure 13 Accuracy of CNN-Transformer model on test set with different hyperparameters

表 2 CNN-Transformer 模型的消融试验
Table 2 Ablation experiment of CNN-Transformer model

序号	因素		准确率/ %	F ₁ 分数/ %
	选择性核卷积 (SKConv)	注意力模块 (NAM)		
0			92.31	92.00
1	✓		95.98	96.00
2		✓	94.72	94.70
3	✓	✓	99.00	99.00

超参数设置。在 CNN-Transformer 模型中采用 SKConv 选择性核卷积代替传统卷积,模型准确率和 F₁ 分数分别提升了 3.67%,4.00%,证明 SKConv 能够有效捕获图像的局部特征。引入 NAM 注意力模块后,模型准确率和 F₁ 分数较基础模型提升了 2.41%,2.70%,证明 NAM 注意力模块能够更好地实现空间和通道间的信息交互,为

模型提供更加有区分性的特征。综合使用 SKConv 选择性核卷积、NAM 注意力模块,模型准确率提升了 6.69%,F₁ 分数提升了 7.00%,证明经过改进的 CNN-Transformer 模型具备较好的特征提取能力和分类效果。

3.5 信息融合性能分析

为了验证试验提出的信息融合模型的有效性,分别采用电子舌、电子鼻数据集,以及融合数据集在 CNN-Transformer 模型上进行训练,利用测试集建立混淆矩阵,结果如图 14 所示,其中 1~5 分别代表西湖龙井、黄山毛峰、信阳毛尖、崂山绿茶和日照绿茶。结果表明,单独采用电子鼻数据集,总体分类准确率为 72%;单独采用电子舌数据集,总体分类准确率为 93%,与电子鼻的响应信号相比,电子舌的响应信号中包含了更丰富的特征信息。采用电子舌、电子鼻融合数据集,总体分类准确率达 99%,表明通过融合多感官特征可以从味觉和嗅觉多感官角度提升模型分辨性能,有助于准确区分不同种类的绿茶样本。

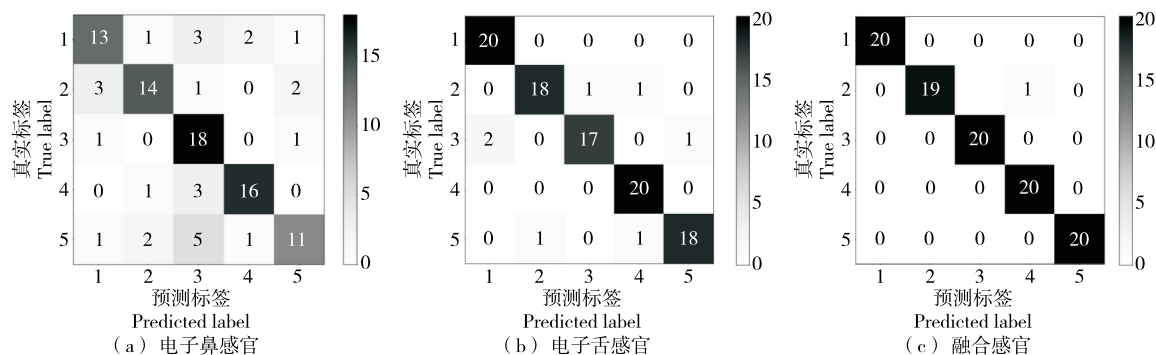


图 14 测试集混淆矩阵对比

Figure 14 Confusion matrix comparison at test set

采用 t-SNE 算法结合 CNN-Transformer 模型对电子鼻、电子舌数据集以及融合数据集提取的特征进行可视化分析^[23],结果如图 15 所示。由图 15 可知,电子鼻、电子舌的每一类样本特征数据点基本上能够形成聚类,但存在一定数量的分类误差,且电子舌特征的可视化展

出了相对于电子鼻更好的聚类效果。每类样本的电子舌和电子鼻融合特征数据点完全形成聚类,且不同类别之间的间隔明显,表明电子舌和电子鼻的数据信息具有互补性,使用融合数据集分辨样本的能力优于单独使用电子舌或电子鼻数据集。

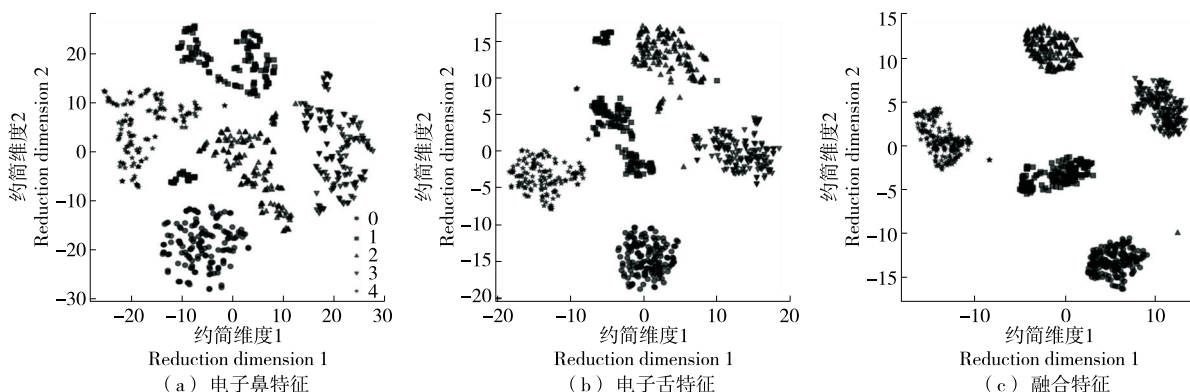


图 15 特征提取数据以及融合数据的 t-SNE 特征可视化

Figure 15 T-SNE feature visualization of feature extraction data and fusion data

3.6 模型对比试验

为了验证试验提出模型的性能,分别采用 CNN、ResNet 50、MobileNet v2、ViT 等深度学习模型与 CNN-Transformer 模型进行性能对比,结合融合感官数据集进行绿茶种类识别。分别采用网格搜索法(grid search)对所有模型的学习率、批处理大小、优化器、损失函数等超参数进行超参数搜索和调优。每个模型在融合感官数据

集上的最佳超参数组合及其测试集结果见表 3,结果表明:CNN-Transformer 模型的各项指标均优于其他模型,其测试集准确率、精确率、召回率以及 F_1 -Score 分别达到 99.00%,99.05%,99.00%,99.00%,相较于其他深度学习模型,CNN-Transformer 模型的准确率提升了 4%~15%,精确率提升了 4.05%~15.02%,召回率提升了 4%~15%, F_1 分数提升了 4%~15%。这是因为 CNN-

表 3 模型分类性能对比

Table 3 Comparison of classification performance of models

模型	学习率	批处理大小	优化器	准确率/%	精确率/%	召回率/%	F_1 分数/%
CNN	0.001	32	Adam	84.00	84.03	84.00	84.00
ResNet 50	0.000 1	16	SGD	92.00	92.03	92.00	92.00
MobileNet v2	0.001	32	Adam	97.00	97.10	97.00	97.00
ViT	0.000 1	20	Adam	95.00	95.04	95.00	95.00
CNN-Transformer	0.001	20	Adam	99.00	99.05	99.00	99.00

Transformer 模型通过引入多头自注意力机制获取上下文信息来提取时频图的全局特征信息,提升了模型的表达能力。同时,CNN 提升了模型的局部特征提取能力,有效缓解了 Transformer 模型忽视图像局部结构信息的问题。CNN-Transformer 模型结合了 CNN、Transformer 模型的优势,分类准确性相较于单独使用 CNN 模型或 Transformer 模型效果更好。

4 结论

研究提出了一种基于电子舌和电子鼻结合 CNN-Transformer 模型的绿茶种类识别方法。结果表明,该方法突破了单一传感器难以获取样本完整信息的局限,所提出的 CNN-Transformer 模型通过引入选择性核卷积和归一化注意力解决了传统的固定卷积核难以适应具有不同方向、尺度特征的数据问题,通过引入多头自注意力机制为特征融合策略增添了新的维度,具有特征提取充分、识别精度高等特点,超越了现有研究中单一深度学习模型的性能。研究的信息融合策略只采用特征级融合一种方法,后期可联用数据级融合、决策级融合方法进行对比试验,寻求最优的信息融合策略。

参考文献

- [1] 陆德彪, 苏祝成. 我国地理标志(产品)保护现状与问题: 以茶叶为例[J]. 中国茶叶, 2009, 31(4): 21-23, 36.
- [2] 童阳, 艾施荣, 吴瑞梅, 等. 茶叶外形感官品质的计算机视觉分级研究[J]. 江苏农业科学, 2019, 47(5): 170-173.
- [3] 郭艺丹. 茶叶检测评审研究进展与关键技术分析[J]. 茶叶学报, 2023, 64(1): 10-20.
- [4] 陈美丽, 唐德松, 龚淑英, 等. 绿茶滋味品质的定量分析及其相关性评价[J]. 浙江大学学报(农业与生命科学版), 2014, 40(6): 670-678.
- [5] 龙立梅, 宋沙沙, 李柰, 等. 3 种名优绿茶特征香气成分的比较及种类判别分析[J]. 食品科学, 2015, 36(2): 114-119.
- [6] 刘奇, 欧阳建, 刘昌伟, 等. 茶叶品质评价技术研究进展[J]. 茶叶科学, 2022, 42(3): 316-330.
- [7] LIU Q, OUYANG J, LIU C W, et al. Research progress of tea quality evaluation technology[J]. Journal of Tea Science, 2022, 42(3): 316-330.
- [8] LIU T, CHEN Y, LI D, et al. Electronic tongue recognition with feature specificity enhancement[J]. Sensors, 2020, 20(3): 772.
- [9] HINES E L, LLOBET E, GARDNER J. Electronic noses: a review of signal processing techniques [J]. IEE Proceedings-Circuits, Devices and Systems, 1999, 146(6): 297-310.
- [10] KUMAR S, GHOSH A. Afeature extraction method using linear model identification of voltammetric electronic tongue [J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(11): 9 243-9 250.
- [11] CHEN J, YANG Y, DENG Y, et al. Aroma quality evaluation ofDianhong black tea infusions by the combination of rapid gas phase electronic nose and multivariate statistical analysis[J]. LWT, 2022, 153: 112496.
- [12] LU L, DENG S, ZHU Z, et al. Classification ofrice by combining electronic tongue and nose[J]. Food Analytical Methods, 2015, 8(8): 1 893-1 902.
- [13] BANERJEE M B, ROY R B, TUDU B, et al. Black tea classification employing feature fusion of e-nose and e-tongue responses[J]. Journal of Food Engineering, 2019, 244: 55-63.
- [14] WEI Z, YANG Y, WANG J, et al. The measurement principles, working parameters and configurations of voltammetric electronic tongues and its applications for foodstuff analysis[J]. Journal of Food Engineering, 2018, 217: 75-92.
- [15] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84-90.
- [16] AZAD R, AL-ANTARY M, HEIDARI M, et al. TransNorm: transformer provides a strong spatial normalization mechanism for a deep segmentation model[M]. [S.l.]: IEEE Access, 2022: 108205.
- [17] WANG E, YU Q, CHEN Y, et al. Multi-modal knowledge graphs representation learning via multi-headed self-attention [J]. Information Fusion, 2022, 88: 78-85.
- [18] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30: 5 998-6 008.
- [19] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale [J/OL]. arXiv, 2020. (2020-10-22) [2023-09-20]. <https://arxiv.org/abs/2010.11929>.
- [20] GAO G W, WANG Z X, LI J C, et al. Lightweight bimodal network for single-image super-resolution via symmetric CNN and recursive transformer[J/OL]. arXiv, 2022. (2022-04-28) [2023-09-20]. <https://arxiv.org/abs/2204.13286>.

(下转第 52 页)

- components of dried shiitake mushroom, *Lentinula edodes* I: relationship between sensory evaluations and amounts of odorous components[J]. Journal of Wood Science, 2004, 50(4): 358-364.
- [32] 张毅航, 方东路, 仲磊, 等. 热风干燥和真空冷冻干燥对猴头菇不同部位风味物质的影响[J]. 食品工业科技, 2022, 43(8): 58-67.
- ZHANG Y H, FANG D L, ZHONG L, et al. Effects of hot air-drying and vacuum freeze-drying on flavor components in different parts of *Hericium erinaceus*[J]. Science and Technology of Food Industry, 2022, 43(8): 58-67.
- [33] TIAN Y T, ZHAO Y T, HUANG J J, et al. Effects of different drying methods on the product quality and volatile compounds of whole shiitake mushrooms[J]. Food Chemistry, 2016, 197: 714-722.
- [34] LIU S X, LAAKSONEN O A, MARSOL-VALL A, et al. Comparison of volatile composition between alcoholic bilberry beverages fermented with non-saccharomyces yeasts and dynamic changes in volatile compounds during fermentation[J]. Journal of Agricultural and Food Chemistry, 2020, 68(11): 3 626-3 637.
- [35] FLAIG M, QI S, WEI G D, et al. Characterization of the key odorants in a high-grade Chinese green tea beverage (*Camellia sinensis*; Jingshan cha) by means of the sensomics approach and elucidation of odorant changes in tea leaves caused by the tea manufacturing process [J]. Journal of Agricultural and Food Chemistry, 2020, 68(18): 5 168-5 179.
- [36] SCALONE G L L, LAMICHHANE P, CUCU T, et al. Influence of free amino acids, oligopeptides and polypeptides on the formation of pyrazines in Maillard model systems[J]. Abstracts of Papers of the American Chemical Society, 2016, 63(22): 5 364-5 372.
- [37] 李小林, 陈诚, 黄羽佳, 等. 顶空固相微萃取—气质联用分析 4 种野生食用菌干品的挥发性香气成分[J]. 食品与发酵工业, 2015, 41(9): 174-180.
- LI X L, CHEN C, HUANG Y J, et al. Analysis of volatile flavors in 4 dried wild edible fungi by HS-SPME-GC IMS[J]. Food and Fermentation Industries, 2015, 41(9): 174-180.
- [38] 李文, 杨焱, 陈万超, 等. 不同干燥方式对香菇含硫风味化合物的影响[J]. 食用菌学报, 2018, 25(4): 71-79.
- LI W, YANG Y, CHEN W C, et al. Effect of drying method on sulfur-containing components in *Lentinula edodes*[J]. Acta Edulis Fungi, 2018, 25(4): 71-79.
- [39] 陈瑞娟, 毕金峰, 周禹含, 等. 干燥方式对胡萝卜超微粉中挥发性风味物质的影响[J]. 中国食品学报, 2015, 15(1): 250-256.
- CHEN R J, BI J F, ZHOU Y H, et al. Effect of different drying methods on volatile components of carrot superfine powder[J]. Journal of Chinese Institute of Food Science and Technology, 2015, 15(1): 250-256.
- [40] CHEN G T, WU F N, PEI F, et al. Volatile components of white *Hypsizygus marmoreus* detected by electronic nose and HS-SPME-GC-MS: influence of four drying methods[J]. International Journal of Food Properties, 2017, 20(12): 2 901-2 910.
- [41] GARCIA E A, ANSORENA D, ASTIASARAN I, et al. Study of the effect of different fiber coatings and extraction conditions on dry cured ham volatile compounds extracted by solid-phase microextraction (SPME)[J]. Talanta, 2004, 64(2): 458-466.
- [42] 杨铭铎, 龙志芳, 李健. 香菇风味成分的研究[J]. 食品科学, 2006, 27(5): 223-226.
- YANG M D, LONG Z F, LI J. Study on flavor compounds in *Lentinus edodes*[J]. Food Sci, 27(5): 223-226.
- [43] FERNANDES A, BARROS L, BARREIRA J C M, et al. Effects of different processing technologies on chemical and antioxidant parameters of *Macrolepiota procera* wildmushroom[J]. LWT-Food Science and Technology, 2013, 54(2): 493-499.
- [44] ZHANG J, YAGOUB A E, SUN Y H, et al. Role of thermal and non-thermal drying techniques on drying kinetics and the physicochemical properties of shiitake mushroom[J]. Journal of the Science of Food and Agriculture, 2022, 102(1): 214-222.
- [45] 张亮, 侯云丹, 黄健, 等. 加热温度对贻贝挥发性成分的影响[J]. 中国食品学报, 2013(9): 227-233.
- ZHANG L, HOU Y D, HUANG J, et al. Effect of temperature on the volatile compounds of mussel[J]. Journal of Chinese Institute of Food Science and Technology, 2013(9): 227-233.
- [46] WU C M, WANG Z. Volatile compounds in fresh and processed shiitake mushrooms (*Lentinus edodes* Sing.) [J]. Food Science & Technology Research, 2000, 6(3): 166-170.
- [47] CHEN D, QIN L, GENG Y, et al. The aroma fingerprints and discrimination analysis of shiitake mushrooms from three different drying conditions by GC-IMS, GC-MS and DSA[J]. Foods, 2021, 10(12): 2 991.
-
- (上接第 42 页)
- [20] YANG L, YANG Y, YANG J, et al. FusionNet: A convolution-transformer fusion network for hyperspectral image classification [J]. Remote Sensing, 2022, 14(16): 4 066.
- [21] ZHANG Q, DENG L. An intelligent fault diagnosis method of rolling bearings based on short-time fourier transform and convolutional neural network[J]. Journal of Failure Analysis and Prevention, 2023, 23(2): 795-811.
- [22] CHINCHOR N. Muc-4 evaluation metrics[C]// Proceedings of the 4th Conference on Message Understanding. [S.l.]: Association for Computational Linguistics, 1992: 22-29.
- [23] 王首程, 于雪莹, 高继勇, 等. 基于电子舌和电子鼻结合 DenseNet-ELM 的陈醋年限检测[J]. 食品与机械, 2022, 38(4): 72-80, 133.
- WANG S C, YU X Y, GAO J Y, et al. Age detection of mature vinegar based on electronic tongue and electronic nose combined with DenseNet-ELM[J]. Food & Machinery, 2022, 38(4): 72-80, 133.