

电子舌和电子眼结合 Informer-Swin Transformer 多层级信息融合的大豆产地溯源检测

李 鑫 王志强 韩天瑞 徐苏超

(山东理工大学计算机科学与技术学院, 山东 淄博 255049)

摘要: [目的] 实现对不同产地大豆的快速溯源检测。[方法] 提出一种基于电子舌和电子眼信息融合的大豆产地快速溯源检测新方法。使用电子舌和电子眼分别采集不同产地大豆的味觉指纹和视觉图像信息, 针对电子舌信号时序特征和电子眼图像空间特征的差异, 分别采用 Informer 和 Swin Transformer 提取两类数据中的关键特征。利用双模特征交互模块对两类特征进行深度交互以获取多尺度层间互信息, 使用交叉注意力融合模块整合层间互信息实现跨模态信息的高效融合与分类识别。[结果] 该模型对不同产地大豆具有较好的分辨能力, 准确率、查全率、召回率和 F_1 -Score 分别达到 98.33%、98.34%、98.33% 和 0.983 3。[结论] 该方法为大豆产地的快速溯源检测提供了一种高效、可靠的新思路。

关键词: 大豆; 电子舌; 电子眼; Informer; Swin Transformer; 交叉融合

Geographical origin tracing of soybeans using electronic tongue and electronic eye combined with multi-level information fusion based on Informer-Swin Transformer

Li Xin Wang Zhiqiang Han Tianrui Xu Suchao

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

Abstract: [Objective] To achieve rapid traceability detection of soybeans from different geographical origins. [Methods] A novel method for rapid soybean origin traceability is proposed based on the fusion of information from electronic tongue and electronic eye. The electronic tongue and electronic eye are used to collect taste fingerprint and visual image information, respectively, from soybeans of different origins. To address the differences between the temporal sequence features of electronic tongue signals and the spatial image features of electronic eye data, Informer and Swin Transformer are respectively employed to extract key features from the two types of data. A dual-modal feature interaction module is designed to perform deep interaction between the two types of features, thereby obtaining multi-scale inter-layer mutual information. A cross-attention fusion module is then applied to integrate the inter-layer mutual information, enabling efficient cross-modal information fusion and classification recognition. [Results] The proposed model has good discriminative ability for soybeans from different origins, achieving an accuracy of 98.33%, precision of 98.34%, recall of 98.33%, and F_1 -Score of 0.983 3. [Conclusion] This method provides an efficient and reliable new approach for the rapid traceability detection of soybean geographical origin.

Keywords: soybean; electronic tongue; electronic eye; Informer; Swin Transformer; cross-integration

大豆是一种重要的经济作物,在世界各地广泛种植和消费^[1]。由于生长环境、气候条件、土壤性质等因素的差异,不同产地的大豆在风味、营养成分和价格上存在显著差异。目前,大豆产地的鉴别主要使用核磁共振^[2]、近

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)

通信作者: 王志强(1977—),男,山东理工大学教授,博士。E-mail: wzq@sdut.edu.cn

收稿日期: 2025-10-28 **改回日期:** 2026-03-19

引用格式: 李鑫,王志强,韩天瑞,等. 电子舌和电子眼结合 Informer-Swin Transformer 多层级信息融合的大豆产地溯源检测[J]. 食品与机械, 2026, 42(9): 65-73.

Citation: Li Xin, Wang Zhiqiang, Han Tianrui, et al. Geographical origin tracing of soybeans using electronic tongue and electronic eye combined with multi-level information fusion based on Informer-Swin Transformer[J]. Food & Machinery, 2026, 42(9): 65-73.

红外光谱^[3]、色谱^[4]等技术。这些检测方法价格昂贵、操作复杂、流程冗长,难以满足大豆产地快速原位检测的需求。

近年来,仿生检测技术因其响应快、可靠性高、成本低等优点被广泛应用于农产品检测和分析。电子舌作为一种新型仿生检测技术,通过多传感阵列与模式识别技术的结合实现对液态样品的定性定量分析,被广泛应用于食品溯源^[5]、医药检测^[6]、环境监测^[7]等领域。电子眼则是一种利用摄像头和图像识别算法对样品进行光学成像分析的检测技术,被广泛应用于农业和工业检测^[8-9]。然而,大豆化学成分复杂、形态多样,其典型特征不仅与蛋白质、脂肪等内部化学组分密切相关,也与色泽、形态等外观特征有关,单独使用电子舌或电子眼进行检测存在信息感知能力单一、样本特征表征不足等问题,容易导致检测结果不可靠。

Orlandi 等^[10]构建了包含电子舌和电子眼的融合数据矩阵,并使用偏最小二乘回归(PLS)实现了葡萄成熟度的监测。Zhou 等^[11]使用皮尔逊相关分析提取了电子舌、电子鼻的主要特征并拼接,应用主成分分析(PCA)区分银耳的发酵程度。但是,传统传感器融合技术主要采用特征串联或加和操作,容易受到局部特征间相互作用的限制,难以实现深层次的信息整合^[12]。同时,传统识别技术主要基于机器学习方法进行检测,往往存在有效信息丢失和过拟合的问题^[13]。

深度学习能从海量数据中自主挖掘核心特征并实现端到端的模式识别,具有强大的表征能力和广泛的适用性。Yang 等^[14]使用卷积神经网络(CNN)结合电子舌和电子眼对普洱茶贮藏时间进行了鉴别。Wang 等^[15]使用 CNN 和注意力机制结合电子舌和电子眼对胡椒产地进行了溯源。然而,卷积神经网络依赖于局部感受野,对全局和长距离特征的捕捉不充分^[16]。Transformer 通过自注意力机制实现全局特征关联,具有泛化能力强、可拓展性高等优点,被广泛应用于自然语言和音频处理领域^[17-18]。但是,传统 Transformer 模型普遍面临计算开销大、长距离特征提取能力有限等问题。Informer 通过稀疏自注意力机制解决了 Transformer 计算复杂度高、长序列信息利用率低的问题^[19],在处理长时间序列信号时具有显著优势。Swin Transformer 则在继承 Vision Transformer 优点的基础上,引入了层级结构和滑动窗口机制,能够有效捕捉图像的局部与全局特征,同时显著降低计算复杂度^[20]。

试验拟提出一种电子舌和电子眼结合 Informer 和 Swin Transformer 的大豆产地溯源检测新方法。利用电子舌和电子眼分别采集来自 5 个不同产区大豆样品的味觉指纹和视觉图像数据,采用 Informer 和 Swin Transformer 模型分别提取两类数据的特征信息。针对传统模型数据融合不充分的问题,提出一种多层次信息融合方法,并应用交叉注意力融合模块对提取的电子舌和电子眼特征进行逐级融合和分类,旨在有效实现大豆产地的溯源检测。

1 材料与方法

1.1 试验材料

大豆样品分别来自云南楚雄、山西忻州、广东阳江、江苏宿迁和河南商丘 5 个不同地区,所有样品均符合 GB 1352—2023《大豆》标准。采集后的大豆样品经过杂质去除处理,通过自然晾晒的方式进行风干,未进行脱皮、破碎或其他精加工处理。干燥后的大豆样本装入无菌食品级聚乙烯塑料袋中,于室温、干燥、避光环境中保存。

1.2 电子舌系统

电子舌系统结构如图 1 所示,该系统由 4 个主要部分组成:① 传感器阵列,采用三电极体系,包含 8 个工作电极(铂、金、钛、钨、银、钨、镍和玻碳),1 个 Ag/AgCl 参比电极和 1 个铂辅助电极;② 信号调理模块,负责驱动三电极体系,实现采集信号的放大与滤波处理;③ 数据采集卡,利用其 D/A 模块输出电极激励信号,通过 A/D 模块完成对传感器阵列响应信号的采集;④ LabVIEW 上位机软件,用于控制检测流程并分析电极响应信号。电子舌系统的工作流程:首先,上位机软件操控数据采集卡发出大幅方波脉冲激励信号,经信号调理模块施加至传感器阵列。各电极在激励信号作用下发生电化学反应并产生微弱电流响应,经过放大与滤波后输入数据采集卡进行模数转换。最终,转换后的信号被传输至上位机进行处理和分析。

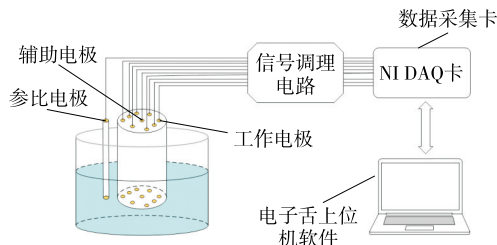


图 1 电子舌系统

Figure 1 Electronic tongue system

1.3 电子眼系统

电子眼系统结构如图 2 所示,该系统由电子眼上位机软件、500 万电子目镜、光电目镜、LED 灯及其调节器、支架和底盘 5 个主要部分组成。电子眼系统的工作流程:将待测样品固定于底盘上,调整目镜放大倍率、LED 光源亮度以及支架高度以获得清晰的图像。随后对样品进行图像采集,并将所得图像传输至上位机软件进行数据处理与分析。

1.4 模型结构

试验提出的模型如图 3 所示,主要包括 Informer 分支、Swin Transformer 分支、双模特征交互模块 DMIF 和交叉注意力融合模块 CFF 4 个部分。首先,将电子舌信号输入到 Informer 分支以充分提取大豆电子舌信号的深层特

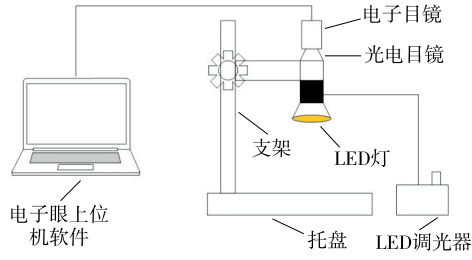


图2 电子眼系统

Figure 2 Electronic eye system

征,该分支由5个 Infrormer stage堆叠而成,每个 stage 输出一组电子舌信号层级特征。电子眼图像输入到 Swin Transformer 分支以提取大豆图像的局部和全局特征,该分支由5个 Swin Transformer stage堆叠而成,每个 stage 输出一组图像层级特征。然后,双模特征交互模块 DMIF 在不同层级上电子舌图谱和电子眼图像特征进行深度交互和整合。最后,采用交叉注意力融合模块 CFF 对整合特征进行逐级融合,融合后的特征被送入全连接层展平,应用 SoftMax 激活函数完成分类任务。

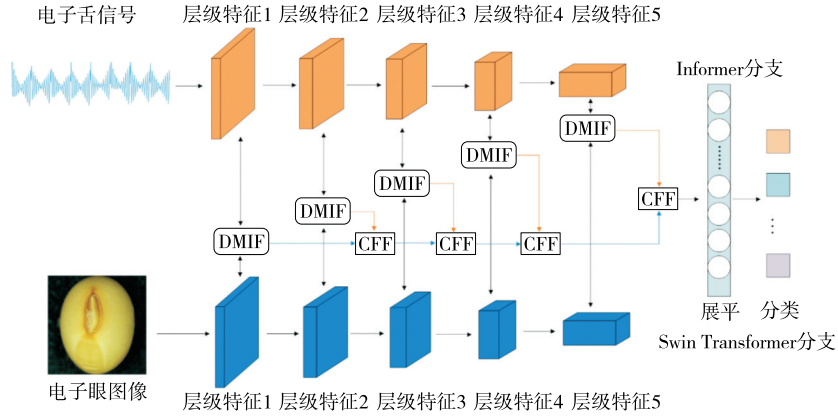


图3 Infor-Swin Transformer模型结构图

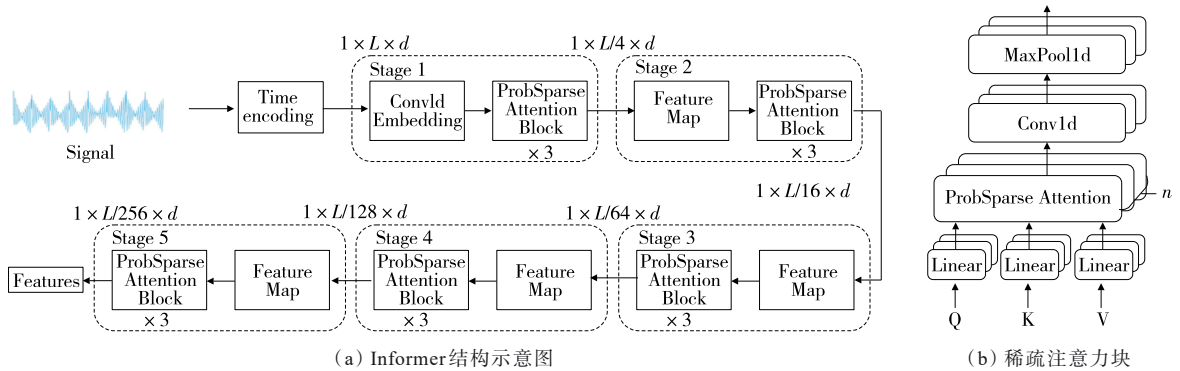
Figure 3 Structure of Infor-Swin Transformer model

1.4.1 Informer Informer是一种针对长序列时间序列预测任务的深度学习模型,Informer在Transformer结构的基础上进行了改进,通过稀疏注意力机制和自注意力蒸馏优化,显著提升了长时序数据的处理效率。试验采用的Informer结构如图4(a)所示。

Informer首先对输入的时间序列信号进行时间编码(Time Encoding)和卷积嵌入(Conv1d Embedding),以提取初始的特征信息。此后,模型通过分阶段堆叠的稀疏自注意力块(ProbSparse Attention Block)构建多尺度特征处理结构,其结构如图4(b)所示,使模型能有效捕捉长序

列中不同时间尺度的重要信息。每个稀疏自注意力块内部引入稀疏自注意力机制,仅对最具代表性的信息进行关注,并使用一维卷积和最大值池化压缩特征维度,大幅降低计算开销。

稀疏自注意力在全局自注意力机制的基础上增加了一个稀疏矩阵来降低计算复杂度,该矩阵仅关注对注意力得分贡献更大的位置。Informer认为低贡献位置的Query近似于均匀分布,如式(1)所示。稀疏自注意力使用KL散度来评估Query的稀疏性,较大的KL散度表示与均匀分布偏差较大,意味着稀疏性增加,按式(2)计算KL散度。



(a) Informer结构示意图

(b) 稀疏注意力块

图4 Informer模型结构和稀疏注意力块示意图

Figure 4 Informer model structure and sparse attention block

$$q(k|q_i) = \frac{1}{L_K}, \quad (1)$$

$$KL(P \parallel Q) = \ln \sum_{j=0}^{L_K} \exp\left(\frac{q_i k_j^T}{\sqrt{d}}\right) - \frac{1}{L_K} \sum_{j=1}^{L_K} \frac{q_i k_j^T}{\sqrt{d}} - \ln L_K, \quad (2)$$

式中:

d ——输入维度;

L_K ——Key 的维度;

q_i ——Query 的第 i 行;

k_j ——Key 的第 j 行。

使用式(3)评估第 i 个 Query 的稀疏性,选取一些贡献度高的 Query 进行自注意力计算,得到稀疏自注意力计算

结果如式(4)所示。

$$\bar{M}(q_i, K) = \max_j \frac{q_i k_j^T}{\sqrt{d}} - \frac{1}{L_K} \sum_{j=1}^{L_K} \frac{q_i k_j^T}{\sqrt{d}}, \quad (3)$$

$$\text{Attention}(Q, K, V) = \text{soft max}\left(\frac{\bar{Q} K^T}{\sqrt{n}}\right) V, \quad (4)$$

式中:

$\bar{M}$ ——单个 Query 的稀疏性;

$\bar{Q}$ ——概率稀疏矩阵(来源于稀疏性 $\bar{M}$)。

1.4.2 Swin Transformer Swin Transformer 源自 Vision Transformer 视觉模型的改进设计,通过分层特征图构建和多尺度窗口注意力机制有效降低了计算复杂度,其结构如图 5(a)所示。

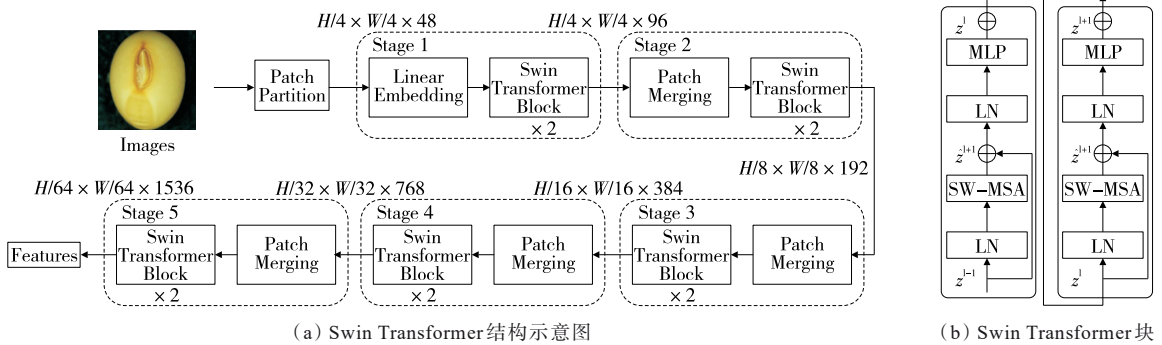


图 5 Swin Transformer 模型结构及其块示意图

Figure 5 Structure and block of Swin Transformer model

Swin Transformer 首先将输入的 RGB 图像分割成互不重叠的块(Patch),然后在首个“Stage”中使用线性嵌入层(Linear Embedding)将这些块投影到更高的维度,这些高维特征块会被送入基于窗口的自注意力机制的 Swin Transformer 块中进行特征提取,其结构如图 5(b)所示。在接下来的“Stage”中,模型会通过 Patch Merging 操作对相邻的 Patch 进行合并,使用若干 Swin Transformer 块处理特征图,从而实现下采样操作,使得模型能够同时捕捉不同尺度的图像特征。

Swin Transformer 块在保持非重叠窗口高效计算的基础上,通过引入跨窗口连接提出了移位窗口划分策略。该策略在相邻的 Swin Transformer 块之间交替使用两种窗口划分方式,分别对应常规窗口和移位窗口的 W-MSA 与 SW-MSA 模块。每个 MSA 模块和 MLP 模块之前都应用层归一化(LayerNorm, LN),并在每个模块之后应用残差连接。Swin Transformer 块的整体计算可表示为:

$$\hat{z}^l = \text{W-MSA}(\text{LN}(z^{l-1})) + z^{l-1}, \quad (5)$$

$$z^l = \text{MLP}(\text{LN}(\hat{z}^l)) + \hat{z}^l, \quad (6)$$

$$\hat{z}^{l+1} = \text{SW-MSA}(\text{LN}(z^{l+1})) + z^{l+1}, \quad (7)$$

$$z^{l+1} = \text{MLP}(\text{LN}(\hat{z}^{l+1})) + \hat{z}^{l+1}, \quad (8)$$

式中:

MLP——多层感知机;

$\hat{z}^l$ ——W-MSA 模块的特征输出;

z^l ——MLP 模块的特征输出;

$\hat{z}^{l+1}$ ——SW-MSA 模块的特征输出;

z^{l+1} ——MLP 模块的特征输出;

W-MSA——常规窗口划分配置的多头自注意力;

SW-MSA——移位窗口划分配置的多头自注意力。

1.4.3 双模特征交互模块 双模特征交互模块 DMIF 的结构如图 6 所示,主要包含空间压缩模块 SC 和交叉特征交互模块 CFI 两个组件。首先,不同模态特征 F_I 和 F_T 经过空间压缩模块 SC 得到压缩空间特征 F_I 和 F_T 。然后,将 F_I 和 F_T 送入交叉特征交互模块 CFI,CFI 模块通过两个并联的交叉特征融合模块 CFF 细化跨模态特征,并通过残差连接添加压缩特征到细化的跨模态特征,得到不同模态的输出特征 F_I' 和 F_T' ,实现不同模态间特征的充分交互。最后,将不同模态特征维度还原并回送,使用 CFF 模块对交互特征 F_I' 和 F_T' 进行最终整合,得到层级融合特征 F_{IT} 。

虽然 Informer 和 Swin Transformer 在提取特征时进行了下采样,但模型的参数规模仍较大,且二维图像特征和一维信号特征间存在维度不齐的问题。为了降低计算复

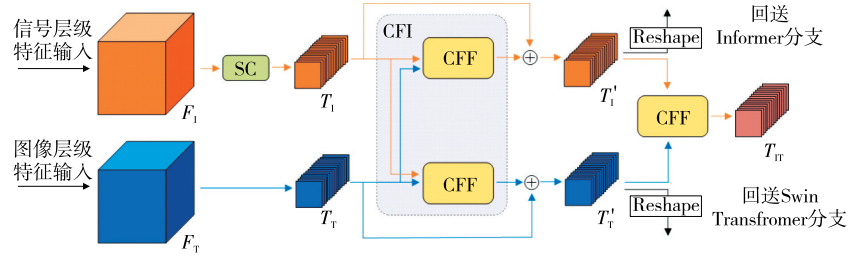


图6 双模特征交互模块

Figure 6 Dual-modal feature interaction module

杂度,同时解决维度不齐的问题,在交叉特征交互模块CFF之前应用空间压缩模块SC。在SC中,使用池化和线性连接两种不同操作对特征图的空间维度进行压缩。对于电子眼特征,首先将二维特征展平[式(9)],然后使用步长为2的最大值池化对输入为 $1 \times L \times C$ 的特征图进行初步压缩[式(10)]。对于电子舌特征,仅使用最大池化进行初步压缩[式(10)]。

$$F_{1 \times L \times C} = \text{Reshape}(F_{H \times W \times C}), L = H \times W, \quad (9)$$

$$F_{\text{pooled}} = \text{MaxPool}_{\text{step2}}(F_{1 \times L \times C}). \quad (10)$$

为了进一步压缩空间特征,同时尽可能地保留细节特征,采用线性连接处理池化后的特征:

$$F_{\text{output}} = \text{Linear}(F_{\text{pooled}}), \quad (11)$$

式中:

F_{pooled} ——池化输出特征图;

F_{output} ——线性连接输出特征图。

1.4.4 交叉特征融合模块 交叉特征融合模块CFF采用交叉注意力机制实现全局不同模态间的信息融合,其结构如图7所示。首先,Informer分支的空间压缩特征 T_I 被投影到两个单独的矩阵 V_I 和 K_I ,Swin Transformer分支的空间压缩特征 T_T 被投影到一个单独的矩阵 Q_T 。然后,通过点积计算得到跨模态的相关性矩阵,并利用softmax函数对其进行归一化处理,以衡量不同模态特征间的相似度。最后,通过关联矩阵与 V_I 相乘获得模态间的细化特征 Z_T , Z_T 经线性层后通过缩放因子 β 和 α 与原始输入特征 T_T 进行残差连接,最终得到包含了来自两个模态的互补信息的融合特征 F_{IT} 。

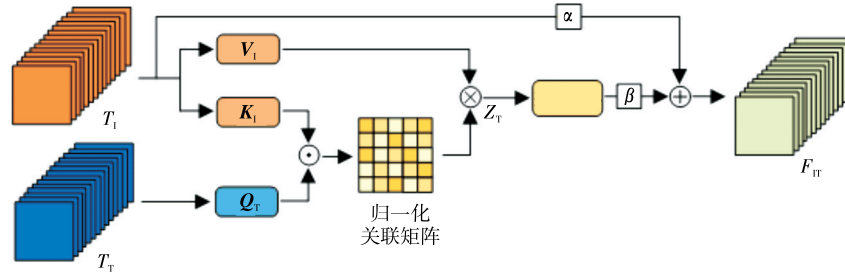


图7 交叉特征融合模块

Figure 7 Cross-feature fusion module

2 试验与结果分析

2.1 电子舌信号采集与处理

在电子舌信号采集过程中,环境温度保持在 $(25 \pm 2)^\circ\text{C}$ 范围。采集信号前,从每个类别中取出50 g完整、干燥的大豆样品,使用高速研磨机将样品研磨成粉末,过200目筛网。取20 g粉末与100 mL蒸馏水混合,搅拌1 min后,使用滤纸过滤溶液两次,使用电子舌系统采集响应信号,大豆样品的电子舌响应信号如图8所示。由于不同电极表面和溶液之间的电化学反应,每个电极都表现出独特的“味觉”特征。每个电极采集1 000个数据点,每个样本共采集8 000个数据点,重复上述步骤,每种大豆采集5 000个电子舌信号。

由图8可知,所有产地样品的信号均呈现周期性波动特征,反映了大豆基质与电子舌电极间的基础电化学反应规律。不同产地大豆的信号波动幅度存在明显区别,如山西忻州产地的大豆在镍电极上的响应信号幅值较低,可能是电极对溶液中不同浓度的氨基酸和蛋白质的敏感度不同,因此,可将其作为区分产地的内部特征依据。

2.2 电子眼图像采集与处理

使用电子眼采集的大豆样品图像的原始分辨率为 $2\,048 \times 1\,536$ 。由于样品在位置和尺寸上存在差异,可能会影响模型训练效果,因此对采集的图像进行缩放和中心裁剪处理。处理后的大豆样品图像如图9所示,分辨率

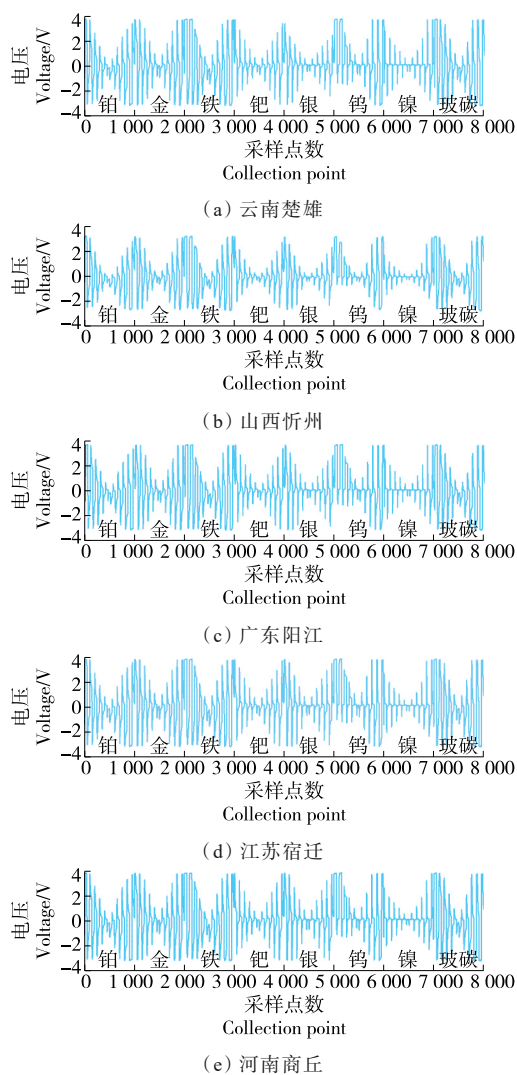


图 8 不同产地大豆电子舌响应信号

Figure 8 Electronic tongue response signals of soybeans from different geographical origins

统一调整至 512×512 , 仅保留大豆主体区域。大豆图像的像素范围为 $[0, 255]$, 直接用于模型训练可能会导致梯度爆炸。为了适应模型训练, 应用最大值归一化将图像原始像素范围映射到 $[0, 1]$, 进而得到背景干扰小且数值标准化的图像数据。

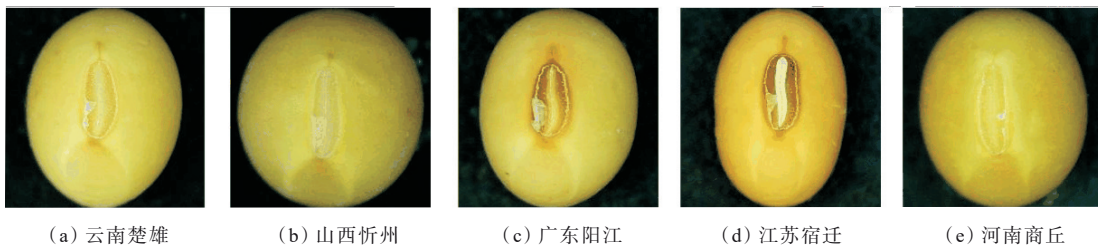


图 9 不同产地大豆的电子眼图像

Figure 9 Electronic eye images of soybeans collected from different geographical regions

由图 9 可知, 不同产地的大豆形态、种脐形状、表皮颜色等存在显著差异, 因此可将其作为区分产地的外观特征依据。

2.3 模型结构优化

在模型结构设计中, Informer 与 Swin Transformer 分支采用成对构建, 以实现多模态特征在相同层级上的对齐与交互。为优化结构深度配置, 试验在保持两个分支 stage 数量一致的前提下, 设定 stage 对数在 $[2, 8]$ 范围内进行网格搜索。通过逐一训练各结构配置, 并在测试集上评估其性能表现, 确定最优的 stage 数量。图 10 为不同 stage 对数配置下模型在测试集上的分类准确率。随着 stage 数的增加, 模型准确率总体呈上升趋势, 并在 stage 数为 5 时达到最高 95.58%, 随后略有下降。说明在一定深度范围内, 增加层级有助于提升特征提取能力, 但过深的结构可能引入过拟合或冗余计算。因此, 选择 stage 数为 5 作为最优配置。

2.4 模型超参数优化

超参数的设置对模型的训练速度和泛化能力有很大影响。模型的超参数通过贝叶斯优化算法 (BOA) 进行确定。BOA 是一种基于贝叶斯理论的全局优化方法, 通过多次评估目标函数来确定全局最优值的超参数组合^[21]。待优化的超参数有学习率、激活函数、优化算法、迭代次数、缩放因子 α 和缩放因子 β , 这些超参数和搜索空间的优化结果见表 1。结果表明, 当学习率为 0.001, 激活函数为 ReLU, 优化算法为 RMSprop, 迭代次数为 80, 缩放因子 α 为 0.6, 缩放因子 β 为 0.4 时, 该模型分类准确率最高。

2.5 模型超参数优化

为了验证试验方法的有效性, 对 3 种策略进行了对比试验: 电子舌结合 Informer、电子眼结合 Swin Transformer 以及双感官融合系统结合 Infor-Swin Transformer。图 11(a) 为训练过程中的损失变化情况, 融合数据集上的训练损失在 80 次迭代后逐渐稳定, 其收敛速度明显快于单一电子舌或电子眼数据集。融合数据集稳定后的损失值为 11.341 8, 低于仅使用电子舌或电子眼数据的情况, 说明融合感官数据能够提升模型的拟合能力。图 11(b) 为测试集上的准确率随迭代次数的变化趋势, 融合感官数据集的准确率上升更快, 最终达到 98.33%, 显著高于电

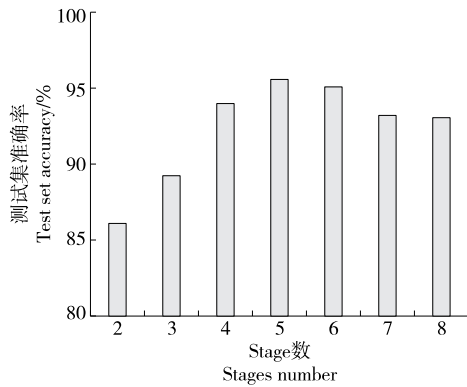


图10 不同stage数量下的测试集准确率

Figure 10 Test set accuracy under different stage numbers

子舌和电子眼的。表明融合感官能充分利用多种信息资源,加快模型收敛,提升分类性能。

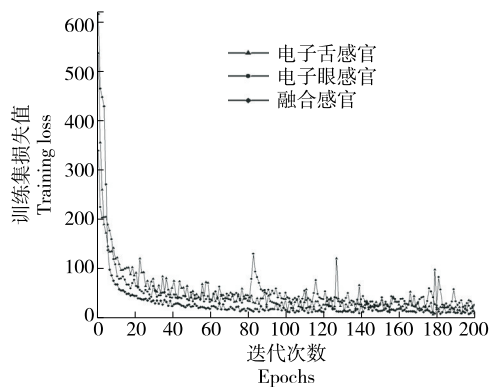
为了评估模型的分类效果,利用测试集建立混淆矩阵,结果如图12所示。模型在单电子舌和单电子眼数据集上分类效果均较差,电子舌数据集的总体分类准确率为79.40%,分类错误主要集中在云南产地,将243个样本

表1 贝叶斯优化算法的搜索空间和搜索结果

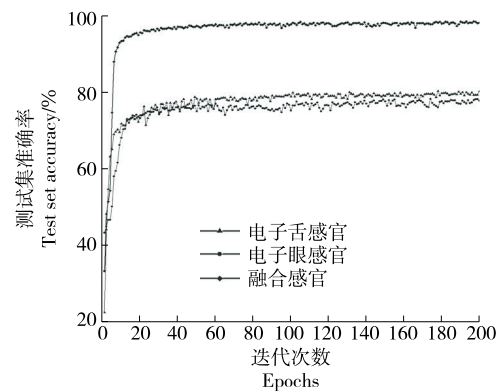
Table 1 Search space and search results of Bayesian optimization algorithm

超参数	搜索空间	搜索结果
学习率	$(1 \times 10^{-4}, 1 \times 10^{-1})$	0.001
激活函数	[ReLU, Leaky ReLU, Sigmoid, tanh]	ReLU
优化算法	[Adam, SGD, RMSprop]	RMSprop
迭代次数	(10, 200)	80
缩放因子 α	(0, 1)	0.6
缩放因子 β	(0, 1)	0.4

错分为广东产地,此外山西、广东、河南产地也分别有2, 29, 33个样本被错误分类。电子眼数据集的总体分类准确率最低,为76.07%,江苏和河南产地分类效果较好,仅有3, 4个样本被错误分类,但云南、山西、广东分别有50, 173, 129个样品被错误分类。而在融合感官数据集上,被错误分类的样本较少,总体分类准确率达到98.33%,说明双模特征交互模块的引入可实现跨模态信息间的充分交互,同时交叉特征融合模块可实现电子舌特征和电子



(a) 训练集损失值



(b) 测试集准确率

图11 训练集损失值和测试集准确率

Figure 11 Training set loss value and test set accuracy

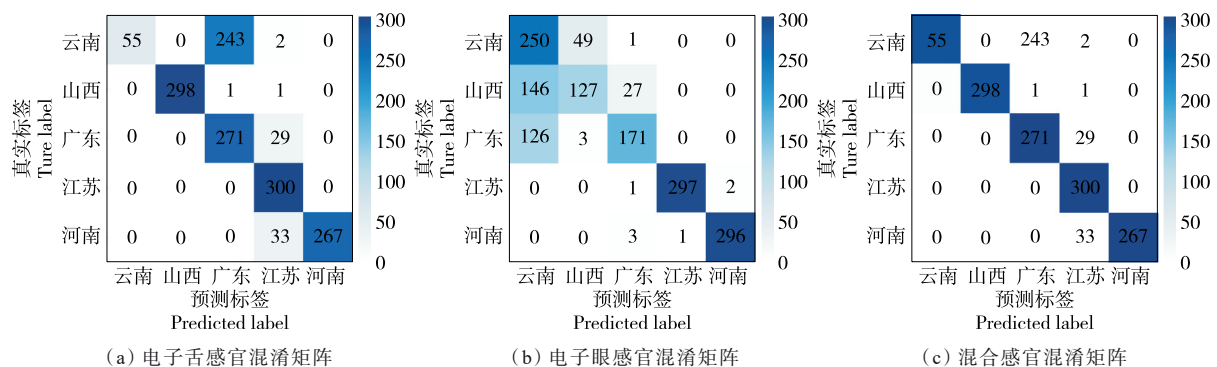


图12 测试集混淆矩阵对比

Figure 12 Testing set confusion matrix comparison

眼特征的有效融合,显著提升了模型性能。

2.6 特征可视化分析

为了直观地显示模型的特征提取效果,采用 t 分布随机近邻嵌入算法(t -SNE)对电子舌、电子眼和融合感官系统的特征提取效果进行可视化分析,结果如图 13 所示。由图 13 可知,电子舌与 Inforner 的组合特征点相对离散,广东和江苏产地未形成比较清晰的类簇,聚类中心不明

确。电子眼和 Swin Transformer 的组合存在特征点混淆现象,云南和广东产地的特征点混叠,同类样本之间的距离也较大。双感官融合系统结合 Infor-Swin Transformer 模型的特征点类别间间距较大,类别内间距较小,同一类别的样本高度聚类,聚类中心明确,说明在信息融合过程中实现了电子舌信息和电子眼信息的互补,具有更好的特征提取能力。

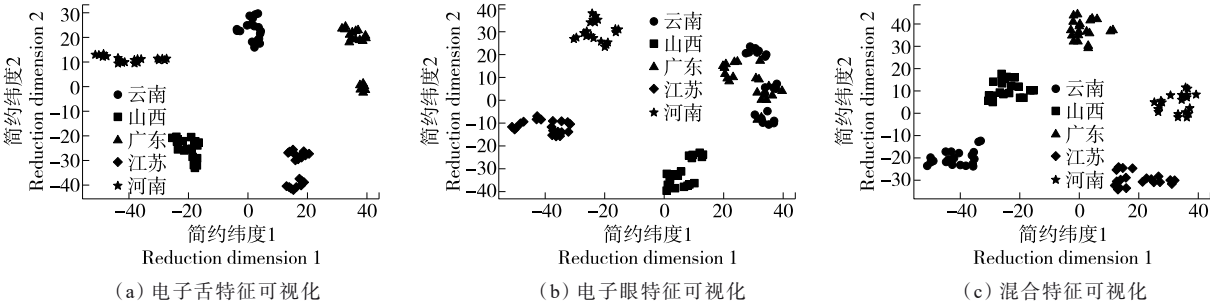


图 13 测试集 t -SNE 特征可视化对比

Figure 13 Comparison of t -SNE feature visualization in testing set

2.7 模型对比分析

为了验证试验模型的性能,分别采用 VGG16、ResNet50、AlexNet、Mamba、Informer 或 Swin Transformer 结合电子舌和电子眼进行大豆产地分类。通过准确率、精确率、召回率和 F_1 -Score 对不同模型的性能进行评价,结果见表 2。

表 2 不同模型的性能对比
Table 2 Comparison of classification performance across various models

数据源	模型	准确率/%	精确率/%	召回率/%	F_1 -Score
电子舌	VGG16	84.20	87.50	84.20	0.832 0
	ResNet50	86.10	88.70	86.10	0.846 5
	AlexNet	81.40	84.00	81.40	0.813 6
	Mamba	88.30	90.50	88.30	0.889 1
	Informer	79.40	86.96	79.40	0.763 1
电子眼	VGG16	72.80	77.00	72.80	0.749 0
	ResNet50	81.30	84.00	81.30	0.826 0
	AlexNet	74.50	78.20	74.50	0.763 4
	Mamba	82.60	85.10	82.60	0.838 2
	Swin Transformer	76.07	80.41	76.07	0.760 3
感官融合	VGG16	94.30	95.10	94.30	0.942 0
	ResNet50	95.80	96.40	95.80	0.961 0
	AlexNet	91.20	92.50	91.20	0.918 5
	Mamba	96.90	97.50	96.90	0.971 6
	Infor-Swin Transformer	98.33	98.34	98.33	0.983 3

由表 2 可知,在单一感官数据源电子舌或电子眼中,不同模型的性能存在一定差异,但整体准确率普遍在 70%~88% 范围内。而在融合电子舌与电子眼后的多模态感官输入中,各模型性能显著提升,准确率普遍 > 90%,其中,融合数据下的 Mamba 模型和 Infor-Swin Transformer 模型表现尤为突出,后者在准确率、精确率、召回率和 F_1 -Score 分别达到 98.33%、98.34%、98.33% 和 0.983 3,各性能指标均优于其他模型,表明其在综合分类能力方面最强。相比之下,单模态下即便是表现较好的 ResNet50 和 Mamba 模型,其准确率仍普遍低于融合模型 5%~20%,证明多源感官融合对提升模型分类精度具有显著作用。

3 结论

研究提出了一种基于电子舌和电子眼结合 Infor-Swin Transformer 模型的大豆产地快速溯源方法。结果表明,与单一电子舌或电子眼系统相比,融合多感官信息的模型能够获得更高的分类准确度。与多种主流深度学习方法相比,试验模型在准确率、精确率、召回率和 F_1 -Score 等指标上均表现出明显优势,可有效区分来自 5 个不同产地的大豆样品。试验中,模型的规模较大,在有限数据集下的训练效率有待提升,未来可以进一步探索高效和轻量级的跨模态特征融合框架,提升模型在有限样本下的训练稳定性并扩展到其他多模态任务中。

参考文献

[1] 张兴政,刘鑫,黄浩捷,等.大豆 ACO 家族基因鉴定、生物信息学分析及盐胁迫下表达模式分析[J].大豆科学,2025,44(3):

- 27-38.
- Zhang X Z, Liu X, Huang H J, et al. Identification, bioinformatics analysis and expression pattern analysis of ACO family gene in *Glycine max* L. under salt stress[J]. Soybean Science, 2025, 44(3): 27-38.
- [2] Huang Z M, Xin J X, Sun S S, et al. Rapid identification of adulteration in edible vegetable oils based on low-field nuclear magnetic resonance relaxation fingerprints[J]. Foods, 2021, 10(12): 3 068.
- [3] 雷渊雄, 夏阿林, 黄炜, 等. 基于近红外光谱结合化学计量学的转基因大豆产地判别[J]. 食品与发酵工业, 2022, 48(12): 275-280.
- Lei Y X, Xia A L, Huang W, et al. Origin discrimination of transgenic soybean based on near infrared spectroscopy and chemometrics[J]. Food and Fermentation Industries, 2022, 48(12): 275-280.
- [4] 苗辰. 超高效液相色谱法测量大豆中大豆异黄酮含量[J]. 食品安全导刊, 2025(15): 55-57.
- Miao C. Determination of soybean isoflavones in soybeans by ultra high performance liquid chromatography[J]. China Food Safety Magazine, 2025(15): 55-57.
- [5] 刘川正, 马景余, 白雪瑞, 等. 基于电子舌和电子鼻结合 CNN-Transformer 模型的绿茶种类识别[J]. 食品与机械, 2024, 40(6): 34-42, 52.
- Liu C Z, Ma J Y, Bai X R, et al. Green tea species recognition based on electronic tongue and electronic nose combined with CNN-Transformer model[J]. Food & Machinery, 2024, 40(6): 34-42, 52.
- [6] 马泽亮, 刘雅倩, 程琦峰, 等. 电子鼻和电子舌结合 LSTM-AM-M1DCNN 检测枸杞产地[J]. 食品与机械, 2024, 40(12): 51-58.
- Ma Z L, Liu Y Q, Cheng Q F, et al. Detection of the origin of wolfberry based on electronic nose and electronic tongue combined with LSTM-AM-M1DCNN[J]. Food & Machinery, 2024, 40(12): 51-58.
- [7] Cetó X, Del Valle M. Electronic tongue applications for wastewater and soil analysis[J]. iScience, 2022, 25(5): 104304.
- [8] Wang Y J, Liu Y, Cui Q Q, et al. Monitoring the withering condition of leaves during black tea processing via the fusion of electronic eye (E-eye), colorimetric sensing array (CSA), and micro-near-infrared spectroscopy (NIRS) [J]. Journal of Food Engineering, 2021, 300: 110534.
- [9] 查天, 张承吉, 朱磊, 等. 基于电子眼和电子鼻的不同包装片烟醇化品质差异分析[J]. 烟草科技, 2024, 57(12): 15-24.
- Zha T, Zhang C J, Zhu L, et al. Differences in aging quality of tobacco strips packaged by different methods based on electronic eyes and electronic noses[J]. Tobacco Science & Technology, 2024, 57(12): 15-24.
- [10] Orlandi G, Calvinini R, Foca G, et al. Data fusion of electronic eye and electronic tongue signals to monitor grape ripening[J]. Talanta, 2019, 195: 181-189.
- [11] Zhou Y F, Zhang Z L, He Y, et al. Integration of electronic nose, electronic tongue, and colorimeter in combination with chemometrics for monitoring the fermentation process of *Tremella fuciformis*[J]. Talanta, 2024, 274: 126006.
- [12] Li Y H, Daho M E H, Conze P H, et al. A review of deep learning-based information fusion techniques for multimodal medical image classification[J]. Computers in Biology and Medicine, 2024, 177: 108635.
- [13] 金鑫宁, 刘铭, 桑恒亮, 等. 基于电子舌和电子眼结合改进 MobileNetv3 的黄芪快速溯源检测[J]. 食品与机械, 2023, 39(6): 37-47.
- Jin X N, Liu M, Sang H L, et al. Fast traceability detection of *Astragalus membranaceus* based on the combination of electronic tongue and electronic eye to improve MobileNetv3 [J]. Food & Machinery, 2023, 39(6): 37-47.
- [14] Yang Z W, Gao J Y, Wang S C, et al. Synergetic application of E-tongue and E-eye based on deep learning to discrimination of Pu-erh tea storage time[J]. Computers and Electronics in Agriculture, 2021, 187: 106297.
- [15] Wang S C, Zhang Q, Liu C Z, et al. Synergetic application of an E-tongue, E-nose and E-eye combined with CNN models and an attention mechanism to detect the origin of black pepper [J]. Sensors and Actuators A: Physical, 2023, 357: 114417.
- [16] Peng Z L, Huang W, Gu S Z, et al. Conformer: local features coupling global representations for visual recognition[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE, 2021: 367-376.
- [17] 蒋雷, 汤海林, 陈瑜瑾. 基于 Transformer 模型的自然语言处理研究综述[J]. 现代计算机, 2024, 30(14): 31-35.
- Jiang L, Tang H L, Chen Y J. A review of natural language processing research based on Transformer model[J]. Modern Computer, 2024, 30(14): 31-35.
- [18] 赵小芬, 彭朋. 基于多模态视听融合的 Transformer 语音识别算法研究[J]. 传感器与微系统, 2025, 44(2): 48-52.
- Zhao X F, Peng P. Research on Transformer speech recognition algorithm based on multimodal audio-visual fusion[J]. Transducer and Microsystem Technologies, 2025, 44(2): 48-52.
- [19] Zhou H Y, Zhang S H, Peng J Q, et al. Informer: beyond efficient transformer for long sequence time-series forecasting [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(12): 11 106-11 115.
- [20] Liu Z, Lin Y T, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE, 2021: 10 012-10 022.
- [21] Rao X, Xiao S Y, Li J X, et al. LSBO-NAS: latent space Bayesian optimization for neural architecture search[C]// 2022 4th International Conference on Control and Robotics (ICCR). Guangzhou: IEEE, 2023: 22-27.