

面向采后赤霞珠葡萄质量分类的传感阵列设计和有效性分析

刘静¹, 窦全胜^{1,2}, 陈炳希², 宁宇宸², 魏广芬²

(1. 喀什大学, 新疆 喀什 844008; 2. 山东工商学院, 山东 烟台 264005)

摘要: 赤霞珠是全球种植最广泛的酿酒葡萄品种之一, 利用电子鼻系统对采后赤霞珠葡萄质量进行有效监测具有重要的现实意义。传感阵列是电子鼻系统的核心, 决定了电子鼻系统的气味识别能力。为此, 设计了一款由8个传感器组成的传感阵列, 用于采集赤霞珠葡萄的挥发性有机化合物(VOC)数据; 同时, 提出了CDI法则和基于多粒度聚类的核心子集发现算法(ESDA-MC), 用于验证传感阵列的有效性。实验结果表明, 在不同聚类粒度下, ESDA-MC提取的核心子集对原始时序数据的覆盖率达73.1%~97.1%, 轮廓系数大于0.5, 验证了ESDA-MC输出的核心子集具有良好的时序连续性与簇间可区分性。特别地, 当聚类数量 k 为3~5时, 核心子集间的轮廓系数达0.74以上、综合指数超过0.84, 输出模式与感官分类一致, 满足CDI法则, 验证了所设计的传感阵列能够有效识别赤霞珠葡萄的VOC模式变化, 为采后赤霞珠葡萄的质量监测提供了技术手段, 也为同类应用中传感阵列的设计和有效性分析提供了一个可以借鉴的解决方案。

关键词: 电子鼻; 传感阵列; 多粒度聚类; 核心子集

中图分类号: TP393

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2026001

Design and effectiveness analysis of a sensor array for post-harvest quality classification of Cabernet Sauvignon grapes

Liu Jing¹, Dou Quansheng^{1,2}, Chen Bingxi², Ning Yuchen², Wei Guangfen²

1. Kashi University, Kashi 844008, China

2. Shandong Technology and Business University, Yantai 264005, China

Abstract: Cabernet Sauvignon is one of the most widely planted wine grape varieties globally, and the effective monitoring of post-harvest Cabernet Sauvignon grape quality using an electronic nose system holds significant practical importance. The sensor array is the core of the electronic nose system, determining the olfactory recognition capability of the electronic nose system. A sensor array composed of eight sensors was designed for collecting data on VOC released during the quality evolution process of Cabernet Sauvignon grapes. A CDI criterion and a essential subset discovery algorithm based on multi-granularity clustering (ESDA-MC) were proposed to validate the effectiveness of the sensor array. Experimental results demonstrated that under different clustering granularities, coverage rates of 73.1% to 97.1% were achieved by the essential subsets extracted by ESDA-MC for the original time-series data, with all silhouette coefficients exceeding 0.5, indicating satisfactory temporal continuity and inter-cluster separation.

收稿日期: 2025-08-18; 修回日期: 2025-11-05

基金项目: 国家自然科学基金(No.61976125); 新疆维吾尔自治区自然科学基金(No.2022D01A237); 新疆维吾尔自治区本科教育教学研究和改革项目(No.XJGXJGPTA-2024011)

Foundation Items: The National Natural Science Foundation of China (No.61976125), Natural Science Foundation of Xinjiang Uygur Autonomous Region (No.2022D01A237), Undergraduate Education Teaching Research and Reform Projects of Xinjiang Uygur Autonomous Region (No.XJGXJGPTA-2024011)

rability. Particularly when the number of clusters was set between 3 and 5, silhouette coefficients were maintained above 0.74, the comprehensive index exceeded 0.84, and the output patterns were found to be consistent with sensory classification, fulfilling the CDI criterion. This verifies that the designed sensor array can effectively identify changes in the VOC patterns of Cabernet Sauvignon grapes, providing a technical means for post-harvest quality monitoring and offering a referable solution for the design and effectiveness validation of sensor arrays in similar applications.

Keywords: electronic nose, sensor array, multi-granularity clustering, essential subset

0 引言

赤霞珠葡萄是全球广泛种植的葡萄品种，因其独特的风味特征成为酿造高品质葡萄酒的核心原料。不同质量状态下的赤霞珠葡萄会释放出不同的挥发性有机化合物（volatile organic compound, VOC），这些 VOC 与赤霞珠葡萄的成熟度、储存时间和腐烂程度密切相关^[1-3]。利用电子鼻系统采集 VOC 数据，并借助智能分析方法，实现对赤霞珠葡萄质量的高效、无损检测，对推进智慧农业具有重要意义，已成为该领域的研究热点之一。

电子鼻系统是一种基于生物嗅觉原理设计的气味检测设备，其核心组件为传感阵列。该阵列能够捕捉目标气体中的化学信息，并将其转化为可测量的电信号，从而实现对特定气体的检测。

由于不同应用场景中目标气体的种类、浓度和环境条件各异，传感阵列的设计与有效验证直接决定了电子鼻系统对目标气体的检测性能，是构建高性能电子鼻系统的关键环节。然而，传感阵列的有效性验证面临诸多挑战：其一，传感器普遍存在的交叉敏感性会导致响应信号重叠，难以准确区分目标气体与干扰气体；其二，气体浓度变化和环境温湿度波动等因素会显著影响传感器响应的稳定性与重复性，使得对阵列性能的客观评估变得复杂；其三，传统评估方法多侧重于单一性能指标，缺乏从数据层面系统验证阵列是

否真正捕捉到目标气体本质特征的综合判据，进一步增加了验证工作的难度。

针对上述问题，本文设计了一款由 8 个传感器组成的传感阵列，利用该阵列可以持续采集赤霞珠葡萄释放的 VOC 数据，用于构建时间序列数据集。为验证传感阵列的有效性，本文提出一种基于 CDI 法则的聚类质量评估验证方法；在此基础上，开发了基于多粒度聚类的核心子集发现算法（essential subset discovery algorithm based on multi-granularity clustering, ESDA-MC）。实验结果表明，本文设计的传感阵列能够有效采集蕴含不同质量状态赤霞珠葡萄特征信息的 VOC 数据。该研究不仅为赤霞珠葡萄的高效无损检测提供了一种可行的技术路径，也为相似场景下电子鼻传感阵列的有效性验证提供了新的思路。

1 相关研究

水果在成熟或腐烂的过程中会释放出特定的 VOC。为捕获此类信息，研究者已针对不同应用场景设计了多种传感阵列，用于采集相应场景下的目标 VOC 数据。文献[4]对具有代表性的应用场景及传感阵列设计方案进行了系统梳理，具体内容汇总于表 1。

从表 1 可以看出，针对苹果品质分级问题，Zou 等^[5]利用 MQ 系列传感器检测苹果释放的醇类和碳氢类 VOC。为判别海棠果是否为药物催

表 1 代表性挥发性有机化合物传感器阵列设计及其应用场景

文献	应用场景	传感阵列	目标 VOC
[5]	苹果品质分级	MQ2, MQ3, MQ6, MQ8, MQ9, MQ135	二氧化碳、醇类、碳氢化合物、脂肪胺
[6]	海棠果催熟判别	TGS2600, TGS2602, TGS2603, TGS2610, TGS2611, TGS2612, TGS2620, GSBT11, WSP2110, MS1100, MP135, MP901	醛类、酮类、酯类
[7]	香蕉成熟度监测	TGS812, TGS822, TGS826, TGS2600, TGS2602, TGS2611, MICS5524, MQ2, MQ4, MQ5, MQ136	萜烯类、硫化物、羧酸类
[8-9]	桃子成熟过程监测	MQ2, MQ3, MQ4, MQ5, MQ6, MQ7, MQ8, MQ9, MQ135, TGS822, TGS2600, TGS2602, TGS2603	挥发性醛类、单萜类、内酯类

熟, Qiao等^[6]设计了一款由12个传感器组成的传感阵列, 实现对海棠果催熟过程中醛类、酮类和酯类物质的识别。Hendrick等^[7]设计的传感阵列包含11个传感器, 用于监测香蕉成熟过程中释放的萜烯类、硫化物和羧酸类VOC。针对桃子成熟度预测问题, Voss等^[8-9]采用由13个传感器组成的阵列, 对桃子成熟过程中产生的挥发性醛类、单萜类和内酯类化合物进行检测。在传感器有效性验证方面, 上述研究大多采用气相色谱-质谱等VOC成分分析技术, 这类方法存在操作流程烦琐、单次检测成本高、检测周期长等问题, 难以满足传感阵列在设计及优化阶段所需的多次实验迭代需求。

赤霞珠葡萄在质量变化过程中同样会释放醇类、醛类、酯类、萜烯类等VOC, 其成分特征与表1中各应用场景产生的VOC存在相似性。因此, 这些设计方案为本文构建面向赤霞珠葡萄的VOC检测传感阵列提供了重要的参考。

2 本文方法

2.1 数据采集

为实现赤霞珠葡萄VOC数据的稳定、精准采集, 本文搭建了一套数据采集装置, 该装置由温湿度控制箱、电子鼻系统和上位机3部分构成^[4,10], 其整体架构如图1所示。

在图1所示的装置中, 温湿度控制箱用于存放赤霞珠葡萄样品, 其通过进气管和排气管与电子鼻系统连接。在进气管和排气管上分别安装了三向阀, 用于控制气体流向, 实现电子鼻系统与温湿度控制箱、电子鼻系统与外部环境之间的气体交换。电子鼻系统的核心组件是一个经过优化设计的传感阵列, 传感阵列位于电子鼻系统的气仓内, 当其暴露于目标VOC时, 其电阻或电导率会发生变化, 这种变化被转换为电信号并被上传、存储在上位机中, 以供后续分析。

本文将电子鼻系统的采样频率设为0.1 Hz, 即每10秒记录一组传感响应数据。每小时进行4次采样, 单个传感器在一次完整的采样过程中可以记录78个数据点。若传感阵列包含 N 个传感器, 则电子鼻系统单次采样可以获得一个维度为 $78 \times N$ 的传感数据矩阵。

定义1 传感阵列响应状态

设系统最大采样次数为 M , D_i 为传感阵列的第 i 次采样, 本文称 D_i 为传感阵列在时刻 i ($i = 1, \dots, M$) 的响应状态, 简称状态 i 。

显然, 传感阵列的响应状态序列 $\{D_1, \dots, D_M\}$ 构成了一个与赤霞珠葡萄质量变化相对应的时间序列。

2.2 传感阵列设计

赤霞珠葡萄在质量变化过程中释放的VOC

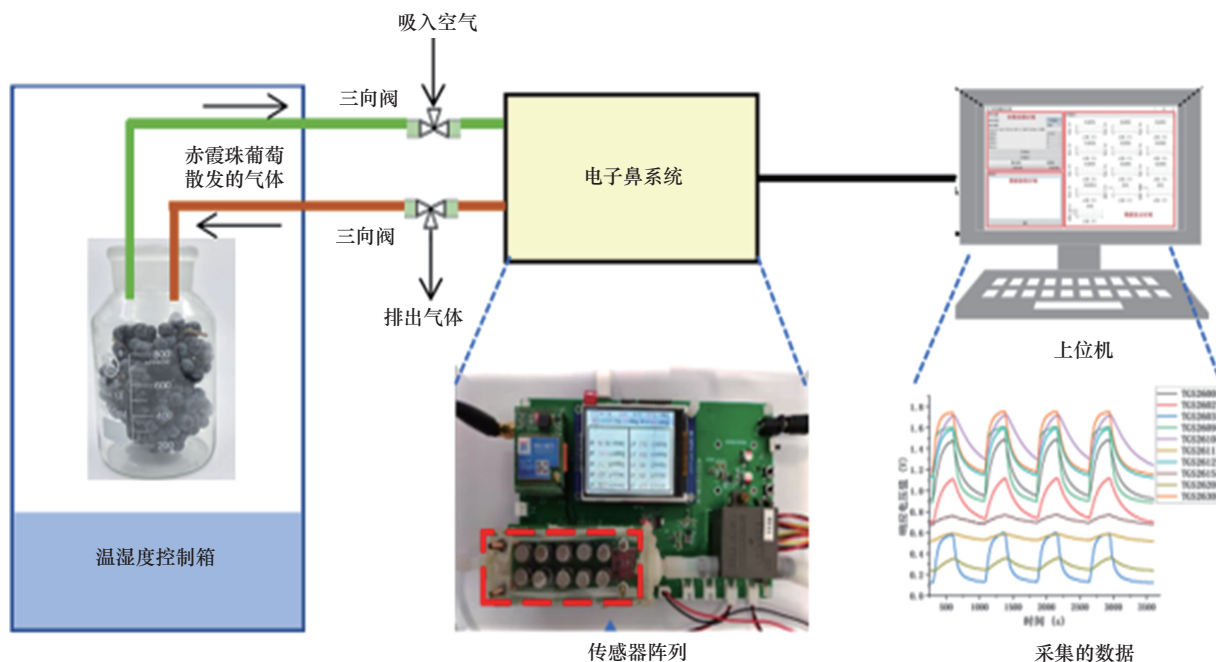


图1 赤霞珠葡萄VOC数据采集装置的整体架构

种类和浓度会持续发生变化。因此,传感阵列的设计应遵循以下原则。

1) 传感阵列中的传感器应对多种 VOC 具有良好的敏感性,能够有效捕捉赤霞珠葡萄在不同质量阶段可能释放的 VOC。同时,传感阵列应具备足够的灵敏度,以适应腐烂过程中 VOC 浓度的动态变化。

2) 针对赤霞珠葡萄腐烂过程中产生的主要 VOC,应选用对相同目标气体敏感但灵敏度和检测范围互补的传感器进行冗余设计,确保全面捕捉赤霞珠葡萄在不同状态下的特征信息。

3) 在传感阵列中适当配置一些非敏感传感器,这些传感器对赤霞珠葡萄产生的 VOC 响应弱,其采集的信号可作为参考基准,用于识别和校正其他传感器的响应,从而提高后续数据处理的准确性。

基于上述原则,结合赤霞珠葡萄 VOC 相关研究^[11-12],本文设计了一款由 8 个气体传感器构成的传感阵列^[4],具体配置见表 2。

在表 2 中,TGS2600、TGS2609 和 TGS2620 传感器用于检测以乙醇为主的醇类挥发物。这 3 个传感器检测的目标 VOC 相似,都是对醇类挥发物具有较高响应。同时,这 3 个传感器具有不同的检测范围和灵敏度,可在不同浓度水平上实现对醇类挥发物的全面覆盖。

TGS2603 传感器对含硫和胺类气体敏感,这类气体通常是赤霞珠葡萄腐烂的典型副产物^[13-14]。TGS2610 传感器对丙烷、丁烷等烷烃敏感,用于检测赤霞珠葡萄在缺氧条件下产生的烷烃类挥发物。在赤霞珠葡萄腐化后期,水分胁迫可能导致甲烷、硫化氢等气体的释放^[16],传感阵列中的

TGS2611 和 WSP7110 传感器可以对这写挥发物进行检测。

TGS2602 传感器检测的主要目标气体是氨气,而氨气并非赤霞珠葡萄释放的常见 VOC,故该传感器采集的响应信号相对稳定。利用这一特点,将 TGS2602 传感器的响应信号作为参考基准,用于辅助校准传感阵列中其他传感器的数据信号。

2.3 传感阵列有效性判定 CDI 法则

利用图 1 所示装置持续采集了 240 h 的赤霞珠葡萄 VOC 数据样本。为避免给细粒度数据分析带来不必要的干扰,本文选择了前 208 h 的数据进行分析,这部分数据不但覆盖了赤霞珠葡萄从新鲜到腐烂的整个演变过程,而且各质量状态的数据分布大致均衡。同时,由于赤霞珠葡萄的质量演变进程缓慢,过于相似的样本会增加后续分析的复杂度,因此在每小时 4 次的采样中随机选择 1 个样本,构成传感阵列响应状态时间序列 $\{D_1, \dots, D_{208}\}$,在该序列中 D_i 为 78×8 矩阵, $i = 1, \dots, 208$ 。

不难看出,时间序列 $\{D_1, \dots, D_{208}\}$ 与赤霞珠葡萄的质量演变过程相对应。然而,在实际应用中,人们习惯将葡萄质量分成不同的状态(如新鲜、轻微腐烂和重度腐烂等)。显然,葡萄的某一特定阶段应与时间序列 $\{D_1, \dots, D_{208}\}$ 中的一段连续子序列相对应。在赤霞珠葡萄质量演变过程中,其释放的 VOC 的种类和浓度会呈现不同的特征。因此,若传感阵列能够有效捕捉这些特征,则一定存在某种测度和聚类算法,将 $\{D_1, \dots, D_{208}\}$ 划分为时间连续且互不相交的簇,每个簇对应赤霞珠葡萄的一个质量状态。

表 2 传感阵列的配置

序号	传感器型号	目标 VOC	检测范围/ppm	灵敏度
Sensor ₁	TGS2600	乙醇、氢气	$1 \times 10^{-6} \sim 10 \times 10^{-6}$	0.08 ~ 0.84
Sensor ₂	TGS2602	氨气	$1 \times 10^{-6} \sim 30 \times 10^{-6}$	0.3 ~ 0.6
Sensor ₃	TGS2603	胺类、含硫化合物	$1 \times 10^{-6} \sim 10 \times 10^{-6}$	0.15 ~ 0.5
Sensor ₄	TGS2609	乙醇、氢气	$1 \times 10^{-6} \sim 50 \times 10^{-6}$	0.1 ~ 0.5
Sensor ₅	TGS2610	丙烷、丁烷	$500 \times 10^{-6} \sim 10\,000 \times 10^{-6}$	0.56 ± 0.06
Sensor ₆	TGS2611	甲烷	$500 \times 10^{-6} \sim 10\,000 \times 10^{-6}$	0.6 ± 0.06
Sensor ₇	TGS2620	乙醇、酯类、酮类等有机溶剂	$100 \times 10^{-6} \sim 2\,500 \times 10^{-6}$	0.5 ~ 0.65
Sensor ₈	WSP7110	硫化氢、乙炔	$40 \times 10^{-6} \sim 4\,000 \times 10^{-6}$	0.55 ~ 0.75

依据上述分析, 本文提出一种基于聚类质量评估的传感阵列有效性验证法则, 简称CDI法则。

CDI法则: 在不引入时序信息的条件下, 若基于合适的相似性度量与聚类算法, 能够将序列 $\{D_1, \dots, D_{208}\}$ 划分为若干个可区分的簇, 并满足以下条件。

1) 连续性 (continuity, C): 各簇中的样本主体来源于原时间序列中的连续区间, 传感数据内部存在与时间演化相关的稳定模式。

2) 差异性 (differentiation, D): 不同簇类间距离显著, 数据中存在明显不同的模式特征。

3) 可解释性 (interpretability, I): 聚类所划分的状态区间与实际观测到的葡萄质量阶段 (如新鲜、轻微腐烂、重度腐烂) 在演变顺序和时间跨度上吻合, 聚类结果具备明确的物理意义。

若同时满足上述3个条件, 则可判定传感阵列能够有效感知采后赤霞珠葡萄VOC的模式变化, 对目标VOC具有较高的敏感性。

需要说明的是, CDI法则为验证传感阵列有效性的充分条件, 该法则建立在相似性度量与聚类算法选择恰当的前提下。若聚类结果未能满足CDI条件, 则不能得出传感阵列无效的结论。

2.4 基于多粒度聚类的核心子集发现算法

设 X 和 Y 为两个维度相同的矩阵, x_{ij} 和 y_{ij} 分别为 X 和 Y 的第 i 行第 j 列元素, X 和 Y 的Frobenius距离可以通过式(1)计算得到。

$$\|X - Y\|_F = \sqrt{\sum (x_{ij} - y_{ij})^2} \quad (1)$$

进一步, 设 C_i 和 C_j 为两个维度相同的矩阵集合, C_i 和 C_j 之间的距离 $\text{dist}(C_i, C_j)$ 定义如下。

$$\text{dist}(C_i, C_j) = \frac{1}{|C_i| \times |C_j|} \sum_{X \in C_i} \sum_{Y \in C_j} \|X - Y\|_F \quad (2)$$

定义2 核心子集和边缘子集

设 S 为时间序列 $\{D_1, \dots, D_{208}\}$ 的任意子集, 称 S 中所有下标连续且元素个数最多的子序列为 S 的核心连续子集, 简称核心子集, 记作 $\text{ES}(S)$ 。称 S 中所有核心子集之外的元素构成的集合为 S 的边缘子集, 记作 $\text{MS}(S)$ 。例如, $S = \{D_1, D_3, D_6, D_7, D_8, D_{10}, D_{11}, D_{14}\}$, 则 $\text{ES}(S) = \{D_6, D_7, D_8\}$, $\text{MS}(S) = \{D_1, D_3, D_{10}, D_{11}, D_{14}\}$

在此基础上, 本文提出了基于多粒度聚类的

核心子集发现算法 (essential subset discovery algorithm based on multi-granularity clustering, ESDA-MC), 如算法1所示。

算法1 基于多粒度聚类的核心子集发现算法

输入 时间序列 $\{D_1, \dots, D_{208}\}$

输出 cluster set C

- 1) initialize: $C_{\text{tmp}} = \{D_1, \dots, D_{208}\}$, $v, k=2$, $C = \Phi$
- 2) do
- 3) compute $\text{dist}(C_i, C_j)$, for all C_i, C_j in C
- 4) merge C_i, C_k into C_{wk} if $\text{dist}(C_i, C_k)$ is the minimum among all distances computed in step 3
- 5) until $|C_{\text{tmp}}| = k$
- 6) if (the proportion of the ESs $\rho \geq v$) then
- 7) $C = \text{getES}(C_{\text{tmp}})$, $k = k+1$, $C_{\text{tmp}} = \{D_1, \dots, D_{208}\}$
- 8) goto 2
- 9) return C

由算法1可以看出, ESDA-MC在层次聚类的框架内嵌入了时间连续性约束。首先, 算法通过层次聚类将数据逐步合并至 k 个簇, 得到聚簇集合 C_{tmp} 。然后, 设定阈值 v , 检验 C_{tmp} 中 k 个簇的核心子集占比 ρ 。若 $\rho > v$, 表明簇中核心子集的占比较高, 对时间序列 $\{D_1, \dots, D_{208}\}$ 的划分粒度过粗, 则可以进一步细化。此时, 利用 getES 函数, 将 C_{tmp} 中每个簇的ES保存在 C 中, 令 $k+1$ 后继续重复上述过程; 当 $\rho \leq v$ 时, 算法终止, 此时所得簇集 C 中的每个簇为单一连续时段, 对应目标VOC的一段连续响应区间。

在算法1中, 对于以下几点需要补充说明。

1) 函数 getLCS 的实现过程较为简单, 因此未在文中给出。

2) 理论上, 若每次聚类得到的聚簇均为时间序列的连续子集, 则 $\rho \geq v$ 始终成立, 算法将持续迭代直至 $k=208$, 此时 $C = \{\{D_1\}, \dots, \{D_{208}\}\}$ 。然而在实际数据中, 这种情况几乎不可能发生。

3) 聚簇中可能出现极少数样本独立成簇的现象, 这通常由数据采集中的瞬时噪声引起。鉴于其所对应的时间片段过短, 不具备实际的状态表征意义, 算法将其视作边缘集合。

上述情况的处理并不复杂, 为保持描述的简洁性, 本文未在算法1中对这些极端情况进行显

式的判别和处理。

3 ESDA-MC 实验结果分析

3.1 核心子集的时间连续性

设参数 $v=0.7$, 该阈值能够有效过滤由噪声引起的碎片化簇, 同时确保输出结果覆盖时间序列 $\{D_1, \dots, D_{208}\}$ 的主体部分, 避免因约束过强而将过渡状态“硬归类”。在时间序列 $\{D_1, \dots, D_{208}\}$ 上运行 ESDA-MC, 输出结果如见 3。

表 3 中, 第 1 列为聚簇数量 k , 第 2 列、第 3 列分别为 ESDA-MC 的输出结果和核心子集在整个时间序列 $\{D_1, \dots, D_{208}\}$ 中的占比。其中, ES_{ij} 表示当 $k=i$ 时, ESDA-MC 输出的第 j 个核心子集。

由表 3 可以得出以下结论。

1) 时间片段过短的簇被视为边缘集合, 因此聚簇数量 k 与 ESDA-MC 实际输出的核心子集个数并不一致。

2) 当 $k=2$ 时, 核心子集占比最高, 达 97.1%。随着 k 值的增大, 不同核心子集间出现更多“混淆”样本, 导致所有核心子集在时间序列中的占比逐渐下降。当 $k=8$ 时, 核心子集占比降至最低, 但仍然覆盖了时间序列 $\{D_1, \dots, D_{208}\}$ 的 73.1%。这说明即便在细粒度簇划分下, 所采集的 VOC 数据仍具有显著的时间连续性。

3) 当 $k=3, 4$ 和 5 时, ESDA-MC 均输出了 3 个核心子集, 但核心子集占比随 k 值的增大而

降低。这是因为聚类粒度细化 (k 值增大) 后, 算法将部分处于模式转换间的边缘样本从原有的核心子集中剥离出来, 每个核心子集中的样本数量减少。当 $k=6$ 时, ESDA-MC 输出了 $ES_{64} = \{D_{147}, \dots, D_{158}\}$, 该集合在 $k=5$ 时属于 ES_{52} 与 ES_{53} 之间的边缘集合子集, 未被独立输出。随着 k 继续增大至 7, ES_{61} 被分割为 ES_{71} 和 ES_{72} ; 同样, 当 k 增大至 8 时, ES_{74} 被分割为 ES_{84} 和 ES_{85} 。这一细分现象表明, 时间序列 $\{D_1, \dots, D_{208}\}$ 中蕴含的模式演变过程并非严格的阶跃式跳变, 在宏观状态内部仍存在连续、细微的模式变化。

上述结果表明, ESDA-MC 在不同 k 值下均能够有效提取具有明显时间连续性的核心子集, 这意味着传感阵列所采集的 VOC 数据内部蕴含了随时间演变的不同模式特征。这些模式在时间上与赤霞珠葡萄质量状态的转变一致, 若进一步验证其可区分性和可解释性, 则可充分证明本文设计的传感阵列是有效的。

3.2 核心子集的簇间差异性

对于任意 $D_i \in \{D_1, \dots, D_{208}\}$, 利用主成分分析 (principal component analysis, PCA) 将其从 78×8 的矩阵降维至 78×2 的矩阵, 降维后的矩阵保留了 D_i 约 87% 的方差^[4]。计算该矩阵两个列向量的数学期望, 进而将 D_i 投影至二维空间。ESDA-MC 输出结果在此二维空间中的分布如图 2 所示。

表 3 不同 k 值下 ESDA-MC 的输出结果和核心子集占比

聚簇数量 k	ESDA-MC 输出结果	核心子集占比
2	$ES_{21} = \{D_1, \dots, D_{160}\}, ES_{22} = \{D_{167}, \dots, D_{208}\}$	97.1%
3	$ES_{31} = \{D_1, \dots, D_{51}\}, ES_{32} = \{D_{60}, \dots, D_{150}\}, ES_{33} = \{D_{160}, \dots, D_{200}\}$	88.4%
4	$ES_{41} = \{D_1, \dots, D_{51}\}, ES_{42} = \{D_{60}, \dots, D_{150}\}, ES_{43} = \{D_{160}, \dots, D_{200}\}$	84.6%
5	$ES_{51} = \{D_1, \dots, D_{51}\}, ES_{52} = \{D_{60}, \dots, D_{143}\}, ES_{53} = \{D_{167}, \dots, D_{200}\}$	81.2%
6	$ES_{61} = \{D_1, \dots, D_{51}\}, ES_{62} = \{D_{68}, \dots, D_{88}\}, ES_{63} = \{D_{95}, \dots, D_{143}\},$ $ES_{64} = \{D_{147}, \dots, D_{158}\}, ES_{65} = \{D_{173}, \dots, D_{200}\}$	77.4%
7	$ES_{71} = \{D_1, \dots, D_{33}\}, ES_{72} = \{D_{34}, \dots, D_{51}\}, ES_{73} = \{D_{68}, \dots, D_{88}\},$ $ES_{74} = \{D_{95}, \dots, D_{143}\}, ES_{75} = \{D_{147}, \dots, D_{158}\}, ES_{76} = \{D_{173}, \dots, D_{200}\}$	77.4%
8	$ES_{81} = \{D_1, \dots, D_{33}\}, ES_{82} = \{D_{34}, \dots, D_{51}\}, ES_{83} = \{D_{68}, \dots, D_{88}\},$ $ES_{84} = \{D_{95}, \dots, D_{107}\}, ES_{85} = \{D_{116}, \dots, D_{143}\},$ $ES_{86} = \{D_{147}, \dots, D_{158}\}, ES_{87} = \{D_{173}, \dots, D_{200}\}$	73.1%

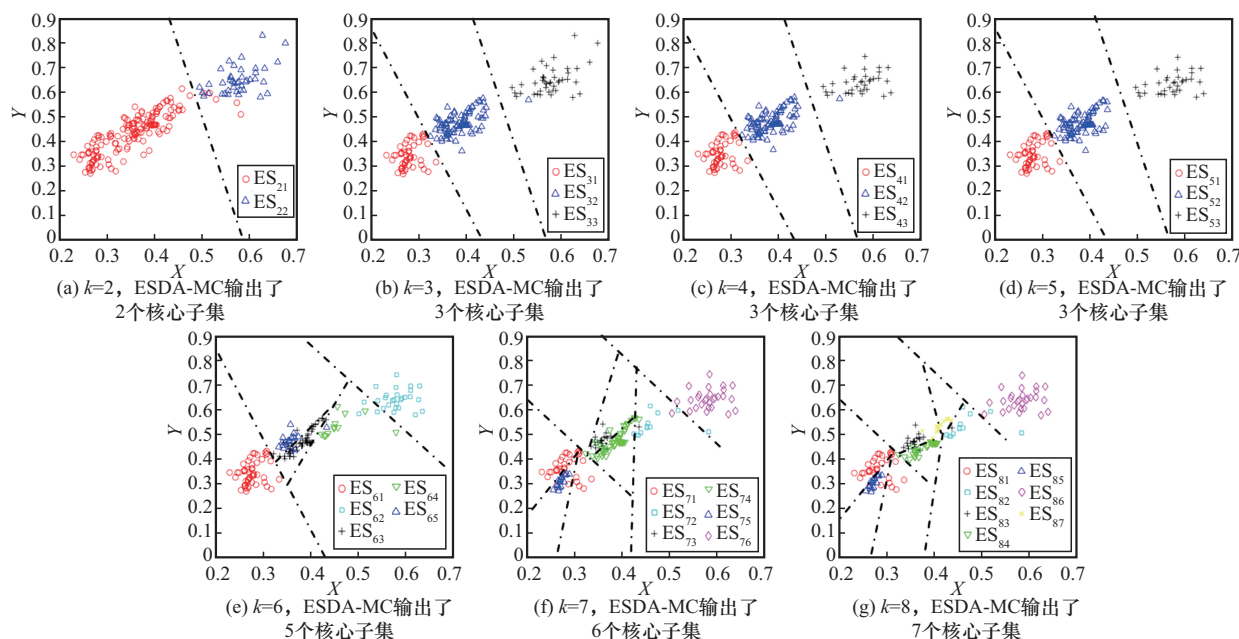


图2 ESDA-MC输出结果在二维空间中的分布

由图2(a)可以看出,当 $k=2$ 时,ESDA-MC输出2个核心子集(ES_{21} 和 ES_{22}),尽管这两个子集中有少量样本重叠,但它们的主体部分在二维空间中仍然保持较好的可区分性,显示出明显的簇间差异性。

观察图2(b)~图2(d)可以看出,当 k 值分别为3、4和5时,ESDA-MC均输出了3个核心子集,且这些核心子集在二维空间中表现出较好的可区分性。特别地,当 $k=5$ 时,由于算法过滤了时间片段过短的簇与边缘样本,子集之间的可区分性更加显著。在此情况下,即使采用如K-近邻这类简单的分类器,也可实现对子集元素的有效识别。

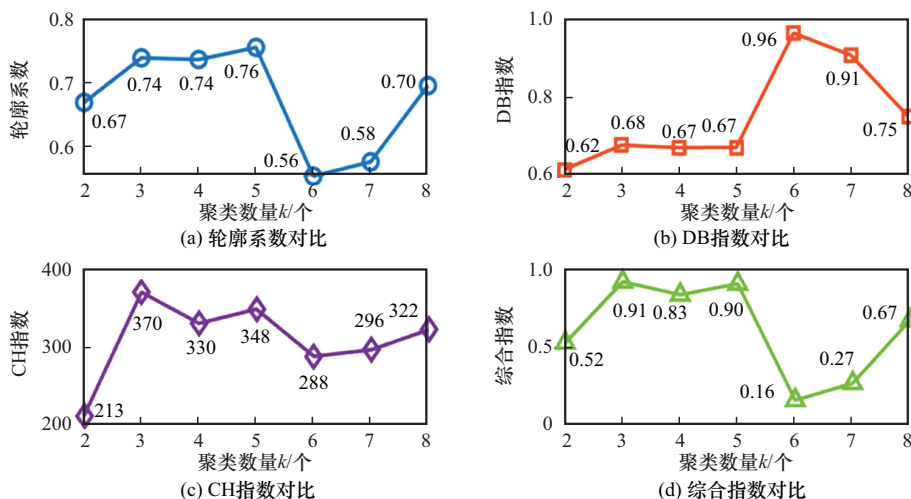
观察图2(e)~图2(g)可以看出,当 k 为6、7和8时,ESDA-MC分别输出了5、6和7个核心子集,这些核心子集在二维空间中呈现可辨识的团簇形态,表明其具有独立的分布特征,理论上具备可区分性。然而,此时各核心子集对应的簇团相互交错,边界区域存在部分重叠,导致核心子集间分类难度增大。在这种情况下,利用K-近邻这类简单的分类器难以实现对核心子集元素的有效判定,需要引入更复杂的分类模型(如基于深度学习或集成学习的模型)才能够实现不同核心子集间的准确区分。

为定量评估ESDA-MC所得核心子集的模式可区分性,本文采用轮廓系数、Calinski-Harabasz

(CH)指数、Davies-Bouldin(DB)指数及综合指数4项内部指标进行量化分析。不同 k 值下ESDA-MC输出的核心子集的各项指标对比如图3所示。

图3(a)为不同 k 值下ESDA-MC输出的核心子集的轮廓系数对比曲线。其中,轮廓系数的取值范围为 $[-1,1]$,其值越接近1,表明核心子集内样本越紧密、子集间可区分性越好。通常认为该值大于0.5时,即表明核心子集间具有较高的可区分性。若轮廓系数大于0.7,则表明不同核心子集之间模式差异显著、可区分性好^[17]。从图3(a)可以看出,所有 k 值对应的轮廓系数均大于0.5,说明核心子集在统计意义上可区分。特别地,当 k 为3、4、5时,轮廓系数分别为0.74、0.74和0.76,表明核心子集之间边界清晰,可区分性显著。

图3(b)为不同 k 值下ESDA-MC输出的核心子集的DB指数对比曲线。DB指数为各核心子集内样本到其质心的平均距离与子集间距离的比值,通过它同样可以量化核心子集间的分离程度。DB指数取值下限为0,其值越小,表明不同核心子集间的分离效果越好。从图中可以看出,当 k 为2、3、4、5时,DB指数分别为0.62、0.68、0.67和0.67,数值相对稳定且保持在较低水平,说明在此区间内ESDA-MC所提取的核心

图3 不同 k 值下ESDA-MC输出的核心子集的各项指标对比

子集保持了较高的分离度。当 $k=6、7、8$ 时，DB指数显著上升，表明此时核心子集之间的边界模糊、可区分性下降。

图3(c)为不同 k 值下ESDA-MC输出的核心子集的CH指数对比曲线。CH指数为核心子集间与核心子集内散度之比。该指数无上限，其值越大，表明核心子集内部结构越紧密且子集间分离度越高。从图中可以看出，当 $k=2$ 时，CH指数为213，处于最低水平。这是因为聚类数量较少时，每个核心子集涵盖的样本范围较广，子集内部离散度增大，簇间与簇内方差之比降低。当 $k=3$ 时，CH指数达到最大值（370），说明此时核心子集内部具有极高的紧凑性，且子集间分离度达到最优。

轮廓系数、CH指数与DB指数分别从簇内一簇间距离关系、方差比和相似度等维度评估核心子集间的可区分性。然而，单一指标难以对核心子集的簇间差异性做出全面客观评定。为此，本文引入综合指数，通过对上述指标进行归一化与加权融合，对核心子集的簇间差异性进行整体评价。图3(d)展示了不同 k 值下ESDA-MC输出的核心子集的综合指数对比曲线，从图中可以看出，当 k 为3、4和5时，ESDA-MC均输出3个核心子集，且综合指数显著高于其他 k 值情况。其中，当 $k=3$ 时，综合指数最高（0.91），表明此时核心子集具有最优的整体可分性。当 $k=4$ 时，由于在 ES_{43} 中删除了边缘集合 $\{D_{201}, \dots, D_{208}\}$ ，虽然提升了 ES_{43} 的内聚性，但也导致 ES_{42} 与 ES_{43} 的簇心距离减小，CH指数下降，综合指数降至

0.83。当 $k=5$ 时， ES_{52} 中删除边缘集合 $\{D_{144}, \dots, D_{150}\}$ ，其内聚性提升与分离性损失大致相当，故而综合指数保持在0.90。当 k 为3、4和5时，综合指数均处于较高水平，表明此时ESDA-MC输出的核心子集间具有显著的差异性，模式可区分性强。

基于本节分析可知，ESDA-MC输出结果在统计意义上具备显著的可区分性。尤其是当 $k=3、4、5$ 时，所得核心子集均表现出较好的可区分性。

参数 k 与VOC的模式划分粒度相关。在需要对葡萄质量进行细致管理的场景中，可以考虑增大 k 值以获得更细致的聚类结果。然而， k 值的增大会使“混淆”状态的数量也相应增加，造成核心子集占比下降；同时，由于在原有核心子集上进行了细分，核心子集间相互交错，这会给后续的模式识别算法带来挑战。

本节实验结果结合第3.1节中关于核心子集时间连续性的分析，可以充分说明传感数据中蕴含了连续演变且彼此可区分的稳定模式，该特性与采后赤霞珠葡萄质量连续变化的物理过程一致。接下来探讨核心子集所对应模式的物理可解释性。

3.3 核心子集的可解释性

在现实场景中，根据人的感官评估，通常将赤霞珠葡萄分为新鲜（fresh state, FS）、轻微腐烂（mild decay, MD）和严重腐烂（severe decay, SD）3种状态。新鲜状态的赤霞珠葡萄果实坚实，表皮光滑，颜色鲜艳，并散发出果香。轻微腐烂的赤霞珠葡萄颜色变暗，表皮开始变软，局

部出现小斑点或损伤。当赤霞珠葡萄严重腐烂时，葡萄萎缩，颜色暗淡，表皮多处破裂，出现大量霉菌，并散发出强烈、难闻的气味。

与实验过程同步，3名实验人员对赤霞珠葡萄的质量状态进行了评估。当评估结果出现矛盾时，通过投票确定葡萄的最终状态。感官评估结果与时间序列的对应关系如下：新鲜对应的区间为 $\{D_1, \dots, D_{67}\}$ 、轻微腐烂对应的区间为 $\{D_{68}, \dots, D_{172}\}$ 、严重腐烂对应的区间为 $\{D_{173}, \dots, D_{208}\}$ 。图4展示了ESDA-MC输出结果与感官评估结果的对比情况。

设 S 为时间序列 $\{D_1, \dots, D_{208}\}$ 的任意子集，记 $\min_idx(S)$ 和 $\max_idx(S)$ 分别为 S 中样本的最小索引和最大索引。从图4的分布可以看出，当 $k=2$ 时，ESDA-MC输出2个核心子集（ ES_{21} 和 ES_{22} ）。由于聚类粒度较大，本应属于轻微腐烂状态的样本被分散划分至 ES_{21} 和 ES_{22} 中。 ES_{21} 和 ES_{22} 对应的模式不能与现实场景中赤霞珠葡萄的质量状态相对应。因此，此类粗粒度划分仅适用于对葡萄质量状态区分要求较低的场景（如仅需将葡萄分为新鲜与腐烂两大类的情形）。

当 k 为3、4和5时，ESDA-MC均输出3个核心子集。随着聚类粒度逐渐细化，部分边缘样本被剔除，但核心子集的整体结构基本保持一致。以 $k=3$ 为例，ESDA-MC输出的3个核心子集 ES_{31} 、 ES_{32} 和 ES_{33} 在时间区间上与感官评估结果

高度重叠，有效对应感官所划分的质量状态。值得注意的是， $\min_idx(ES_{32}) < \min_idx(MD)$ 且 $\min_idx(ES_{32}) < \min_idx(SD)$ ，即感官判定为轻微腐烂和严重腐烂的起始时间均晚于相应核心子集的起始索引。这一现象表明，本文设计的传感阵列对赤霞珠葡萄VOC模式变化的灵敏性高于人类感知，能更早地感知赤霞珠葡萄质量状态的变化。

当 k 为6、7和8时，MD和SD对应的核心子集逐步被细分，ESDA-MC分别输出5、6和7个核心子集。此时核心子集间重叠度增加、整体可区分性下降。细分后得到的核心子集对应了赤霞珠葡萄腐烂过程中更细微的VOC动态变化特征，对状态监测要求更精细的应用场景具有意义。

通过本节分析可以得出：当 $k=2$ 时，ESDA-MC输出了2个核心子集，适用于对赤霞珠葡萄质量状态区分要求较低的场景。当 k 为3、4和5时，ESDA-MC均输出了3个核心子集，核心子集对应的时间区间与依据感官划分的新鲜、轻微腐烂和严重腐烂状态高度一致，且传感阵列对VOC模式变化的灵敏性优于感官评估。当 k 为6、7和8时，ESDA-MC分别输出了5、6和7个核心子集，此时核心子集对应的模式虽然可以揭示更精细的VOC动态变化，但核心子集间重叠度增加、可区分性下降，需要借助更复杂的分类模型才能实现核心子集的有效区分。

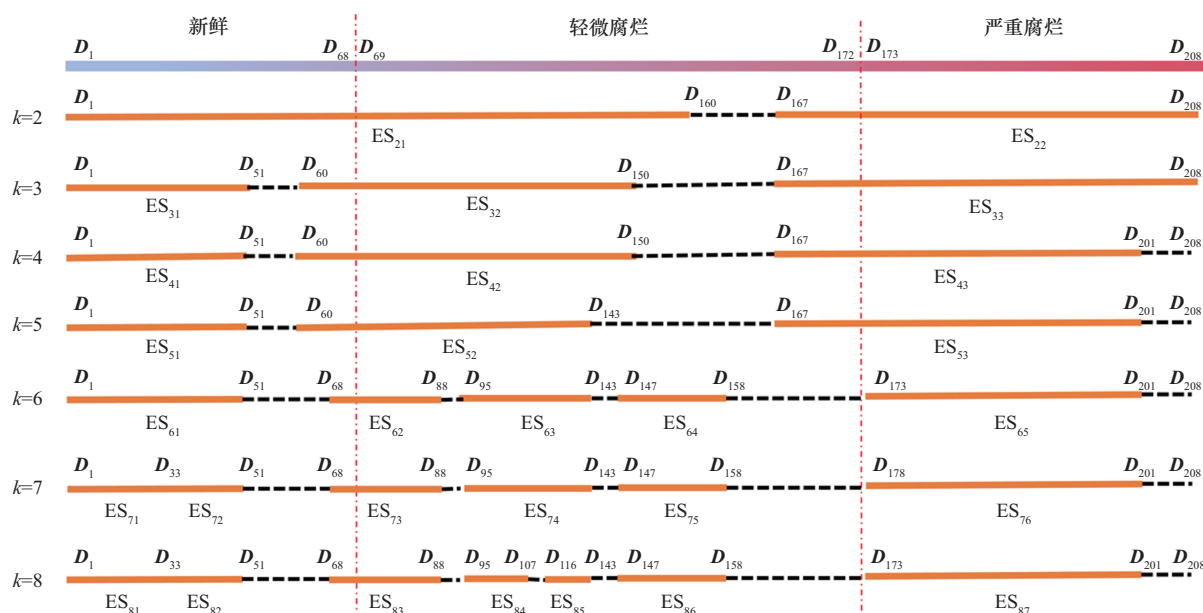


图4 ESDA-MC输出结果与感官评估结果对比

上述结果表明, ESDA-MC 提取的核心子集展现出明确的物理可解释性, 能够有效解释赤霞珠葡萄质量演变的实际过程。

4 结束语

本文设计了一款由 8 个传感器组成的传感阵列, 用于采集赤霞珠葡萄采后过程中释放的 VOC 数据; 同时, 提出了结合连续性、差异性和可解释性的评估法则, 以及基于多粒度聚类核心子集发现算法, 用于验证传感阵列的有效性。

实验结果表明, ESDA-MC 在不同聚类粒度下 ($k=2, \dots, 8$), 输出的核心子集对原始数据具有较高的覆盖率 (73.1%~97.1%), 证明传感数据内部蕴含具有时序连续性的稳定模式。

利用轮廓系数、DB 指数、CH 指数和综合指数对 ESDA-MC 输出结果进行分析, 结果显示, ESDA-MC 输出的核心子集间的轮廓系数大于 0.5。特别地, 当 k 为 3、4 和 5 时, 核心子集间的轮廓系数不小于 0.74、综合指数不低于 0.83, 证明核心子集间具有良好的可区分性。

将 ESDA-MC 的输出结果与人的感官评估结果对比发现, 当 k 为 3、4 和 5 时, 核心子集对应的时间区间与赤霞珠葡萄的新鲜、轻微腐烂、严重腐烂 3 种状态高度吻合, 且 ESDA-MC 识别的状态转换时间早于感官判断, 这不仅验证了 ESDA-MC 输出的核心子集具有明确的可解释性, 还证明了传感阵列对赤霞珠葡萄 VOC 的变化更加灵敏, 对其质量变化具有更早的预警能力。

综上所述, 本文设计的传感阵列所采集的数据满足 CDI 法则中连续性、差异性与可解释性要求, 能够有效表征赤霞珠葡萄 VOC 的动态变化, 为赤霞珠葡萄采后质量监测提供依据; 同时, 本文提出的 CDI 法则为类似问题中传感阵列的有效性验证提供了可借鉴的解决方案。

未来的研究工作将扩大数据采集范围, 探索更加高效的模式识别算法, 进一步验证本文设计的传感阵列在更复杂场景下的适用性。

参考文献:

[1] Tao Y S, Li H, Wang H, et al. Volatile compounds of young Cabernet

Sauvignon red wine from Changli County (China) [J]. *Journal of Food Composition and Analysis*, 2008, 21(8): 689-694.

[2] Wei R T, Ding Y T, Chen N, et al. Diversity and dynamics of microbial communities during spontaneous fermentation of Cabernet Sauvignon (*Vitis vinifera* L.) from different regions of China and their relationship with the volatile components in the wine[J]. *Food Research International*, 2022, 156: 111372.

[3] Ntuli R G, Saltman Y, Ponangi R, et al. Impact of fermentation temperature and grape solids content on the chemical composition and sensory profiles of Cabernet Sauvignon wines made from flash détente treated must fermented off-skins[J]. *Food Chemistry*, 2022, 369: 130861.

[4] Dou Q S, Ning Y C, Chen B X, et al. Customized e-nose sensor array configuration and VOC pattern analysis for Cabernet Sauvignon grape post-harvest quality monitoring[J]. *Journal of Agriculture and Food Research*, 2025, 22: 102029.

[5] Zou X G, Wang C Y, Luo M M, et al. Design of electronic nose detection system for apple quality grading based on computational fluid dynamics simulation and K-nearest neighbor support vector machine[J]. *Sensors*, 2022, 22(8):2997.

[6] Qiao J L, Su G Q, Liu C, et al. Study on the application of electronic nose technology in the detection of the artificial ripening of crab apples[J]. *Horticulturae*, 2022, 8(5): 1-13.

[7] Hendrick, Efrizon, Yultrina, et al. E-nose application for detecting banana fruit ripe levels using artificial neural network backpropagation method[J]. *International Journal of Data Science*, 2022, 3(1): 11-18.

[8] Voss H G J, Stevan S L, Ayub R A. Peach growth cycle monitoring using an electronic nose[J]. *Computers and Electronics in Agriculture*, 2019, 163: 104858.

[9] Voss H G J, Ayub R A, Stevan S L. E-nose prototype to monitoring the growth and maturation of peaches in the orchard[J]. *IEEE Sensors Journal*, 2020, 20(20): 11741-11750.

[10] Wei G F, Gao C, He A X, et al. A spatiotemporal correlation-based data recovery method for improving the performance of electronic noses[J]. *IEEE Transactions on Instrumentation and Measurement*, 2024, 74: 9501811.

[11] Delgado N, Olivera M, CáDiz F, et al. Volatile organic compounds (VOC) produced by *Gluconobacter cerinus* and *Hanseniaspora osmophila* displaying control effect against table grape-rot pathogens[J]. *Antibiotics*, 2021, 10(6): 663.

[12] Huang P W, Li J, Gong Q H, et al. Characterization and analysis of dynamic changes of microbial community associated with grape decay during storage[J]. *Food Microbiology*, 2024, 123: 104581.

[13] 张清安, 陈博宇. 葡萄酒中与风味相关 4 类含硫化合物的研究进展[J]. *中国农业科学*, 2020, 53(5): 1029-1045.

Zhang Q A, Chen B Y. Research progress of four sulfur compounds related to red wine flavor[J]. *Scientia Agricultura Sinica*, 2020, 53(5): 1029-1045.

[14] Liu S X, Lou Y, Li Y X, et al. Aroma characteristics of volatile compounds brought by variations in microbes in wine making[J]. *Food Chemistry*, 2023, 420: 136075.

[15] Bauza T, Kelly M T, Blaise A. Study of polyamines and their precursor amino acids in Grenache noir and Syrah grapes and wine of the Rhone Valley[J]. *Food Chemistry*, 2007, 105(1): 405-413.

- [16] 赵峥, 周德平, 褚长彬, 等. 不同葡萄废果组成对厌氧发酵转化活性物质的影响[J]. 农业环境科学学报, 2022, 41(5): 1116-1121.

Zhao Z, Zhou D P, Chu C B, et al. Effects of different waste grape fruit components on the transformation of active substances by anaerobic fermentation[J]. Journal of Agro-Environment Science, 2022, 41(5): 1116-1121.

[作者简介]



刘静（1981-），女，新疆焉耆人，喀什大学副教授，主要研究方向为深度学习与模式识别等。



窦全胜（1971-），男，黑龙江大庆人，博士，山东工商学院教授，主要研究方向为机器学习、数据挖掘、知识工程与知识处理、智能理论与方法等。



陈炳希（1998-），男，山东枣庄人，山东工商学院硕士生，主要研究方向为计算机视觉、目标检测、机器学习和视觉语言模型等。



宁宇宸（2002-），男，山东泰安人，山东工商学院硕士生，主要研究方向为人工智能、深度学习、知识图谱和大语言模型等。



魏广芬（1978-），女，山东诸城人，博士，山东工商学院教授，主要研究方向为电子鼻系统等。