

西北地区干红葡萄酒产地与酒龄的理化判别方法

白雪冰^{1,2}, 杨佳宁¹, 姜醒睿¹, 赵宇¹, 武运³, 陶永胜^{1,2*}

(1. 西北农林科技大学葡萄酒学院, 杨凌 712100; 2. 西北农林科技大学宁夏贺兰山东麓葡萄酒试验示范站, 永宁 750104;
3. 新疆农业大学食品科学与药学学院, 乌鲁木齐 830052)

摘要:西北地区具有独特的风土条件, 适宜酿酒葡萄的种植, 已形成“大产区大品牌”的产业规划。但西北地区葡萄酒的地域风格尚不明晰, 导致产业的规范性不足, 可持续发展存在隐患。该研究分析西北地区干红葡萄酒的 23 项重要的色泽-味感理化指标, 旨在开发一种基于机器学习技术的葡萄酒判别方法, 实现西北地区干红葡萄酒产地和酒龄的精准判别。首先, 以西北地区 200 款干红葡萄酒为研究对象, 通过理化试验测定了总酚、总花色苷、滴定酸等 23 项色泽-味感理化指标; 然后, 采用 Pearson 相关系数分析了西北产区葡萄酒质量特征的一致性, 并耦合随机森林(Random Forest, RF)分析理化指标对干红葡萄酒产地和酒龄表征的贡献度; 最后, 基于人工神经网络(Artificial Neural Networks, ANN)构建了西北地区干红葡萄酒产地和酒龄的精准判别模型。结果表明, 色泽相关指标的模型贡献率合计 31%, 花色苷相关指标的模型贡献率合计 26%, 酚类物质相关指标的模型贡献率合计 21.1%。模型对宁夏产区酒样判别的灵敏度(Sensitivity, SEN)为 98.72%, 准确率(Accuracy, CCR)98.72%; 对新疆产区酒样判别的 SEN 为 95.45%, CCR 为 100%; 对甘肃产区酒样判别的 SEN 为 100%, CCR 为 95.45%。该方法可以实现西北产区干红葡萄酒产地和酒龄的精准判别, 可为中国西北地区优质葡萄酒的生产和产品的市场监管提供科学依据。

关键词:主成分分析; 模型; 干红葡萄酒; 产地判别; 酒龄判别; 理化指标; 随机森林; 人工神经网络

doi: 10.11975/j.issn.1002-6819.2022.13.035

中图分类号: TP391

文献标志码: A

文章编号: 1002-6819(2022)-13-0319-08

白雪冰, 杨佳宁, 姜醒睿, 等. 西北地区干红葡萄酒产地与酒龄的理化判别方法[J]. 农业工程学报, 2022, 38(13): 319-326.
doi: 10.11975/j.issn.1002-6819.2022.13.035 <http://www.tcsae.org>

Bai Xuebing, Yang Jianing, Jiang Xingrui, et al. Discrimination method for the origin and age of dry red wine by physico-chemical indices in northwest China[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(13): 319-326. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2022.13.035 <http://www.tcsae.org>

0 引言

地域性是决定葡萄酒风格的主要因素之一, 研究不同风土下的葡萄酒质量特征与风格, 不仅有助于揭示葡萄酒酿造品质与调控工艺的独特性, 而且对于地区优质葡萄酒生产与管控也有重要指导作用^[1]。西北地区独特的风土条件适宜于酿酒葡萄的生长, 造就了该地区葡萄极佳的酿造特性和葡萄酒典型的质量特征, 形成了新疆天山南麓、甘肃河西走廊、宁夏贺兰山东麓等酿酒葡萄优质产区^[2]。目前, 西北地区已形成“大产区大品牌”的产业规划, 实现优质葡萄酒的规模化生产, 但其地域性色泽味感特征尚未明确^[3]。综合研究新疆、甘肃、宁夏三地葡萄酒的色泽味感质量, 构建西北产区葡萄酒指纹特征, 有利于规范西北地区葡萄酒酿造行业质量标准, 实现西北地区葡萄酒产品的地域保护, 推动优质葡萄酒的可持

续发展。Lan 等^[4]通过 CIELab 色表体系分析了中国 5 个产区马瑟兰干红的颜色差异, 研究发现胶东半岛和渤海湾产区的亮度值更高, 怀涿盆地、黄土高原和新疆产区的红绿度更高, 以此为依据可初步判断马瑟兰干红的产地。Tsiakkas 等^[5]连续监测塞浦路斯岛 Yiannoudi 和 Maratheftiko 两个本土葡萄品种干红陈酿 3 a 的花色苷变化, 新酒中的总花色苷含量最高可达(509.18±8.18)mg/L, 而陈年酒中总花色苷含量最低可达(48.29±1.52)mg/L, 新酒与陈年酒的总花色苷含量平均差异为 84%, 证明了总花色苷含量可作为干红葡萄酒年份鉴别的重要依据。Cheng 等^[6]汇总了华南地区 9 种干红葡萄酒的 CIELab 表色体系和花色苷相关参数, 绘制了中国华南地区干红葡萄酒的颜色指纹图谱, 用于鉴别干红葡萄酒的产地。Geană 等^[7]解析了干红葡萄酒多酚和挥发酸在中红外波段的映射关系, 并通过特征波段表征了 2009—2017 年份干红葡萄酒的差异信息, 基于线性判别分析算法建立了干红葡萄酒的酒龄鉴别模型。Murru 等^[8]分析了 Bodegas Vidas 酒庄主要种植的 5 种酿酒葡萄采摘时和发酵完成时的单宁变化与联系, 为该酒庄的干红葡萄酒提供了从原料到成品的追溯方法, 实现了该地区干红葡萄酒的原产地保护。Vinciguerra 等^[9]采用三维荧光分析法记录巴西南部干红葡萄酒在不同 pH 值环境下的荧光激发发射矩阵, 实现了 8 种酚酸的精准检测, 以此构建了巴西 Serra

收稿日期: 2022-02-27 修订日期: 2022-06-12

基金项目: 国家重点研发计划项目(2019YFD1002504); 宁夏回族自治区科技重点研发计划项目(2018BBF02001); 新疆维吾尔自治区科技支疆项目计划(2022E02011)

作者简介: 白雪冰, 博士, 讲师, 研究方向为葡萄酒风味化学与智能检测技术。Email: bxb@nwfufu.edu.cn

*通信作者: 陶永胜, 博士, 教授, 博士生导师, 研究方向为葡萄酒酿造与风味化学。Email: taoyongsheng@nwsuaf.edu.cn

Gaucha 产区的认证方法。葡萄酒的质量特征参数可有效反映葡萄酒色泽、香气、味感等感官特性,进而表征葡萄酒的产地和酒龄等信息^[10]。目前,国内外相关研究更多强调特定酒庄或小产区的单一信息,用于横跨经纬度较宽的西北地区干红葡萄酒产地与酒龄判别缺乏系统性和普适性。因此,本研究以西北产区近十年的干红葡萄酒为研究对象,基于现存有效的理化分析方法测定 23 项重要的色泽-味感理化指标,进而量化西北产区葡萄酒的质量特征;分别采用 Pearson 相关系数和随机森林(Random Forest, RF)分析西北产区葡萄酒质量特征的一致性和对干红葡萄酒产地和酒龄表征的贡献度,构建人工神经网络(Artificial Neural Networks, ANN)判别模型,以期形成系统性的西北产区干红葡萄酒产地与酒龄判别方法,为电子舌等葡萄酒质量快速解析方法提供算法支撑。

1 材料与方法

1.1 干红葡萄酒样

本研究采集了中国西北产区 200 款干红葡萄酒,如表 1 所示。每款干红葡萄酒制备 3 个平行酒样,共计获得 600 个干红葡萄酒样品。

表 1 干红葡萄酒样信息
Table 1 Information of dry red wine samples

产区 Origin	品种 Varieties	酒龄 Age	类型 Style	数量 Number
宁夏 Ningxia	赤霞珠、梅鹿辄、蛇龙珠、 品丽珠、马瑟兰	2~7 a	陈酿型、 非陈酿型	112
新疆 Xinjiang	赤霞珠、梅鹿辄、品丽珠、 西拉、马瑟兰	1~4 a、 6~7 a	陈酿型、 非陈酿型	45
甘肃 Gansu	赤霞珠、梅鹿辄、蛇龙珠、 黑比诺、西拉、烟 73	2、3、5 a	陈酿型、 非陈酿型	43
共计 Total				200

1.2 仪器与试剂

紫外可见分光光度计(UV-1780),购置于日本岛津公司,波长范围为 190~1 100 nm,分辨率为 0.5 nm,杂散光≤0.05%。

无水乙醇、浓盐酸、冰醋酸、乙醛、酒石酸、氢氧化钠、明胶、三氯化铁、十二烷基硫酸钠、三乙醇胺、甲醇、氯化钠等均为分析纯,购于西安化学试剂厂;没食子酸、橡黄素、咖啡酸等均为分析纯,购于天津一方科技有限公司;锦葵啉-3-葡萄糖苷、原花青素、Trolox、DPPH(1,1-Diphenyl-2-picrylhydrazyl radical)等均为分析纯,购于北京 Sigma 公司。

1.3 理化检测方法

1.3.1 常规指标检测

参照李运奎等^[11]的方法,计算酒样的 L^* (明度)、 a^* (红/绿通道)、 b^* (黄/蓝通道)、 C_{ab}^* (色度)、 h_{ab} (色调)、 ΔE_{ab}^* (色差)等参数;参考 Cliff 等^[12]的方法,计算酒样的色度和色调,获取物理表征酒样颜色的 CIELab 色表体系参数。

参照郭玉呈等^[13]的方法,分别基于 280 nm 处吸光度值计算总酚含量,320 nm 处吸光度值计算酒石酸酯含量,360 nm 处吸光度值计算黄酮醇含量,520 nm 处吸光度值

计算总花色苷含量。

参照 Alexandre-Tudo 等^[14]的方法,以酒样、SO₂溶液和乙醛溶液的吸光度值的混合运算获得辅色花色苷、单体花色苷和多聚体花色苷的含量;以沸水浴和常温酒样在 550 nm 下吸光度值计算总单宁含量;通过 520 nm 测定酒样吸光度 A_1 、酒样与 NaHSO₃ 混合液吸光度 A_2 、酒样与 HCl 混合液吸光度 A_3 和酒样、NaHSO₃ 与 HCl 混合液吸光度 A_4 混合计算离子化指数。

DPPH 抗氧化活性的测定参考翁夏蒙等^[15]的方法,根据 Trolox 浓度和 517nm 处的吸光值绘制 DPPH 标准曲线,计算求得,结果以 Trolox 的当量值 TEAC 表示,单位为 $\mu\text{mol/L}$ 。

参照 Teixeira 等^[16]的方法,通过全自动电位滴定仪记录消耗 0.1 mol/L 的氢氧化钠滴定液体积换算得到 pH 值和滴定酸值。

1.3.2 乙醇指数

乙醇指数表示干红葡萄酒中与多糖结合的单宁比例,其测定参照 Figueiredo-González 等^[17]的方法。按照酒样:无水乙醇=1:9 体积比配制混合溶液,加蒸馏水稀释 4 倍后测 280 nm 下的吸光度 A_5 ;将混合溶液静置 24 h,离心后取上清液,加蒸馏水稀释 4 倍后测 280 nm 下的吸光度 A_6 ;则乙醇指数为:

$$\text{乙醇指数}=(A_5-A_6)/A_5\times 100 \tag{1}$$

1.3.3 明胶指数

明胶指数表示干红葡萄酒中单宁寡聚体的比例,其测定参照 Figueiredo-González 等^[17]的方法。按照酒样:蒸馏水=1:19 体积比配制混合溶液,测 280 nm 下吸光值 A_7 ;取 4 mL 混合液,加入 2%明胶溶液,在 4℃下静置 48 h,离心后取上清液,测 280 nm 下吸光值 A_8 ;则明胶指数为:

$$\text{明胶指数}=(A_7-A_8)/A_7\times 100 \tag{2}$$

1.3.4 盐酸指数

盐酸指数表示干红葡萄酒中单宁高聚体的比例,其测定参照 Figueiredo-González 等^[17]的方法。按照酒样:蒸馏水:浓盐酸=2:1:3 体积比配制混合溶液,加蒸馏水稀释 15 倍后测 280 nm 下的吸光度 A_9 ;将混合溶液静置 7 h,离心后取上清液,加蒸馏水稀释 15 倍后测 280 nm 下的吸光度 A_{10} ;则盐酸指数为:

$$\text{盐酸指数}=(A_9-A_{10})/A_9\times 100 \tag{3}$$

1.4 数据分析

1.4.1 基于 KNN 的异常值检测

由于滴定误差等不可控因素,样本集中的某些测量值会明显偏离真实值,从而形成异常值。K 邻近算法(K-Nearest Neighbor, KNN)可有效识别并剔除异常值,提高样本集的置信度。KNN 的原理是对未知样本相近的 K 个已知样本进行分类,占比最大的类别即为未知样本的归属类别^[18],具体算法如下:

设异常点判定阈值为 m ,依次计算每个样本点与 K 个最近样本的欧氏距离,求各样本点欧氏距离的平均值。若平均值大于 m ,则认为该样本为异常样本。样本点 x_1 与 x_2 之间的欧氏距离 $D(x_1, x_2)$ 计算公式如下:

$$D(x_1, x_2) = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2} \quad (4)$$

式中 x_1 和 x_2 为 n 维向量, x_{1i} 为 x_1 的第 i 维向量值, x_{2i} 为 x_2 的第 i 维向量值。

1.4.2 基于 RF 的贡献率分析

试验采集了干红葡萄酒样的 23 项色泽-味感理化指标, 可将酒样 x 表示为 $x=(x_1, x_2, \dots, x_{23})$ 。自变量维数过高会增加判别模型的空间复杂度和时间复杂度, 影响判别效率及准确性。通过 RF 算法评估 23 项色泽-味感理化指标对表征西北地区干红葡萄酒产地和酒龄信息的贡献度, 进而选取贡献度较高的指标作为模型的输入量, 可有效降低模型的复杂度。RF 的原理是根据样本特征构建多个弱分类树, 基于每个分类树的结果分析组建一个强分类器以选取最佳特征^[19], 具体步骤如下: 1) 计算出各指标的重要性, 并降序排列。2) 从叶子节点处的特征开始, 依次剔除一个特征并得到相应的特征集。以此迭代进行, 直至剩余 2 个特征位置。3) 比较 2) 中得到的各个特征集对应的袋外误差率, 选取袋外误差率最低的特征集作为最佳特征体系。

1.4.3 基于 ANN 的判别模型构建

以贡献度较高的色泽-味感理化指标为输入量, 产地和酒龄为输出量, 构建基于 ANN 的西北地区干红葡萄酒判别模型, 具体方法如下:

设输入量为 $x=(x_1, x_2, \dots, x_n)$, 其中 $n<23$ 。Sigmoid 函数为:

$$\omega_j^{k+1} = \omega_j^k + \lambda(y_i - \hat{y}_i^k)x_{ij} \quad (5)$$

式中 ω_j 为 x_j 的第 j 个属性的权重; ω_j^k 为第 k 次计算得到

的 ω 值; λ 为学习率常数; y_i 为第 i 个数据的类别, \hat{y}_i^k 为利用第 k 次计算得到的 y_i 值, x_{ij} 是第 i 个数据的第 j 个属性。进而得到输出量 $y'=(\text{产地}+\text{酒龄})$, 如图 1 所示。

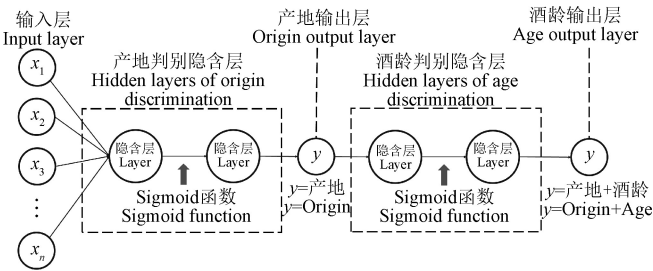


图 1 干红葡萄酒产地与酒龄的 ANN 判别流程图

Fig.1 ANN discrimination flow chart of dry red wine origin and age

为更加科学的评估 ANN 判别模型性能, 分别计算模型的灵敏度 (Sensitivity, SEN) 和准确率 (Accuracy, CCR), 如式 (6) 和 (7) 所示。设 TP 为被正确分类的正例, FN 为本来是正例, 错分为负例, FP 为本来是负例, 被错分为正例。

$$SEN=TP/(TP+FN) \quad (6)$$

$$CCR=TP/(TP+FP) \quad (7)$$

上述数据分析及模型建立均基于 Matlab R2021a 数学软件平台 (美国 MathWorks 公司) 实现。

2 结果与分析

2.1 色泽-味感理化指标解析

试验测定了 600 个干红葡萄酒样品的 23 项色泽-味感理化指标, 如表 2 所示。

表 2 西北产区干红葡萄酒色泽-味感理化指标范围

Table 2 Color and taste parameters ranges of quality characteristics of dry red wine in northwest China

指标 Parameters	宁夏 Ningxia	新疆 Xinjiang	甘肃 Gansu
明度 L^*	(58.55, 82.08)	(50.52, 79.17)	(48.67, 88.75)
红/绿通道 a^*	(15.78, 38.42)	(20.85, 46.32)	(10.68, 47.22)
黄/蓝通道 b^*	(5.83, 22.34)	(0.81, 18.54)	(4.03, 14.37)
色度 C_{ab}^*	(19.41, 41.77)	(22.20, 46.76)	(14.04, 40.43)
色调 h_{ab}	(11.27, 43.71)	(0.92, 34.43)	(6.37, 44.07)
色差 ΔE_{ab}^*	(7.74, 35.42)	(6.36, 32.35)	(4.84, 43.73)
色度 Chrominance	(0.50, 1.32)	(0.52, 1.50)	(0.33, 1.53)
色调 Hue	(0.7, 1.09)	(0.61, 0.94)	(0.61, 1.07)
总花色苷 Total anthocyanins/(mg·L ⁻¹)	(51.65, 210.16)	(51.07, 451.98)	(68.52, 296.57)
辅色花色苷 Accessory anthocyanin/%	(1.66, 15.36)	(0.37, 81)	(0.51, 11.89)
单体花色苷 Monomer anthocyanin/%	(5.35, 50.12)	(7.61, 53.31)	(16.1, 53.82)
多聚体花色苷 Polymeric anthocyanins/%	(46.24, 86.6)	(37.24, 82.65)	(42.41, 73.67)
离子化指数 Ionization index/%	(16.28, 73.59)	(13.67, 84.12)	(10.39, 53.50)
总酚 Total phenol/(mg·L ⁻¹)	(537.21, 1227.33)	(486.36, 2646.31)	(578.54, 1651.52)
黄酮醇 Flavonol/(mg·L ⁻¹)	(338.54, 947.94)	(382.75, 1528.26)	(385.75, 1194.36)
总单宁 Total tannin/(g·L ⁻¹)	(0.84, 5.58)	(1.42, 7.08)	(1.86, 5.32)
乙醇指数 Ethanol index/%	(6.91, 50.23)	(2.58, 68.76)	(5.92, 28.66)
明胶指数 Gelatin index/%	(0.84, 47.52)	(2.32, 51.31)	(1.64, 36.07)
盐酸指数 Hydrochloric acid index/%	(3.98, 39.09)	(3.72, 37.07)	(0.87, 39.05)
DPPH/(μmol·L ⁻¹)	(715.71, 916.27)	(709.35, 932.69)	(735.01, 947.11)
酒石酸酯 Tartaric acid ester/(mg·L ⁻¹)	(459.82, 975.36)	(444.31, 1491.90)	(452.49, 1354.63)
pH 值 pH value	(3.48, 3.94)	(3.63, 4.04)	(3.66, 4.00)
滴定酸 Titrable acid/(g·L ⁻¹)	(4.30, 6.22)	(4.43, 6.97)	(4.53, 7.32)

从 CIELab 色表体系分析可知,宁夏产区酒样 L^* 和 b^* 的取值范围均高于新疆产区,而 a^* 值的范围低于新疆产区;甘肃产区酒样的 L^* 和 a^* 值分布范围较广,而 b^* 值分布范围较窄,反映了西北地区不同子产区干红葡萄酒的颜色特征具有地域辨识性。从呈色物质分析可知,西北产区酒样的总花色苷含量差异较大,其中新疆产区酒样含量 (51.07~451.98 mg/L),可作为地域性判别的重要依据;多聚体花色苷含量明显高于单体花色苷,这是由于在发酵与储藏过程中单体花色苷为非稳态,会在乙醛介导下与黄烷醇聚合生成多聚体。总酚含量最高的酒样来自新疆产区 (2 646.31 mg/L),甘肃和宁夏产区酒样总酚含量最高为 1 651.52 mg/L 和 1 227.33 mg/L,但甘肃产区酒样的 DPPH 测定值最高,宁夏与新疆产区的 DPPH 值相近,这说明葡萄酒清除自由基的能力与酚类物质含量有关,但也受其他因素影响。从涩感相关指标分析,各子产区酒样的总单宁、乙醇指数和明胶指数的测定分布范围差异较大,其中最大值均取自新疆产区酒样,最小值都取自宁夏产区的酒样;而各子产区酒样的盐酸指

数测定范围趋于一致。

由图 2 可知, CIELab 色表体系中 L^* 和 a^* 呈显著负相关,但与 b^* 相关性较低,这说明颜色偏黄会影响酒体明亮度。 C_{ab}^* 和色度的相关性为 0.99,这是由于 C_{ab}^* 和色度在一定程度上都反映颜色的饱和度; h_{ab} 和色调同理。 C_{ab}^* 和 ΔE_{ab}^* 的相关性为 -0.89,这是由于饱和度过高会影响色差的表征程度,与兰圆圆等^[20]对西北产区干红葡萄酒颜色指标的关联分析结果一致。单体花色苷与多聚体花色苷之间呈显著负相关,但与总花色苷之间相关性较弱,这说明花色苷单体与聚合体之间存在相互转化的平衡体系,并且不受总量影响。总酚与黄酮醇呈强正相关,可推断葡萄酒的酚类物质主要为黄酮醇;而与总单宁的相关性较弱,这是因为单宁为黄烷醇类聚合物。总单宁、乙醇指数、盐酸指数和明胶指数之间的相关性较弱,这说明黄烷醇与糖苷、蛋白质或自身缩合三种化学反应之间没有明显的制约关系,赵宇等^[21]在西北地区干红葡萄酒味感质量解析中也证实了这一特性。

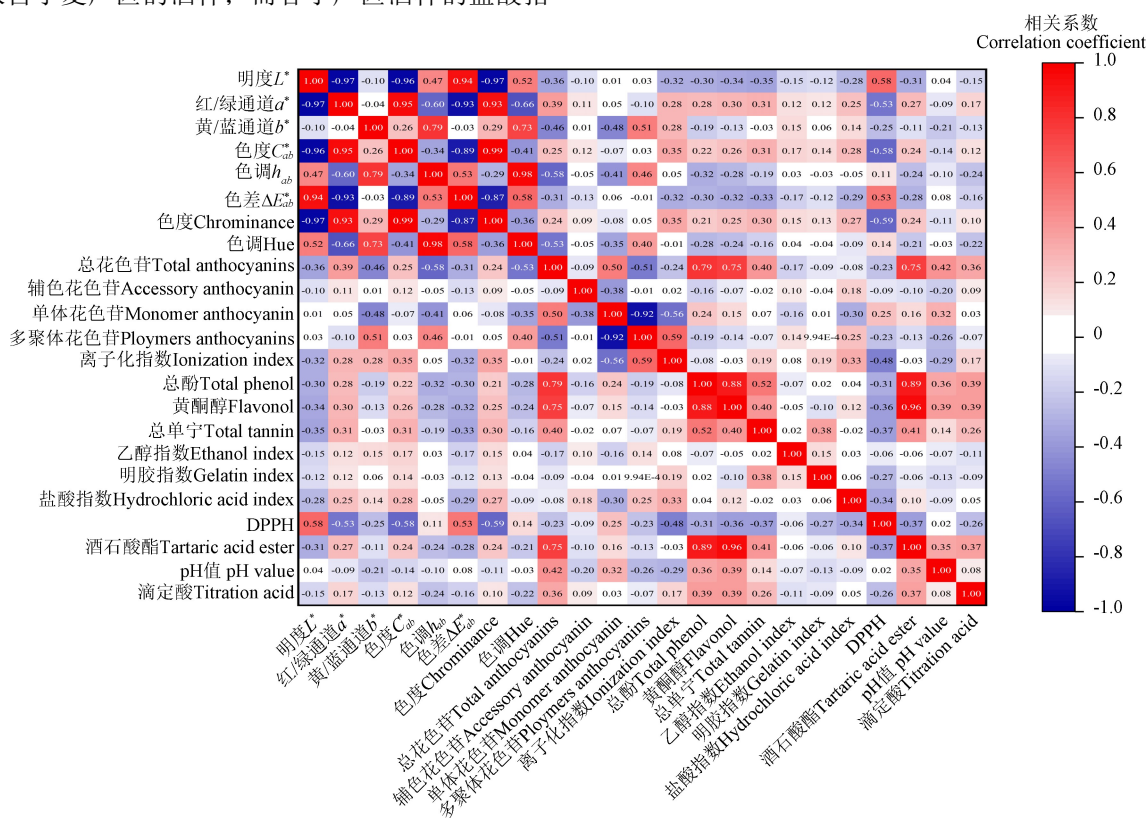


图 2 干红葡萄酒色泽-味感理化指标的相关性分析

Fig.2 Correlation analysis for the color-taste physicochemical indexes of dry red wine

2.2 色泽-味感理化指标的贡献率分析

23 项色泽-味感理化指标的 RF 分析结果如图 3 所示。CIELab 色表体系是表征干红葡萄酒颜色特征的重要指标,合计贡献率达 31%。其中 ΔE_{ab}^* 的贡献率最高 (5.4%),这是由于 ΔE_{ab}^* 表示色差的整体变化,在一定程度上综合反映了 L^* 、 a^* 和 b^* 的差异。总花色苷的贡献度为 13%,花色苷相关指标的贡献度合计为 26%,在所有指标类别中占比最大。这是由于西北地区风土条件不同,导致酿

酒葡萄原料累积的花色苷存在差异^[22],成为了反映干红葡萄酒地域性特征的重要指标。酚类物质相关指标的贡献率合计为 21.1%,其中涩感分析指标在产地判别中的贡献率合计为 15.6%,相对较低,说明西北产区葡萄酒的味感特征相似度较大。另外, pH 值的贡献率为 8.9%,在单指标贡献率中仅次于总花色苷,说明宁夏、新疆、甘肃三个产区葡萄酒的 pH 值存在明显差异,可作为葡萄酒产地和酒龄判别的重要信息。

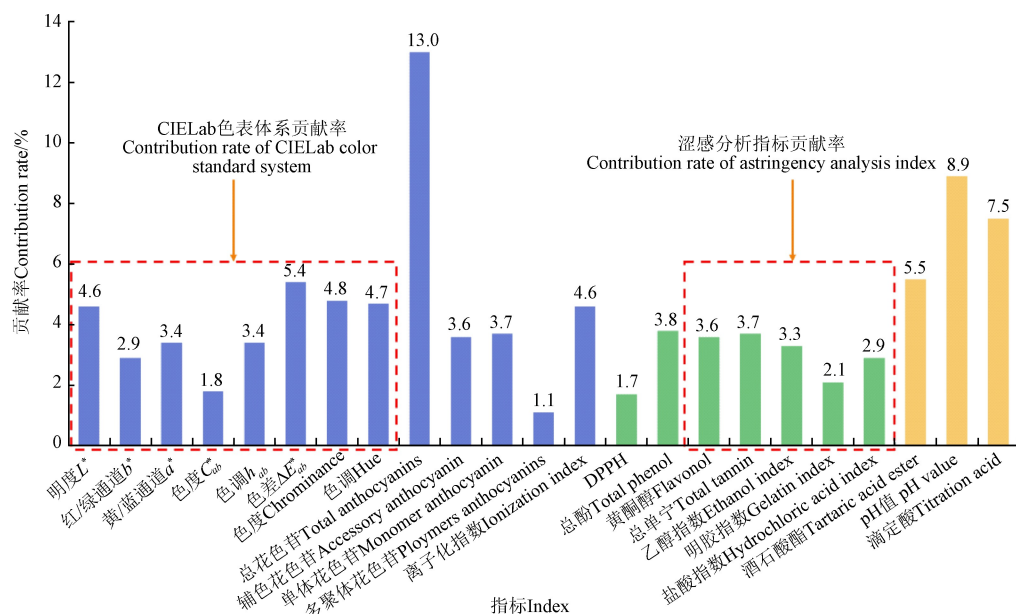


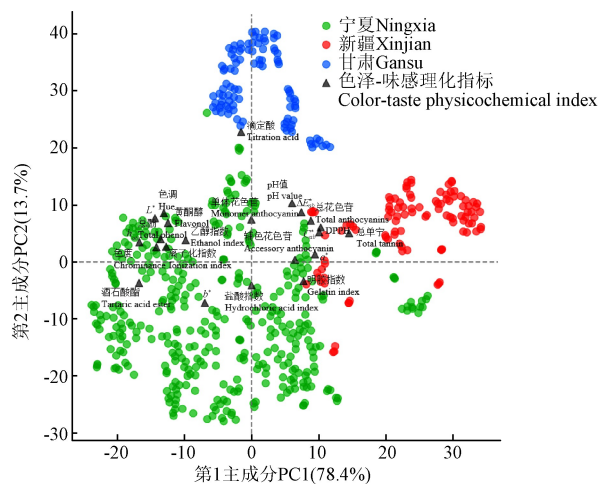
图 3 色泽-味感理化指标的贡献度

Fig.3 Contribution degree of the color-taste physicochemical indexes

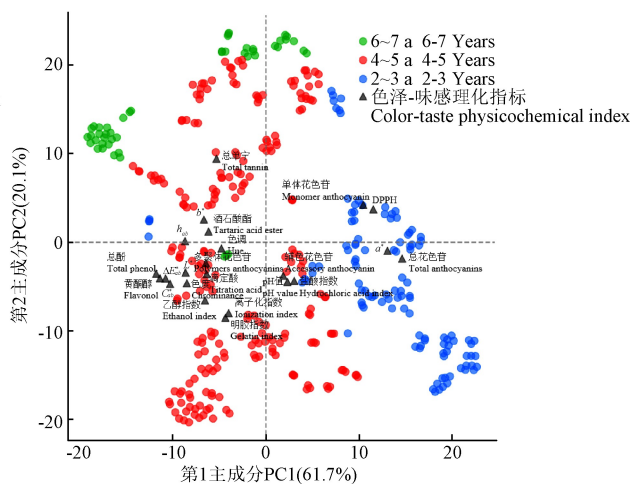
2.3 西北地区干红葡萄酒判别模型分析

将 600 个酒样数据按 4:1 均匀划分为校正集和预测集。基于 ANN 算法构建三层卷积结构的西北地区

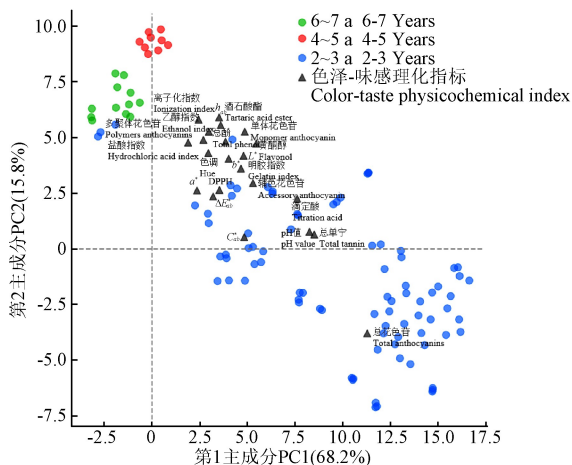
干红葡萄酒产地判别模型及各产地葡萄酒的酒龄判别模型。产地和酒龄判别的 ANN 荷载分布结果如图 4 所示。



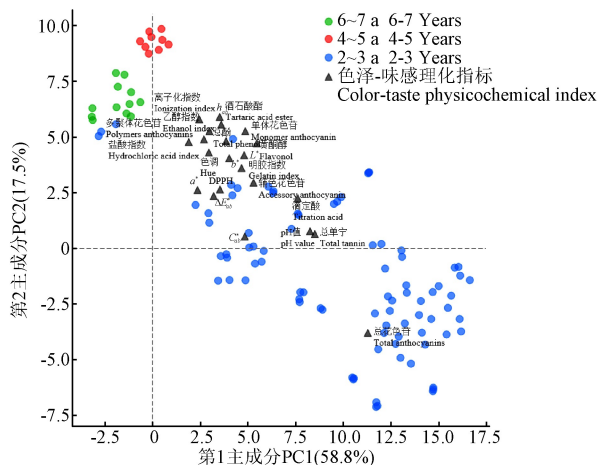
a. 酒样原产地判别



b. 宁夏产区酒龄判别



c. 新疆产区酒龄判别



d. 甘肃产区酒龄判别

图 4 基于贡献度分析的酒样 ANN 荷载分布

Fig.4 ANN's load distribution of wine samples based on contribution analysis

由图 4a 可知,宁夏酒样在每个象限都有分布,CIELab 色表体系和涩感系列指标对宁夏酒样判别的贡献率较高;甘肃酒样主要分布于第一、二象限,总酸含量对甘肃酒样判别起关键作用;新疆酒样集中分布于第一象限,总单宁和 DPPH 对新疆酒样判别具有较大的贡献率。从 600 个酒样的 ANN 荷载分布可知,西北地区各产区酒样分布差异较为明显,仅有少量异常样本,这说明试验获取的 23 项色泽-味感理化指标可以作为西北地区干红葡萄酒产区判别的重要依据。

图 4b~图 4d 分别为宁夏、新疆和甘肃产区干红葡萄酒酒龄判别模型的 ANN 荷载分布,对比可知新酒和陈年酒的分布差异明显。其中,宁夏酒样中陈年酒主要分布于 x 轴的负轴,表现出多数的质量特征;新酒主要分布于 x 轴的正轴,仅在 a^* 、总花色苷、单体花色苷和 DPPH 等有明显的表征。新疆酒样的新酒和陈年酒分布于 x 轴的正负轴,与宁夏产区相同;但大多理化指标分布于新酒附近,与宁夏产区相反。甘肃酒样新酒和陈年酒分布于 x 轴的负轴和正轴,与宁夏和新疆产区相反;大多理化指标在新酒中表征,这一特点与新疆产区相同。ANN 分类器对预测集酒样产地和酒龄的判别结果如表 3 所示。

表 3 基于 ANN 分类器的西北地区干红葡萄酒样判别分析
Table 3 Discriminant analysis of red wine samples in Northwest China based on ANN classifier

产地 Origin	酒龄 Age	宁夏 Ningxia			新疆 Xinjiang			甘肃 Gansu		
		2~3 a	4~5 a	6~7 a	1~2 a	3~4 a	6~7 a	2~3 a	5 a	7 a
宁夏 Ningxia	2~3 a	17								
	4~5 a		50	1						
	6~7 a			9			1			
新疆 Xinjiang	1~2 a				14					
	3~4 a					1				
	6~7 a						6			
甘肃 Gansu	2~3 a							13		
	5 a								7	
	7 a			1						

根据表 3 可知,宁夏产区 78 个酒样分类中,有 1 个酒样被误判为甘肃产区,且有 1 个新疆产区的酒样被误判为宁夏产区,所以模型对宁夏产区酒样的 SEN=98.72%,CCR=98.72%,这说明模型对宁夏产区的酒样较为敏感,且判别准确率较高。新疆产区 22 个酒样分类中,有 1 个酒样被误判为宁夏产区,所以模型对新疆产区酒样的 SEN=95.45%,CCR=100%,这说明模型对新疆酒样较为敏感,且判别准确率很高。甘肃产区 20 个酒样分类中,有 1 个宁夏产区的酒样被误判为甘肃产区,所以模型对甘肃产区酒样的 SEN=100%,CCR=95.24%,这说明模型对甘肃酒样的敏感性很高,且判别准确率也较好。此外,在各产区的酒龄判别分析中,宁夏产区将 1 个酒龄 7 a 的样本误判为酒龄 5 a,酒龄判别的准确率为 98.7%,其他产区酒龄判别结果无误。总体判别结果分析可知,模型对 120 个预测酒样的产地判别准确率可达 98.33%,能够实现西北地区干红葡萄酒的产地和酒龄精准判别。

3 讨论

3.1 干红葡萄酒产地与酒龄判别的依据

中国西北地区独特的风土条件适宜酿酒葡萄的种植,但各产区的土壤与气候条件各不相同,不同年份的日照时长和降雨量有所差异,这就促成了特定产地及年份的酿酒葡萄中花色苷含量和色素形成具有唯一性,进而在干红葡萄酒的直观表现特征为颜色。根据试验结果与分析,宁夏酒样的 L^* 居中,但 b^* 较高;新疆酒样的 L^* 较低,但 a^* 较高;甘肃酒样的 L^* 较高,与吴璐璐等^[23]对宁夏产区干红葡萄酒色泽透亮,黄度更高,偏于砖红色;新疆产区干红葡萄酒色泽偏暗,红度更高,呈樱桃红色;甘肃产区干红葡萄酒色泽明亮,蓝度更深,偏于紫红色的表述一致。由此论证,颜色指标参数可作为构建干红葡萄酒产地与酒龄判别模型的重要依据。

此外,不同产区及年份的酿酒葡萄在糖酸比、单宁含量的方面也具有独特性,从而形成干红葡萄酒味感方面的差异。研究结果中,宁夏酒样的盐酸指数居高,甘肃酒样的盐酸指数偏低,新疆地区的乙醇指数和明胶指数都高于其他产区,与章冉等^[24]对宁夏产区干红葡萄酒口感柔顺,结构温和;新疆产区干红葡萄酒收敛性强、酒体结构强劲;甘肃产区干红葡萄酒涩感较强、柔和感低的评价对应,进而说明味感系列指标也是构建干红葡萄酒产地与酒龄判别模型所需考虑的重要因素。

3.2 干红葡萄酒产地与酒龄判别的方法

色泽-味感理化指标能够可视化描述干红葡萄酒的质量属性,但对于普通消费者感知缺乏直观性。基于机器学习算法建立色泽-味感理化指标与产地和酒龄之间的非线性关系,形成干红葡萄酒产地和酒龄的快速精准判别方法,将客观表征干红葡萄酒产品的原产地和年份属性。色泽指标分析中,CIELab 色表体系和花色苷系列指标均反映干红葡萄酒的颜色属性,但两组指标的相关性较低,这表明二者之间不存在显著的线性关系;此外,单体花色苷和多聚体花色苷之间呈负相关,但二者与总花色苷的相关性都较低,这是由于总花色苷含量的增加并不能增强干红葡萄酒颜色的稳定性。味感指标分析中,总单宁、乙醇指数、盐酸指数和明胶指数等反映多酚物质含量与形态指标之间的相关性也较弱,黄烷醇的缩合方式之间不存在直接的制约关系^[25]。综上可推断,表征同一属性的各个指标对于干红葡萄酒的产地和酒龄判别具有唯一性作用,合理分配 23 项色泽-味感理化指标的贡献率权重是构建西北地区干红葡萄酒产地与酒品判别模型的先行基础。

该研究融合 RF 贡献率分析和 ANN 分类算法建立了西北地区干红葡萄酒产地与酒龄的精准判别模型。由模型解析可知,颜色系列指标对干红葡萄酒产地和酒龄判别的贡献率较高(57%),其中 CIELab 系列指标的合计贡献率高于花色苷系列指标,但单项指标贡献率占比最高的为总花色苷。味感系列指标对干红葡萄酒产地和酒龄判别的贡献率也高达 43%,其中涩感属性指标占比 15.6%,单项指标贡献率占比最高的为 pH 值。色泽-味感

理化指标的贡献率权值分配结果与干红葡萄酒产地与酒龄判别的依据分析一致, 证明了本文构建的判别模型具有一定的理论基础。从判别结果来看, 模型对 120 个预测酒样产地的判别准确率可达 98.33%; 对各产区酒样酒龄的进一步判别准确率分别为宁夏 98.7%、新疆和甘肃 100%, 这表明模型在正常运行下能够准确的判别出酒样的产地来源和年份归属。因此, 研究建立的 ANN 模型可以实现西北地区干红葡萄酒产地与酒龄的精准判别, 可为地区优质葡萄酒的生产和产品的市场监管提供重要的判别依据。

4 结 论

本研究以 600 个干红葡萄酒样品为研究对象, 基于 23 项色泽-味感理化指标的贡献度分析构建了西北地区干红葡萄酒产地和酒龄的判别模型。研究发现, 总花色苷含量是判别西北地区干红葡萄酒所属产地和酒龄的最高贡献率指标。试验结果表明, 模型判别宁夏产区酒样的灵敏度 (Sensitivity, SEN) 为 98.72%, 准确率 (Accuracy, CCR) 为 98.72%; 判别新疆产区酒样的 SEN 为 95.45%, CCR 为 100%; 判别甘肃产区酒样的 SEN 为 100%, CCR 为 95.45%, 可以实现西北地区干红葡萄酒产地和酒龄的精准判别。本研究可为电子舌等葡萄酒质量快速解析方法提供算法支撑, 推动西北地区葡萄酒产地保护及市场监管的发展。

[参 考 文 献]

- [1] 刘巍, 战吉成, 黄卫东, 等. 基于近红外光谱技术的葡萄酒原产地辨识方法[J]. 农业工程学报, 2010, 26(1): 374-378.
Liu Wei, Zhan Jicheng, Huang Weidong, et al. Application of near infrared spectroscopy and clustering analysis to classify wines from different origins[J]. Transactions of the Chinese Society of Agricultural Engineering, 2010, 26(1): 374-378. (in Chinese with English abstract)
- [2] 祝霞, 赵丹丹, 李俊娥, 等. 河西走廊产区酒球菌酯酶活性对葡萄酒酯类香气物质的影响[J]. 农业工程学报, 2021, 37(1): 315-322.
Zhu Xia, Zhao Dandan, Li Jun'e, et al. Effects of esterase activity of *Debaryomyces hansenii* in Hexi Corridor production areas on ester aroma compounds in wine[J]. Transactions of the Chinese Society of Agricultural Engineering, 2021, 37(1): 315-322. (in Chinese with English abstract)
- [3] Liu A, Song H. Analysis and forecasts of the demand for imported wine in China[J]. Cornell Hospitality Quarterly, 2021, 62(3): 371-385.
- [4] Lan Y, Liu M, Zhang X, et al. Regional variation of chemical characteristics in Young Marselan (*Vitis vinifera* L.) red wines from five regions of China[J]. Foods, 2022, 11(6): 787.
- [5] Tsiakkas O, Escott C, Loira I, et al. Determination of anthocyanin and volatile profile of wines from varieties Yiannoudi and Maratheftiko from the Island of Cyprus[J]. Beverages, 2020, 6(1): 4.
- [6] Cheng G, Zhou S H, Wen R D, et al. Anthocyanin characteristics of wines in *Vitis* germplasms cultivated in southern China[J]. Food Science and Technology, 2017, 38: 513-521.
- [7] Geană E I, Ciucure C T, Apetrei C, et al. Application of spectroscopic UV-Vis and FT-IR screening techniques coupled with multivariate statistical analysis for red wine authentication: Varietal and vintage year discrimination[J]. Molecules, 2019, 24(22): 4166.
- [8] Murru C, Chimeno-Trinchet C, Díaz-García M E, et al. Artificial neural network and attenuated total reflectance-fourier transform infrared spectroscopy to identify the chemical variables related to ripeness and variety classification of grapes for protected. Designation of origin wine production[J]. Computers and Electronics in Agriculture, 2019, 164: 104922.
- [9] Vinciguerra L L, Böck F C, Schneider M P, et al. Geographical origin authentication of southern Brazilian red wines by means of EEM-pH four-way data modelling coupled with one class classification approach[J]. Food Chemistry, 2021, 362: 130087.
- [10] Jin Y, Shu N, Xie S, et al. Comparison of 'Beibinghong' dry red wines from six producing areas based on volatile compounds analysis, mineral content analysis, and sensory evaluation analysis[J]. European Food Research and Technology, 2021, 247(6): 1461-1475.
- [11] 李运奎, 韩富亮, 张子林, 等. 基于 CIELAB 色空间的红葡萄酒颜色直观表征[J]. 农业机械学报, 2017, 48(6): 296-301.
Li Yunkui, Han Fuliang, Zhang Yulin, et al. Visualization for representation of red wine color based on CIELAB color space[J]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(6): 296-301. (in Chinese with English abstract)
- [12] Cliff M A, King M C, Schlosser J. Anthocyanin, phenolic composition, colour measurement and sensory analysis of BC commercial red wines[J]. Food Research International, 2007, 40(1): 92-100.
- [13] 郭玉呈, 李桂善, 王俊婷, 等. 不同类型发酵柿子果酒的质量和抗氧化活性[J]. 食品与发酵工业, 2017, 43(12): 108-113.
Guo Yucheng, Li Guishan, Wang Juntong, et al. Quality evaluation and antioxidant activity of fermented persimmon wines[J]. Food and Fermentation Industries, 2017, 43(12): 108-113. (in Chinese with English abstract)
- [14] Alexandre-Tudo J L, du Toit W. A chemometric approach to the evaluation of the ageing ability of red wines[J]. Chemometrics and Intelligent Laboratory Systems, 2020, 203: 104067.
- [15] 翁夏蒙, 鲁光耀, 王鸯妮, 等. 白及愈伤组织总酚含量测定及抗氧化作用研究[J]. 中药材, 2013, 36(1): 32-35.
Weng Xiameng, Lu Guangyao, Wang Yuanni, et al. Determination of total phenol content in white callus and its antioxidant activity[J]. Journal of Chinese Medicinal Materials, 2013, 36(1): 32-35. (in Chinese with English abstract)
- [16] Teixeira Dos Santos Cláudia A, Páscoa Ricardo N M J, Porto Patrícia ALS, et al. Raman spectroscopy for wine analyses: A comparison with near and mid infrared spectroscopy[J]. Talanta, 2018, 186: 306-314.
- [17] Figueiredo-González M, Cancho-Grande B, Simal-Gándara J. Garnacha Tintoreria-based sweet wines: Chromatic properties and global phenolic composition by means of UV-Vis spectrophotometry[J]. Food Chemistry, 2013, 140(1/2): 217-224.
- [18] Wang B, Mao Z. A dynamic ensemble outlier detection model based on an adaptive k-nearest neighbor rule[J]. Information Fusion, 2020, 63: 30-40.
- [19] Speiser J L, Miller M E, Tooze J, et al. A comparison of random forest variable selection methods for classification prediction modeling[J]. Expert Systems with Applications, 2019, 134: 93-101.
- [20] 兰圆圆, 陶永胜, 张世杰, 等. 我国多产区干红葡萄酒颜色相关指标的关联分析[J]. 食品科学, 2013, 34(11): 1-4.
Lan Yuanyuan, Tao Yongsheng, Zhang Shijie, et al. Correlation analysis of color parameters and chemical components of Chinese red wines from different growing

- regions[J]. Food Science, 2013, 34(11): 1-4. (in Chinese with English abstract)
- [21] 赵宇, 沙青, 孔彩琳, 等. 西北地区干红葡萄酒质量相关理化指标的判别功能解析[J]. 食品科学技术学报, 2021, 39(3): 129-139.
Zhao Yu, Sha Qing, Kong Cailin, et al. Discriminant analysis of physicochemical indexes related to quality of dry red wines from northwest China[J]. Journal of Food Science and Technology, 2021, 39(3): 129-139. (in Chinese with English abstract)
- [22] Wang Z, Yin H, Yang N, et al. Effect of vineyard row orientation on microclimate, phenolic compounds, individual anthocyanins, and free volatile compounds of Cabernet Sauvignon (*Vitis vinifera* L.) in a high-altitude arid valley[J]. European Food Research and Technology, 2022, 248(5): 1365-1378.
- [23] 吴璐璐, 范舒悦, 张煜, 等. 不同产区陈酿干红葡萄酒花色苷与颜色相关性分析[J]. 中外葡萄与葡萄酒, 2022(2): 1-7.
Wu Lulu, Fan Shuyue, Zhang Yue, et al. Correlation between anthocyanins and color of aging red wines from different regions of China[J]. Sino-Overseas Grapevine & Wine, 2022(2): 1-7. (in Chinese with English abstract)
- [24] 章冉, 田湑, 安然, 等. 卵清蛋白沉淀法分析干红葡萄酒的涩感质量[J]. 食品科学技术学报, 2016, 34(4): 80-84.
Zhang Ran, Tian Tian, An Ran, et al. Astringency analyzing of dry red wines with ovalbumin assay[J]. Journal of Food Science and Technology, 2016, 34(4): 80-84. (in Chinese with English abstract)
- [25] Li S Y, Duan C Q. Astringency, bitterness and color changes in dry red wines before and during oak barrel aging: An updated phenolic perspective review[J]. Critical Reviews in Food Science and Nutrition, 2019, 59(12): 1840-1867.

Discrimination method for the origin and age of dry red wine by physico-chemical indices in northwest China

Bai Xuebing^{1,2}, Yang Jianing¹, Jiang Xingrui¹, Zhao Yu¹, Wu Yun³, Tao Yongsheng^{1,2*}

(1. College of Enology, Northwest A&F University, Yangling 712100, China; 2. Ningxia Helan Mountain's East Foothill Wine Experiment and Demonstration Station, Northwest A&F University, Yongning 750104, China; 3. College of Food Science and Pharmacy, Xinjiang Agricultural University, Urumqi 830052, China)

Abstract: Northwest China has formed into a benign industrial area named “big appellation and big brand”. The unique black soil can also be a benefit for wine grape planting. However, the regional flavors of dry red wine vary greatly in the different regions in Northwest China. Potential risk can remain for the sustainable development of the wine industry. In this study, a wine discrimination system was developed to accurately identify the origin and age of dry red wine using machine learning. 23 physical and chemical parameters were determined about the wine color, taste, and aroma. Firstly, 600 samples of dry red wine were collected from the different wineries in Ningxia, Gansu, and Xinjiang areas. The total phenols, total anthocyanin, and titrated acids were then measured during the experiment. Secondly, the Pearson correlation coefficient formula was selected to evaluate the consistency of 23 dry red wine parameters. Subsequently, random forest (RF) was used to calculate the percentage contributions of each parameter for the identification of the origin and age of dry red wines. Finally, an accurate discrimination system was developed to identify the origin and vintage of dry red wines using an artificial neural network (ANN) classifier. The results showed that the eight color-related parameters provided with 31% reference to distinguish the origin and year of dry red wine. Specifically, the wine samples from Ningxia's wineries showed higher yellowness and brick red. The wine samples from Xinjiang's wineries showed a high degree of redness, like the peach red. The wine samples from Gansu's wineries showed higher blueness and purple. Five anthocyanin and seven phenolic-related parameters were provided with 26% and 21% reference, respectively. In terms of taste, Ningxia's wine samples had less astringency, while Gansu's wine samples showed a stronger astringency, and Xinjiang's wine samples tasted the most astringent. The loading analysis demonstrated that the Ningxia wine shared a certain aging potential because the old wine from the Ningxia showed more flavor characteristics. By contrast, the new wine from Xinjiang presented strong flavor characteristics, but the old wine showed weak flavor characteristics, indicating the quick dissipation of flavor as time increased. Similar flavor characteristics were achieved in the old and new Gansu wine, indicating the slow quality loss of Gansu wine during aging. The sensitivity (SEN) and accuracy (CCR) of the wine's origin discriminant model were 98.72% and 98.72% for the wine samples from Ningxia's wineries, respectively, 95.45% and 100% for the wine samples from Xinjiang's wineries, respectively, while 100% and 95.45% for the wine samples from Gansu's wineries, respectively. The correct rate of the wine's age discriminant model for each region was 98.7% for the Ningxia wine, and 100% for the Xinjiang wine and Gansu wine. The discrimination model can be expected to make accurate discrimination of wine origin and age from Northwest China. The finding can provide scientific support for the production of premium wine in the market supervision in these regions.

Keywords: principal component analysis; models; dry red wine; origin discrimination; age discrimination; physico-chemical index; random forest; artificial neural networks