基于近红外快检和高斯过程回归的单-多茶类数字化拼配模型构建

俞秋雯[1,2]，周洁[1,2]，张浩[1,2]，高海燕[1,2]，许凌[1,2]，王家鹏[1]，杜颖颖[1,2*]

（1.中华全国供销合作总社杭州茶叶研究所，浙江杭州 310016）（2.国家茶叶质量检验检测中心，浙江杭州 310016）

摘要：传统茶叶拼配高度依赖人工感官经验，存在耗时费力、主观偏差大等固有局限，难以满足当前茶叶消费升级背景下，市场对多茶类跨界融合及风味创新拼配日益增长的快速响应需求。本研究构建了基于近红外光谱与高斯过程回归（Gaussian Process Regression，GPR）的数字化拼配模型：以 886 个全国主产区茶叶样品为对象，采用偏最小二乘法建立含水分、水浸出物、茶多酚、咖啡碱、粗纤维、游离氨基酸 6 项品质指标的近红外定量模型，模型定标相关系数高于 0.92，外部验证决定系数（R-Square，R^2）大于 0.98。在此基础上，选取 27 种不同茶类、外形的原料茶，设计涵盖多类场景的拼配方案，以近红外数据为输入，训练径向基函数与白噪声核组合的 GPR 模型进行不同比例拼配样的品质预测，其模型独立测试集 R^2 为 0.95，均方误差 0.12，五折交叉验证 R^2 为 0.92±0.04。部分拼配茶样国标法检测实例结果表明，除茶多酚含量外，国标法实测 5 种茶叶内含物质含量基本落在 GPR 模型预测区间值，且在跨多茶类复杂场景下噪声指标最低，噪声整体均值为 2.79。本研究结果为茶叶拼配由经验主导型向数据驱动型转变提供技术支撑。

关键词：近红外光谱；高斯过程回归；茶叶拼配；数字化拼配；算法

DOI: 10.13982/j.mfst.1673-9078.2027.9.1136

Design of Digital Blending Model of Single-and Multi-Category Teas Using Near-Infrared Spectroscopy Detection and Gaussian Process Regression

YU Qiuwen[1,2], ZHOU Jie[1,2], ZHANG Hao[1,2], GAO Haiyan[1,2], XU Ling[1,2], WANG Jiapeng[1], DU Yingying[1,2*]

(1.Hangzhou Tea Research Institute, CHINA COOP, Hangzhou 310016, China)

(2.National Center for Tea Quality Inspection and Testing, Hangzhou 310016, China)

Abstract: Traditional tea blending which is heavily dependent on human sensory experience, is inherently time-consuming, labor-intensive, and subject to considerable subjective bias. These limitations make it increasingly inadequate to address the escalating market demand for rapid adaptation to multi-category teas flavor innovation, especially in the context of the current upgrading of tea consumption patterns. Thus, a digital blending model based on near-infrared spectroscopy and gaussian process regression (GPR) had been developed. A total of 886 tea samples from major production regions across China were used to establish a near-infrared quantitative model with six quality indicators via partial least squares method, which includes moisture, water extracts, tea polyphenols, caffeine, crude fiber and free amino acids. The calibration correlation coefficient of the model was higher than 0.92, and the external validation R-Square (R^2) was greater than 0.98. Furthermore, 27 raw tea materials of different categories and shapes were selected to design blending schemes covering various scenarios. With near-infrared data as input, a GPR model combining radial basis function and white noise kernels was trained for quality prediction, which achieved an R^2 of 0.95, a mean square error of 0.12 on the independent test set, and a five-fold cross-validation R^2 of 0.92±0.04. The results of actual tests on several blended tea samples using national standard methods indicated that, despite the content of tea polyphenols, other five tea chemical components tested by the national standard methods basically fell within the prediction intervals provided by GPR model. Moreover, the noise index was the lowest in the complex scenario spanning multiple tea categories, with an overall mean noise of 2.79. This study provides technical support for transforming tea blending from an experience-driven approach to a data-driven one.

Key words: near-infrared spectroscopy; gaussian process regression; tea blending; digital blending; algorithm

收稿日期：2026-07-30；修回日期：2026-10-08；接受日期：2026-10-08

基金项目：浙江省农业重大技术协同推广计划（2024ZDXT05-05）

作者简介：俞秋雯（1999-），女，硕士，研究实习员，研究方向：茶叶感官风味及品质化学研究，E-mail：m15827457152@163.com

通讯作者：杜颖颖（1983-），女，博士，研究员，研究方向：茶叶感官风味、品质化学和标准化研究，E-mail：duyy83@sina.com

茶叶是全球消费量最大的健康饮品之一，因其品质受茶树生长环境、鲜叶加工季节等因素影响，其从生产端流向消费端过程中所必需的风味稳定性，一般通过拼配技术实现[1-3]。传统茶叶拼配高度依赖经验丰富的审评专家，然而随着新茶饮赛道的爆发式增长以及消费者多元化的产品需求，传统依赖人工经验的拼配模式正面临严峻挑战[3-6]。一方面，专业拼配人才培养周期长，产业面临人才短缺困境，另一方面，跨茶类风味融合等特需创新使茶叶拼配复杂度与品质要求特异性持续提升，而传统经验模式应对市场迭代反应迟缓，因此亟需能够快速响应的数据驱动型拼配技术[7,8]。

近红外光谱检测是一种快速、无损的检测手段，其光谱结果可结合化学计量学分析，共同实现大豆、咖啡豆、茶叶等产品的多组分快速测定[9-13]。在茶叶品质检测中该技术已趋成熟，可快速测定水分、茶多酚、咖啡碱、游离氨基酸等核心品质成分[14-16]，这些指标与茶叶滋味品质高度相关，可为拼配过程提供量化数据支撑[17,18]。通过广泛采集茶叶近红外光谱，基于定量模型可快速获取茶叶内含品质数据，为数字化拼配模型构建提供高质量数据基础[19-23]。

机器学习是计算机系统通过发现和利用数据统计规律以提高自身性能的方法[24]，已成为茶叶智能化拼配研究的核心算法，在祁门红茶拼配预测、眉茶拼配预测等生产场景中均有应用探索[25-28]。然而，上述模型大多聚焦于单一茶类内部拼配，面对新茶饮业态下跨茶类拼配的复杂需求仍显不足。高斯过程回归是一种基于贝叶斯理论的非参数概率模型，具有实施简便、预测结果自带不确定性评估及随数据累积同步更新的能力，能适应原料茶品类和库存的动态变化[29]，该模型在产品成分检测[30]、单一茶类拼配研究中表现出应用潜力[31,32]，但尚缺少光谱技术结合高斯回归过程在多茶类茶叶拼配中的应用开发[33,34]。

本研究以近红外快速检测与高斯过程回归为核心，构建了面向茶叶新业态的单茶类及多茶类数字化拼配方案。近红外建模阶段以 886 个茶叶样品为对象，采用国标法测定水分、水浸出物、茶多酚、咖啡碱、粗纤维、游离氨基酸含量，经预处理与算法优化建立高精度定量预测模型。基于该模型，选取 27 个不同品类、外形的原料茶，设计不同比例组合，构建 GPR 拼配预测模型，实现对源自在库原料茶的任意组合拼配茶样内含品质的精准预测。研究成果为面向市场需求的快速拼配提供技术支撑，有助于推动茶叶拼配从“经验主导”向“数据驱动”的智能化转型。

1 材料与方法

1.1 材料与试剂

试验前期收集包含六大茶类和不同等级抹茶在内的 886 个茶叶样品，样品遍及全国所有茶叶产区，包括浙江、福建、安徽、江苏、广东、广西、云南、西藏等。每类茶叶含有 100 个以上不同生产时间和等级的样品。除以上样品外，另选 20 个茶叶样品用于模型的外部预测和效果验证。

用于构建拼配模型的 27 个不同品类和外形的原料茶，包括绿茶、红茶、白茶、黄茶、抹茶等，其编码和具体产地信息见表 1。

表 1 不同品类原料茶编号及产地信息

Table 1 Serial numbers and origin information of different categories of tea

Category	Shape/Grade	Sample ID	Geographical origin
绿茶	扁形（龙井茶）	P1、P5、P6、P8、P10	浙江
	颗粒形	KL1、KL2、KL3	
	卷曲形	U1、U2	
红茶	条形	RT1、RT2、RT3	浙江、福建、云南
	红碎茶	RP1、RP2、RP3	
黄茶	片状和芽叶型	H1、H2、H3	浙江
白茶	牡丹和寿眉	B1、B3	福建
抹茶	粉末状	M1、M2、M3	浙江
花茶	卷曲形和条形	ML1、ML2、ML3	浙江、福建、四川

标准品及化学试剂：硫酸、氢氧化钠、盐酸、乙醇、丙酮、甲醇、碳酸钠、福林酚、没食子酸、氧化镁、磷酸氢二钠、磷酸二氢钾、茚三酮、二水合氯化亚锡、茶氨酸（分析纯），均购于国药集团化学试剂有限公司；甲醇

（色谱纯），德国 Merck 公司；咖啡碱（纯度>99.8 wt.%），均购于中国计量科学研究院。

1.2 仪器与设备

Thermo U3000 液相色谱仪，美国 Thermo Fisher Scientific 公司；Tu-1901 紫外可见分光光度计，北京普析通用仪器有限责任公司；SOP 电子天平，赛多利斯科学仪器（北京）有限公司；DGG-9140A 电热恒温鼓风干燥箱，上海精宏实验设备有限公司；SC3610 低速离心机，安徽中科中佳科学仪器有限公司；DK-S26 电热恒温水浴锅，上海精宏实验设备有限公司；Direct3UV 超纯水仪，法国 Millipore 公司；SupNIR-2700 近红外分析仪（配有 RIMP 测量分析软件），聚光科技（杭州）股份有限公司。

1.3 实验方法

1.3.1 茶叶关键品质指标测定

水分含量的测定按照 GB 5009.3-2016《食品安全国家标准 食品中水分的测定》；水浸出物的测定按照 GB/T 8305-2013《茶 水浸出物》；茶多酚的测定按照 GB/T 8313-2013《茶叶中茶多酚和儿茶素类含量的检测方法》；咖啡碱的测定按照 GB/T 8312-2013《茶 咖啡碱测定》；粗纤维含量的测定按照 GB/T 8310-2013《茶 粗纤维测定》；游离氨基酸总量的测定按照 GB/T 8314-2013《茶 游离氨基酸总量的测定》。

1.3.2 茶叶近红外光谱采集

SupNIR-2700 近红外分析仪开机预热 30 min 后，用白色参比盘校准。设置光谱波长范围 1 000~1 800 nm，循环测量 10 次，间隔 10 s，系统自动计算 10 条光谱的平均值作为样品的最终光谱。每个样品重复装样检测 3 次，填样及检测操作参考李文萃等[16]的检测方法。

1.3.3 茶叶近红外定量模型构建

光谱数据预处理：参照 GB/T 37969-2019《近红外光谱定性分析通则》中光谱数据预处理一节规定的内容，根据茶样近红外光谱的实际情况，采用标准正态变量变换、去趋势校正、Savitzky-Golay 平滑、Savitzky-Golay 一阶导数、均值中心化等方法对光谱进行预处理。

模型建立方法：将每个样品的光谱数据与其经国标法测定的 6 项品质指标（水分、水浸出物、茶多酚、咖啡碱、粗纤维、游离氨基酸）的检测值一一对应，作为建模输入。应用 RIMP 测量分析软件，将包含 886 个茶叶样品光谱的数据集随机划为校正样品集和验证样品集，采用偏最小二乘法（Partial Least Squares，PLS）交叉验证，通过优选因子数和剔除异常样本达到最佳模型效果，建立集成式的含有茶叶 6 项品质指标的近红外快速检测模型。

1.3.4 近红外定量模型验证

随机选取 20 个独立于近红外建模样品集之外的茶样，代入 1.3.3 已完成构建的茶叶近红外定量模型，按照 1.3.2 采集方法上机检测后获取其水分、水浸出物、茶多酚、咖啡碱、粗纤维、游离氨基酸含量的近红外检测值，并与茶样按 1.3.1 方法测定的化学检测值进行比较，以验证近红外定量模型的有效性和准确性。

1.3.5 原料茶及拼配样近红外检测

将 27 种原料茶按 1.3.2 方法采集近红外光谱，代入 1.3.3 所建定量模型，得到各品质指标预测值，与原料茶基础信息合并存储为基础原料茶数据集。选取同一茶类或跨茶类的 2~3 种原料茶，按不同质量比例拼配，共获得 70 个拼配茶样。将这些拼配茶样同样经近红外检测，得到品质指标预测值，与拼配比例一同存储为拼配茶数据集。两份数据集将用于后续的拼配预测模型构建和数据分析。

1.3.6 茶叶近红外定量模型评价

所建模型的评价指标主要包括定标标准偏差、定标相关系数、内部交互验证标准偏差、性质值方差与内部交互验证标准偏差的比值和主因子数，评价指标由 RIMP 测量分析软件建模同时生成。

1.3.7 茶叶近红外定量模型的外部验证评估

对于 1.3.4 中的 20 个茶样的化学测定值和近红外检测值进行对比分析，以评估水分、水浸出物、茶多酚、咖啡碱、粗纤维、游离氨基酸模型的预测值和真实值之间的吻合度。

1.3.8 高斯过程回归算法建立单-多茶类数字化拼配模型

本研究采用高斯过程回归（Gaussian Process Regression，GPR）模型，采用核函数为径向基函数（Radial Basis Function，RBF），以原料茶拼配比例和拼配后的茶叶关键内含物质含量为输入量进行训练：

$$K_{RBF}(x,x') = exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right) \quad (1)$$

式中:

$K_{RBF}(x,x')$——核函数值;

x——特征向量，独立样本点;

x'——特征向量，不同于 x 的另一个的独立样本点;

$\|x-x'\|^2$——两个向量 x 和 x' 在空间中的直线距离的平方;

exp——自然指数函数;

σ ——带宽参数，控制函数变化的平滑程度。

1.4 数据处理与分析

茶样的近红外光谱数据由 RIMP Client V2.0 收集、存储并统一构建预测模型。原料茶及其拼配样数据由 RIMP Client V2.0 模型预测获得后，使用 Microsoft Excel 2016 软件进行数据处理，采用 Python 建立高斯过程回归模型并完成模型相关分析，设定噪声评价指标用以评估模型性能：

$$N = \frac{E}{|\hat{y}-y_{avr}|} \quad (2)$$

式中:

E——置信区间半宽，wt.%;

$\hat{y}$ ——GPR 点预测值，wt.%;

y_{NIR} ——近红外实测均值，wt.%。

2 结果与讨论

2.1 近红外定量模型评价

定标标准偏差（Standard Error of Calibration，SEC）、定标相关系数（Correlation Coefficient of Calibration，RC）、内部交互验证标准偏差（Standard Error of Cross-Validation，SECV）、性质值方差与 SECV 的比值（Ratio of Performance to Deviation of Calibration，RPDC）是综合评估模型预测效果的评价指标。由表 2 可见，水分、水浸出物、茶多酚、咖啡碱、粗纤维、游离氨基酸 RC 均高于 0.9，SEC 和 SECV 在各性质指标中数值均小于 1.6 且比值较小，说明该模型的近红外分析结果与化学检测结果吻合度高，模型预测效果良好[16]。

集成模型预测效果整体良好，但各性质成分的预测性能存在差异。其中，咖啡碱模型的 SECV（0.37）与 SEC（0.30）最为接近，RPDC 达到 4.66，且 RC 为 0.99，表明咖啡碱模型具有优秀的定量预测精度。游离氨基酸模型相关系数 RC 为 0.92，RPDC 值为 2.07，略低于其他性质但仍大于 2，适用于茶样游离氨基酸含量的定量预测。

表 2 不同品质成分的模型参数

Table 2 Model parameters of different quality components

Component	SEC	RC	SECV	SECV/SEC	RPDC	Factors	Range/%	Samples
水分	0.55	0.98	0.61	1.10	4.05	12	1.12~12	459
水浸出物	1.44	0.96	1.56	1.08	3.46	11	23.6~52.4	518
茶多酚	1.20	0.97	1.27	1.06	3.75	7	4.7~24.3	171
咖啡碱	0.30	0.99	0.37	1.24	4.66	9	0.25~5.2	142
粗纤维	0.92	0.96	1.19	1.29	2.60	5	6.27~18.79	157
游离氨基酸	0.55	0.92	0.68	1.22	2.07	9	0.6~6.4	130

注：表中性质范围%为质量分数 wt.%。

2.2 近红外定量模型的外部验证

对 20 个独立于近红外建模样品集之外的茶样编号为 1~20，绘制其真实值（国标法化学检测值）和预测值（近红外模型检测值）折线图，结果如图 1 所示。从图中看出，20 个外部样品不同性质的真实值与预测值分布存在一定差异，由折线图走势可见，水分和咖啡碱的真实值和预测值重合程度最高。线性拟合结果显示，样品水分含量的预测值和真实值线性拟合结果为 y=0.958 8x，R^2=0.996；样品咖啡碱含量的预测值和真实值线性拟合结果为 y=0.987 5x，R^2=0.997，在外部验证中均表现出良好的线性关系。

部分样品粗纤维和游离氨基酸含量的真实值和预测值存在一定偏差，但粗纤维含量线性拟合结果为 y=0.975 6x，R^2=0.997；游离氨基酸含量线性拟合结果为 y=0.950 2x，R^2=0.984。此外，样品水浸出物含量的预测值和真实值的线性拟合结果为 y=1.022 4x，R^2=0.998；样品茶多酚含量的预测值和真实值的线性拟合结果为 y=0.962 7x，R^2=0.995，整体拟合效果较好，但在茶多酚含量 16%~22%（质量分数）的较高含量区间，样品的茶多酚含量真实值略高于近红外预测值。以上结果表明，所有性质的真实值与预测值的线性回归系数范围为 0.95~1.02，且决定系数 R^2 均大于 0.98，近红外模型的预测解释能力强。

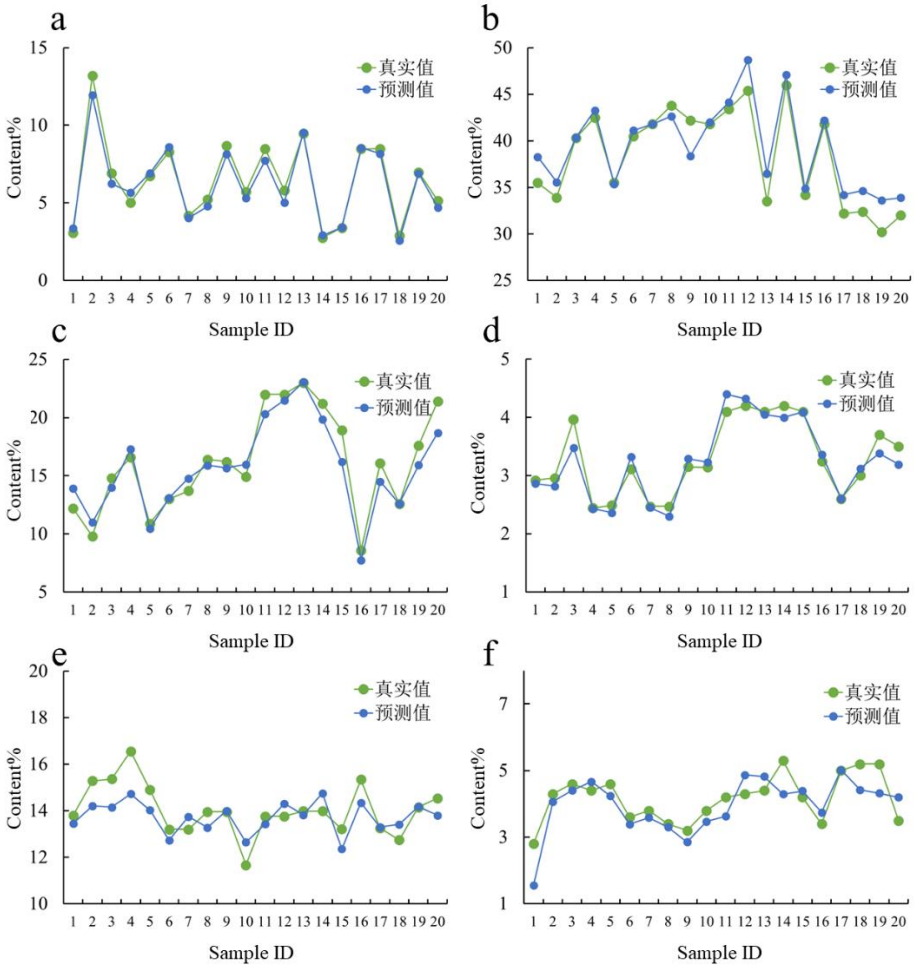


图 1 茶叶各品质成分模型预测值与真实值验证

Fig.1 Verification of predicted and actual values of tea quality component models

注:（a）水分;（b）水浸出物;（c）茶多酚;（d）咖啡碱;（e）粗纤维;（f）游离氨基酸; 纵坐标中%为质量分数 wt.%。

2.3 原料茶及拼配样的 NIRS 数据分析

将不同品类、外形和产地的 27 个原料茶及拼配茶样，装填入样品盘中，近红外分析仪采集得到的近红外光谱信息如图 2 所示。在近红外波长 1 000~1 400 nm 区间，样品的吸光度分布较为均匀，在近红外波长 1 440~1 800 nm 区间，样品的吸光度分成两个区域，其中 a 区样品的吸光度均高于 0.75，b 区样品的吸光度均低于 0.75。中波近红外谱区中 1 440、1 450、1 500 和 1 720 nm 分别对应 H_2O、O-H、N-H 和 C-H 的二倍频振动[35]。

分析 a、b 区的样品编号，其中 14 种原料茶位于 NIR 光谱曲线图中的 a 区，样品品类为卷曲形、颗粒形绿茶，

红茶和茉莉花茶；13 种原料茶位于 b 区，品类为扁形绿茶、黄茶、抹茶和白茶，原料茶存在可区分的内部差异，且样品分布均匀合理。

将光谱数据代入到茶叶近红外定量模型中，其水分、水浸出物、茶多酚、咖啡碱、粗纤维、游离氨基酸的含量分布如图 3 所示。其中，a 区和 b 区原料茶的内质含量区间高度重叠，但内部变异格局不同。a 区原料茶中，游离氨基酸和茶多酚的含量是主要变异源。a 区原料茶的游离氨基酸含量（质量分数）为 0.39%~5.46%，茶多酚含量（质量分数）为 8.09%~21.6%，含量变异系数分别为 40.88%和 28.00%。b 区原料茶中，粗纤维含量（质量分数）为 7.57%~20.66%，是主要变异源，变异系数为 36.33%。此外，与 a 区原料茶相比，b 区原料茶的茶多酚含量分布更均一，含量集中在 10.98%~16.35%，变异系数仅为 12.45%，且含量区间较窄，包含于 a 区原料茶茶多酚含量分布区间。

这些内含物质的含量差异表明，a 区的卷曲形、颗粒形绿茶，红茶和茉莉花茶与 b 区的扁形绿茶、黄茶、抹茶和白茶虽然都被归于原料茶大类，但不是均一群体。光谱数据反映，不同茶类的化学成分含量既存在波动差异，也表现出一定的集群性，该结果提示在设计拼配实验时，应跨区间选择原料茶品类，合理设计拼配比例以贯通不同原料茶类群，构建具有代表性的拼配样数据集。

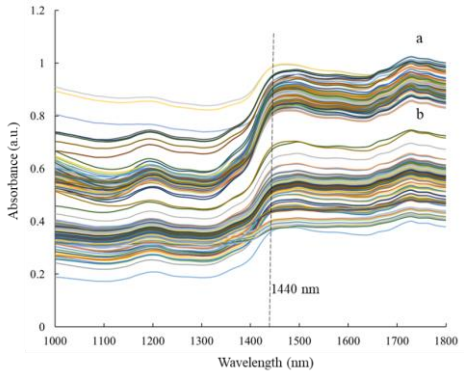


图 2 原料茶及其拼配样 NIR 光谱曲线

Fig.2 NIR spectral curves of base teas and blended samples

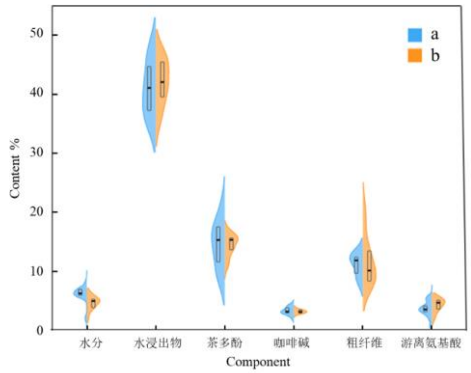


图 3 27 种原料茶的近红外检测性质含量豆荚图

Fig.3 Bean plot of near-infrared detected contents of 27 base teas

2.4 原料茶数字化拼配模型构建

高斯过程回归（Gaussian process regression，GPR）模型，能够量化预测的不确定性，为预测结果提供可靠的置信区间[29]。模型以 97 个有效拼配样本为研究对象，在贝叶斯推断框架下，根据给定的协方差函数和观测数据，推出新输入特征向量所对应预测输出的后验分布，将后验分布的均值作为点预测值，并将协方差矩阵的对角线元素作为预测值的置信区间。

训练完成后的模型在独立测试集上展现出优异的预测性能，测试集上的决定系数 R^2 为 0.95，均方误差为 0.12，平均绝对误差为 0.25。采用五折交叉验证评估模型的泛化能力与稳定性所得决定系数 R^2 为 0.92±0.04，表明模型，未出现对训练集的过拟合现象，具备良好的泛化能力与稳健性。

表 3 GPR 模型训练结果

Table 3 Results of trained GPR model

Performance metrics	Calculated value
测试 R^2	0.947 9
测试 MSE	0.115 4
测试 MAE	0.248 7
交叉验证 R^2	0.921 4±0.043 3

2.5 原料茶拼配模型预测结果

以测试集决定系数 R^2 为评价指标，由图 4 可见模型对于不同化学性质的预测效果不同，其中模型对粗纤维、水浸出物、茶多酚和水分的预测精度最高，决定系数 R^2 均高于 0.95，表明这些性质的光谱特征与拼配比例之间存在稳定映射关系，其次是游离氨基酸和咖啡碱。

由图 5 可见，模型中各成分的均方误差（Mean Squared Error，MSE）取值范围为 0.04~0.28，整体处于极低水平，进一步证实模型预测值围绕真实值的离散程度小。平均绝对误差（Mean Absolute Error，MAE）取值范围为 0.08~0.43。结合各性质的实测均值，以 MAE 除以观测值均值，计算得到相对 MAE（Relative Mean Absolute Error，rMAE），不同性质的 rMAE 均在 1.02%~4.30%间，说明预测误差相对于各成分的近红外检测实际含量较小，模型表现出高精度、低误差和强泛化能力，能够为后续的拼配使用提供可靠的定量预测支持。

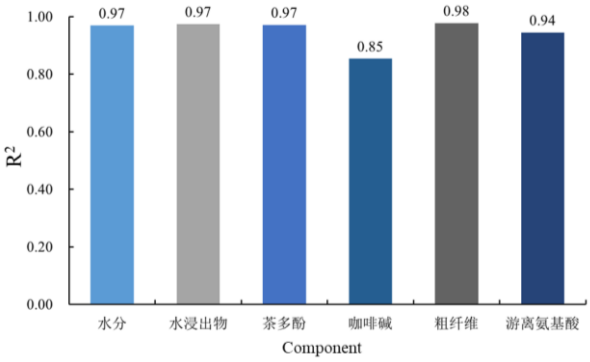
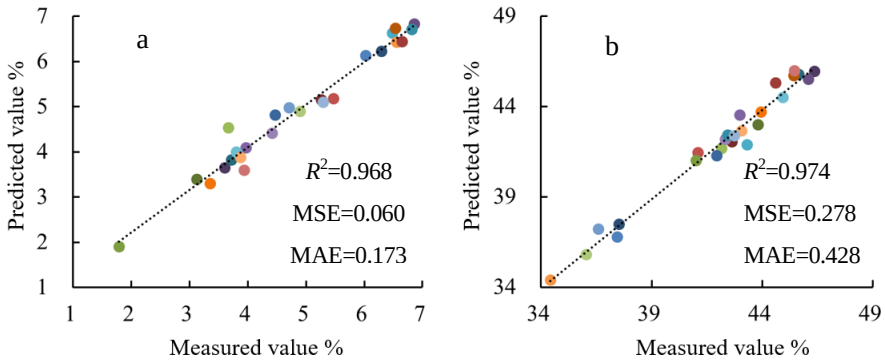


图 4 GPR 模型中茶叶品质成分决定系数 R^2 得分

Fig.4 R^2 scores of various tea quality components in GPR model



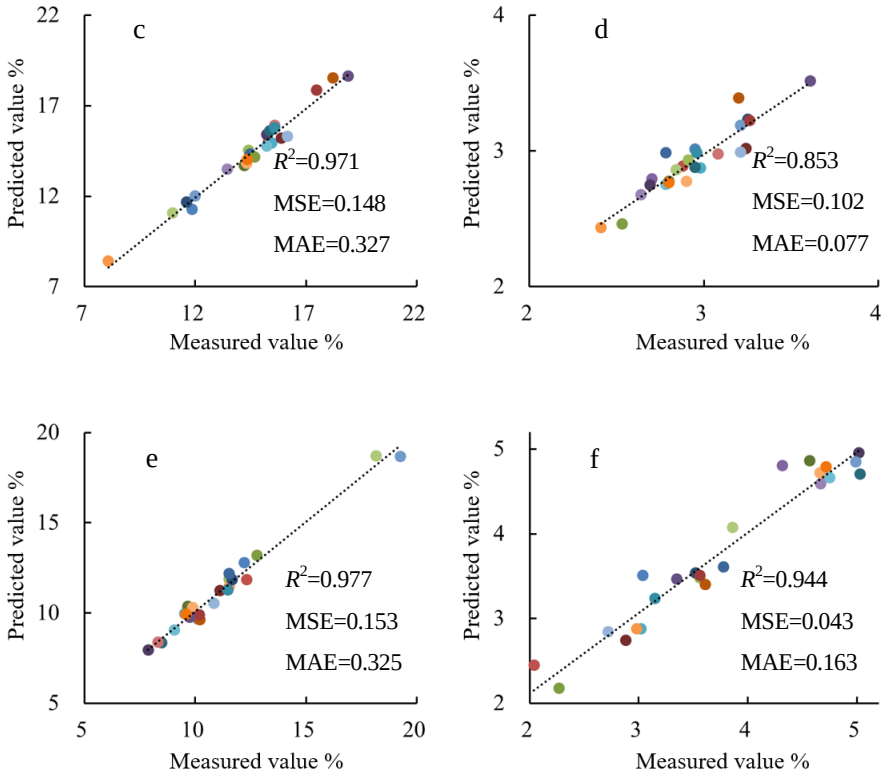


图 5 GPR 模型中茶叶品质成分模型预测值与真实值验证

Fig.5 Validation of predicted values and actual values of quality components of teas in GPR model

注:（a）水分;（b）水浸出物;（c）茶多酚;（d）咖啡碱;（e）粗纤维;（f）游离氨基酸; 纵坐标中%为质量分数 wt.%。

2.6 验证模型的拼配茶及其比例选择

选取属于不同近红外分区的原料茶样品 P1、P8、H1（a 区）和 KL2、RT2（b 区）设计拼配方案，拼配方案 TS1-TS12 选用的原料和拼配比例如表 4 所示，12 种验证用拼配方案按照其特性可以细分为同茶类/同外形拼配，同茶类/不同外形拼配，跨茶类拼配和跨多茶类拼配。原料茶及代表性拼配样 TS2、TS4、TS7、TS9 和 TS12 的外形如图 6 所示，龙井原料茶 P1、P8 外形尚匀整，有宽扁条和青壳等，黄茶原料茶 H1 外形欠匀、碎片和茎梗较多，KL2 和 RT1 外形相对匀整，跨茶类拼配样品 TS7、TS9、TS12 经拼配混匀后装填至样品盘后表面匀整度较差。

表 4 模型验证原料茶及拼配样组成比例

Table 4 Model validation of raw teas and the composition ratio of blended samples

Tea	Sample ID	Material 1	Ratio 1	Material 2	Ratio 2	Material 3	Ratio 3	Material 4	Ratio 4	Material 5	Ratio 5	Type/Shape
Material Teas	P1	P1	1	-	-	-	-	-	-	-	-	绿茶/扁形/龙井
	P8	-	-	P8	1	-	-	-	-	-	-	绿茶/扁形/龙井
	KL2	-	-	-	-	KL2	1	-	-	-	-	绿茶/颗粒形
	RT1	-	-	-	-	-	-	RT1	1	-	-	红茶/条形
	H1	-	-	-	-	-	-	-	-	H1	1	黄茶/片状
Blend Teas	TS1	P1	0.2	P8	0.8	KL2	-	RT1	-	H1	-	同茶类/同外形
	TS2		0.8		0.2		-					
	TS3		0.5		0.5		-					

TS4	P1	0.5	P8	-	KL2	0.5	RT1			同茶类/不同外形
TS5		-		0.5		0.5				
TS6		0.25		0.25		0.5				
TS7		0.5						0.5	H1	-	跨茶类/两类
TS8		0.5						-		0.5	
TS9		-						0.5		0.5	
TS10		0.33						0.33		0.34	跨茶类/三类
TS11		0.3						0.5		0.2	
TS12		0.5						0.3		0.2	

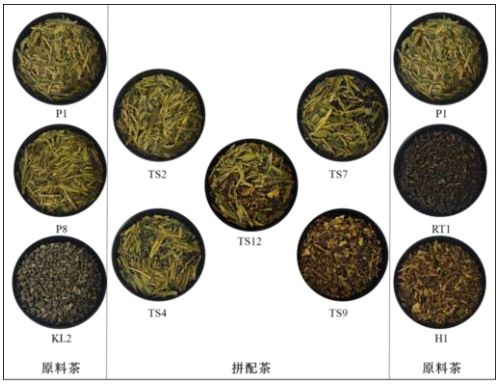


图 6 不同品类和外形的原料茶及其拼配茶示例

Fig.6 Examples of raw teas and blended teas of different categories and shapes

2.7 高斯过程回归模型验证结果

12 个拼配样品的 GPR 模型预测值和近红外光谱实测值数据如表 5 所示，其中 GPR 模型预测值以预测均值±95%置信区间（两倍半宽）表示，近红外光谱实测值以平均值±标准差表示。不同成分的预测不确定性在 GPR 模型中存在差异，其中咖啡碱的预测置信区间半宽（以质量分数计）最小，为 0.22%~0.24%，对应相对波动范围（半宽/预测均值）为 6.83%~7.60%；水浸出物的预测置信区间半宽（以质量分数计）最大，为 1.40%~1.60%，但由于水浸出物含量高，因此经计算得到的相对波动范围仅为 3.20%~3.68%。整体而言，除游离氨基酸和粗纤维外，其余成分的相对波动范围均在 10%以内，表明 GPR 模型对各品质成分均能提供具有统计学意义的概率化预测。此外，12 个拼配样品的近红外实测结果显示，其测量值的相对波动范围（标准差/均值）普遍小于 GPR 模型给出的预测置信区间，除水分 1.11%~9.07%和粗纤维 1.12%~7.32%外，其余成分的相对波动均低于 5%，表明近红外实测具有较高的重复测量精密度。

尽管 GPR 模型提供的预测置信区间及其相对波动范围在数值上普遍大于近红外重复测量的标准差，但如图 7a 所示，GPR 模型的点预测值在 12 个拼配样品中的变化趋势与近红外实测值一致。除水浸出物的 GPR 预测值整体略高于近红外检测值外，茶多酚、粗纤维、游离氨基酸等成分的预测值与近红外检测值在各样品中的分布趋势吻合，表明 GPR 模型对拼配茶样中六种品质成分的定量预测准确性高，系统偏差处于可控范围内。

四组数据的噪声指标强度结果如图 7b~e 所示。同茶类/同外形拼配方案中，茶叶原料的理化性质相近，光谱特征差异较小，如图 7b 所示，各成分的噪声指标分布范围为 1.17~35.83，最大值 35.83 出现于 TS3 样本的粗纤维含量性质中，其次是游离氨基酸 20.74 出现在 TS2 样品中。除粗纤维外，其余指标多集中在 1~10 区间，且整体均值为 5.09，中位数为 3.89。表明在原料外形相似的同一茶类拼配场景下，GPR 模型对多数成分能给出较为紧凑的置信区间，但对粗纤维和游离氨基酸的预测偏差相对较大，导致噪声指标升高。

同茶类/不同外形拼配方案中，外形差异导致光谱散射效应变化，如图 7c 所示，该组拼配样品的噪声指标范围为 0.80~17.53，其中茶多酚在 TS5 样本中为 17.53，粗纤维在 TS5 样本中为 13.00，游离氨基酸在 TS6 样本中为 11.86；水分和咖啡碱的指标均低于 3.6，整体均值为 3.52，中位数为 2.79，与 b 组相比，极端高值有所下降。

跨茶类拼配方案中，如图 7d 所示，T9 样本中的游离氨基酸预测偏差达到 64.18，但剔除极端值后，其余指标

的均值为 3.38。由于 GPR 模型给出的置信区间半宽并未显著增加，表明跨茶类拼配中，游离氨基酸噪声指标异常升高原因是其对游离氨基酸的预测偏差小，同时 GPR 模型对其他成分仍保持较高的可靠性。

跨多茶类拼配方案中，三种茶类混合时，光谱重叠作用较为复杂，如图 7e 所示，该组拼配样品的噪声指标范围为 1.08~11.12，其中水浸出物在 TS10 样本中为 11.12，咖啡碱在 TS11 样本中为 7.17，其余性质指标大多低于 6.0。全组样品整体均值为 2.79，中位数为 2.50，是四组中最低的一组。

为进一步验证模型的准确性和适用性，选取 TS2、TS5、8、TS11 作为同茶类/同外形、同茶类/不同外形、跨茶类/两类、跨茶类/三类的代表性样品，通过国标法检验得到其内含成分真实值如表 6 所示。

TS2、TS5、TS8、TS11 的 GPR 模型预测结果、近红外实测结果及国标法测定结果的分项绘图比较结果如图 8 所示。4 个拼配茶中，除 TS2 样品的咖啡碱国标法测定含量，TS8 样品的水分、水浸出物国标法测定含量微高出 GPR 预测模型提供的置信区间上限外，其他拼配茶样品的水分、水浸出物、咖啡碱、粗纤维、游离氨基酸含量的国标法检测值均位于 GPR 模型的预测范围区间，表明 GPR 预测模型在以上内含物质分项中输出的预测结果与国标法检测值吻合度高。且由图 8b 可见，在水浸出物一项中，GPR 模型预测中心值与国标法测定含量基本吻合，展示出高于 NIRS 实测的准确度。但由图 8c 可见，用于验证的拼配茶的茶多酚真实含量均高于 GPR 模型和近红外预测值，其国标法检测值平均高于预测值约 3%（质量分数），这是背离模型的准确性的结果，但其茶多酚和含量分布又与近红外实测和 GPR 模型预测结果基本一致。这表明在本试验中，近红外模型 GPR 模型在预测茶多酚含量集中在 16%~20%的原料茶时存在系统性误差，该项不足可以通过后续统一校正模型消除。

综合分析模型维度上的噪声评价指标值和实验维度上的拼配茶的 NIRS 实测、GPR 模型预测和国标法测定值比较分析结果可见，尽管跨多茶类的拼配最为复杂，但其噪声指标低于前三种拼配方案，且其内含物质含量的国标法检测值基本处于 GPR 模型给出的预测置信区间中，这体现了 GPR 模型预测分布自带不确定性估计，可随数据累积动态更新模型以适应输入空间的变化的核心优势[29]，在跨茶类的复杂拼配场景中表现出良好的适用性。与 Ling 等[31]的研究相比，本模型拓宽了拼配茶的原料来源，在未显著影响预测偏差的基础上，将噪声指标维持较低水平，且加入了国标法检测结果，通过数据比对验证了 GPR 模型的准确性，在复杂拼配场景下 GPR 模型自适应的不确定性估计能力将有助于满足新业态下的复杂化拼配需求[28]。

表 5 拼配样品的 GPR 模型预测与近红外检测结果数据表

Table 5 Data of GPR model predictions and near-infrared detection results for blended samples

Sample ID	Moisture/%	Water extract/%	Tea polyphenols/%	Caffeine/%	Crude fiber/%	Free amino acids/%	Data origin
TS1	3.83±0.64	45.39±1.60	15.80±1.28	3.01±0.24	8.52±1.50	4.65±0.78	
TS2	4.17±0.64	45.37±1.60	15.61±1.28	2.87±0.24	8.30±1.50	4.76±0.78	
TS3	3.79±0.56	45.47±1.46	15.54±1.16	2.94±0.22	8.48±1.30	4.81±0.72	
TS4	4.74±0.56	46.07±1.44	16.79±1.16	3.45±0.22	8.14±1.28	4.97±0.70	
TS5	4.64±0.56	45.97±1.46	16.93±1.16	3.53±0.22	8.25±1.30	4.82±0.72	GPR 模型
TS6	4.62±0.60	46.42±1.50	16.93±1.18	3.45±0.22	8.12±1.38	4.89±0.72	预测值，取
TS7	5.24±0.54	41.63±1.40	13.93±1.12	3.29±0.22	10.02±1.24	4.39±0.68	95%置信区
TS8	4.41±0.54	43.08±1.40	15.28±1.12	3.02±0.22	11.07±1.22	3.78±0.70	间
TS9	5.92±0.54	39.17±1.44	13.79±1.14	3.47±0.22	12.77±1.26	3.21±0.70	
TS10	5.04±0.56	41.73±1.46	14.47±1.18	3.19±0.22	11.27±1.30	3.63±0.72	
TS11	5.39±0.56	41.02±1.46	14.05±1.16	3.29±0.22	10.87±1.30	3.75±0.72	
TS12	4.94±0.56	42.44±1.46	14.59±1.16	3.12±0.22	10.24±1.28	3.96±0.72	
相对波动	11.42~13.54	3.20~3.68	7.33~8.32	6.83~7.60	12.63~15.47	16.27~18.08	
TS1	3.90±0.05	43.58±0.59	15.69±0.17	3.12±0.05	8.07±0.57	5.25±0.25	
TS2	3.63±0.12	43.41±0.23	15.49±0.20	2.82±0.10	8.57±0.21	4.72±0.13	
TS3	3.60±0.13	43.15±0.40	15.26±0.13	2.75±0.04	8.45±0.26	4.99±0.10	拼配样品近
TS4	4.40±0.23	43.97±0.94	16.64±0.58	3.18±0.11	8.60±0.34	5.08±0.11	红外检测实
TS5	4.48±0.14	43.81±0.16	16.86±0.32	3.65±0.01	8.15±0.12	5.21±0.25	测值
TS6	4.20±0.14	43.12±0.46	15.89±0.44	3.29±0.09	8.48±0.23	4.95±0.02	

TS7	4.68±0.27	41.53±0.52	14.34±0.44	3.00±0.09	9.83±0.27	4.90±0.09
TS8	3.98±0.07	42.17±0.15	15.37±0.05	2.82±0.06	10.12±0.11	4.13±0.05
TS9	5.53±0.10	38.46±0.21	14.24±0.08	3.45±0.05	13.04±0.31	3.20±0.04
TS10	4.95±0.10	39.14±0.44	14.00±0.21	3.08±0.05	11.74±0.38	3.93±0.17
TS11	5.61±0.07	39.27±0.27	14.24±0.33	3.32±0.02	12.07±0.15	3.61±0.16
TS12	5.28±0.41	40.28±0.83	14.31±0.38	3.18±0.07	10.91±0.72	3.68±0.20
相对波动	1.11~9.07	0.36~2.25	0.33~3.82	0.32~3.50	1.12~7.32	0.45~5.59

注：水分、水浸出物、茶多酚、咖啡碱、粗纤维、游离氨基酸含量%为质量分数 wt.%，相对波动为%。

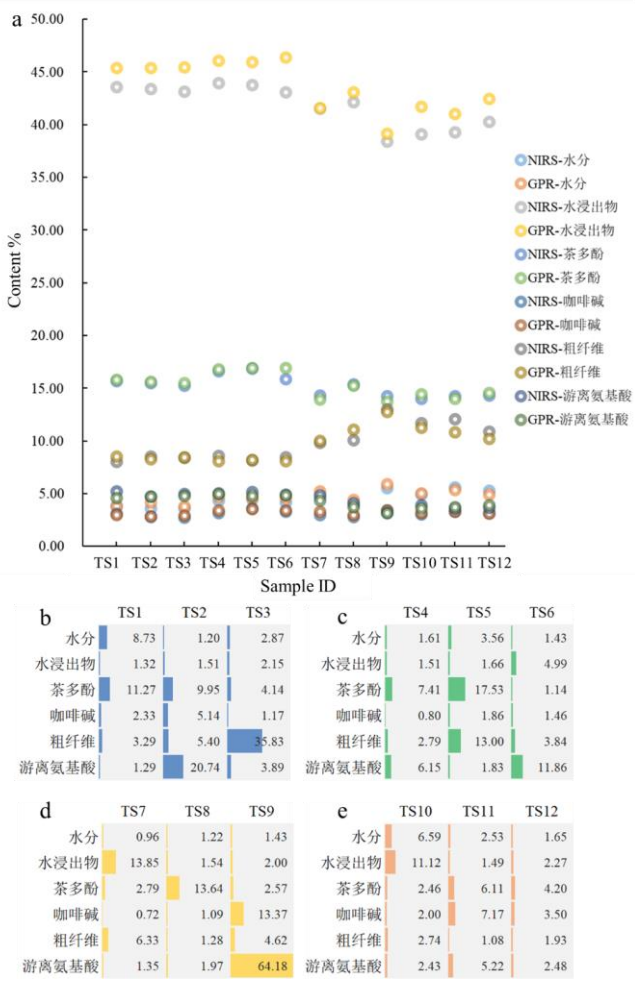


图 7 GPR 模型预测结果和近红外实测结果及不同拼配组的噪声比较

Fig.7 Comparison of GPR model prediction and near-infrared measured results with noise of different blending groups

注：（a）GPR 模型预测和近红外实测结果;（b）同茶类/同外形拼配组噪声指标;（c）同茶类/不同外形拼配组噪声指标;（d）跨茶类拼配组噪声指标;（e）跨多茶类拼配组噪声指标。

表 6 拼配样品 TS2、TS5、TS8、TS11 的国标法检测结果

Table 6 Tested results of blended samples TS2、TS5、TS8、TS11 based on national standard methods

Sample ID	Moisture/%	Water extract/%	Tea polyphenols/%	Caffeine/%	Crude fiber/%	Free amino acids/%	Data origin
TS2	3.98±0.01	46.15±0.07	18.95±0.64	3.30±0.01	9.66±0.01	4.90±0.01	GB 5009.3-2016; GB/T 8305-2013; GB/T 8313-2018; GB/T 8312-2013; GB/T 8310-2013;
TS5	4.66±0.08	46.1±0.14	20.45±0.49	3.65±0.07	8.58±0.01	4.55±0.21	
TS8	5.10±0.08	44.85±0.21	18.55±0.35	3.10±0.01	12.00±0.01	3.60±0.01	

TS11	5.89±0.05	40.8±0.01	15.85±0.21	3.30±0.01	11.90±0.01	3.40±0.01	GB/T 8314-2013

注：水分、水浸出物、茶多酚、咖啡碱、粗纤维、游离氨基酸含量%为质量分数 wt.%。

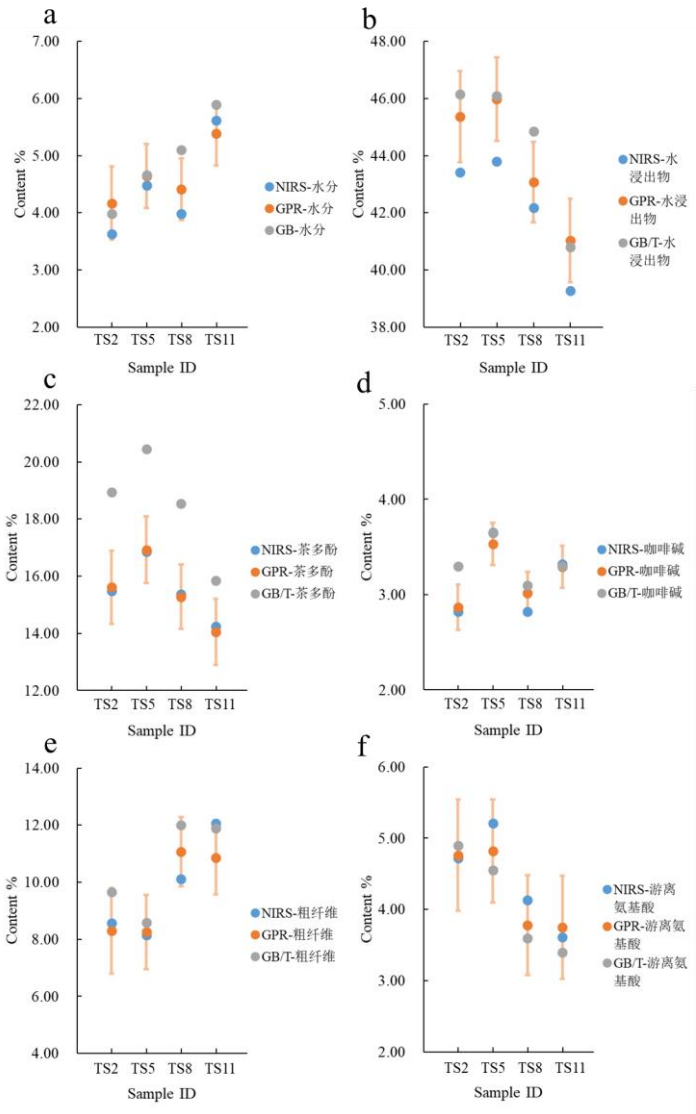


图 8 GPR 模型预测结果、近红外实测结果及国标法测定结果的比较

Fig.8 Comparison of GPR model predictions、near-infrared measured results and tested results based on national standards

注:（a）水分;（b）水浸出物;（c）茶多酚;（d）咖啡碱;（e）粗纤维;（f）游离氨基酸; 纵坐标中%为质量分数 wt.%。

3 结论

本研究建立了基于近红外光谱快速检测与高斯过程回归概率预测的拼配茶品质数字化模型。茶叶近红外多组分定量模型可为拼配模型提供涵盖 6 项品质指标的可靠成分数据，GPR 模型预测的拼配茶品质成分含量与近红外实测结果相比，测试集 R^2 达 0.95，并能以 95%置信区间量化输出预测不确定性。在四组复杂度由易到难的原料茶拼配组中，GPR 模型的预测噪声呈现自适应变化，在同茶类拼配中均值小于 5.09，在跨茶类中最低预测噪声为 2.79。进一步的，各拼配组中的代表性拼配茶 TS2、TS5、TS8、TS11 内含物质的国标法检测值与 GPR 模型预测值的比较分析结果显示，除在茶多酚含量预测上 GPR 模型存在系统误差外，该模型具有准确预测输出原料茶水分、水浸出物、咖啡碱、粗纤维、游离氨基酸含量的综合能力，展示出 GPR 模型的输出结果带有自适应预测置信区间的性能优势。

需要指出的是，本研究未以同一数据集为基础，实施其他算法模型如偏最小二乘回归算法、人工神经网络算法等与基于近红外快检和高斯过程回归的单-多茶类数字化拼配模型的性能比较。但拼配验证实例佐证了该模型在

应用层面的便利性和稳健性，模型初步具备循环应用“预设配方——GPR 模型输出配方内含成分——人工实施、调整及验证——确认配方”流程的能力。后续研究将校正模型在茶多酚含量预测输出中的系统误差、并开展系统性的模型能力比较，以明确 GPR 模型在不同茶叶拼配场景中的使用优势与适用边界，并将优化后的 GPR 模型应用于现实中形成拼配实施例。综合后续措施，实现将大量试拼的时间精力成本压缩在模型的数据层，从而在现实中解放专业人员的茶叶拼配生产力，进一步推动茶叶拼配从“经验主导”向“数据驱动”的智能化转型。

参考文献

[1] YIN J F, CHEN J X, HU Y R, et al. Progress and trends on innovation and application of tea blending technology [J]. Journal of Tea Communication, 2025, 52(2): 137-144.

[2] 许姗姗,李姝寰,宁井铭.茶叶加工中拼配技术的研究进展[J].中国茶叶加工,2020,1:34-37+59.

[3] 郑际雄.外销滇红工夫茶精制加工与拼配技术[J].中国茶叶加工,2017,Z2:57-60.

[4] 尹军峰.新式茶饮业现状与发展趋势[J].中国茶叶, 2021,43(8):1-6.

[5] 阎鹏翔,彭康,陈富桥,等.年轻群体新茶饮购买意愿提升实证研究——基于顾客感知价值视角[J].茶叶科学,2024,44(6):1023-1036.

[6] 陈丁丁,陈根生,张颖彬,等.茶叶数字化拼配技术及智能化应用系统研究进展[J].中国茶叶,2026,48(1):23-28.

[7] GUO J, ZHANG K, ADADE S Y S S, et al. Tea grading, blending, and matching based on computer vision and deep learning [J]. Journal of the Science of Food and Agriculture, 2025, 105(6): 3239-3251.

[8] WANG S P, ALTANER C, FENG L, et al. A review: Integration of NIRS and chemometric methods for tea quality control—principles, spectral preprocessing methods, machine learning algorithms, research progress, and future directions [J]. Food Research International, 2025, 205: 115870.

[9] DANKOWSKA A, WŁODARSKA K, MANDAL A, et al. Spectroscopic techniques for quality assessment of tea and coffee [M]. Advanced Spectroscopic Techniques for Food Quality. Cambridge: Royal Society of Chemistry, 2022.

[10] TURGUT S S, KÜÇÜKÖNER E, KARACABEY E. TeaPot: A chemometric tool for tea blend recipe estimation [J]. Applied Food Research, 2021, 1(1): 100006.

[11] FERNANDES J M, ROUX K, RENARD C M G C, et al. Predicting the composition of multiple soybean varieties from whole and ground seeds using Fourier transform near-infrared spectroscopy (FT-NIRS) and machine learning [J]. Food Chemistry, 2025, 479: 143856.

[12] NURROHMAN A M, ROHMAN A, WIDODO S, et al. Prediction accuracy of near infrared spectroscopy coupled with adaptive machine learning methods for simultaneous determination of chlorogenic acid and caffeine on intact coffee beans [J]. Journal of Agriculture and Food Research, 2024, 16: 101173.

[13] GRASSI S, GIRAUDO A, NOVARA C, et al. Monitoring chemical changes of coffee beans during roasting using real-time NIR spectroscopy and chemometrics [J]. Food Analytical Methods, 2023, 16(5): 947-960.

[14] 唐雪平,刘淑岚,郭小华,等.近红外光谱技术在茶叶品质分析中的研究进展[J].轻工科技,2020,36(5):12-15.

[15] 梁建华,郭嘉明,乔小燕.近红外光谱技术在茶叶生产质控与品质评价中的研究进展[J].食品安全质量检测学报,2024,15(8):254-261.

[16] 李文萃,周新奇,范起业,等.便携式近红外茶叶品质快速检测仪设计与试验[J].现代食品科技,2021,37(5):303-309.

[17] 李春霖.基于化学计量学和近红外光谱技术的龙井茶感官及化学品质评价研究[D].杭州:浙江大学,2020.

[18] 刘蕾,罗文文,龚淑英,等.采用近红外光谱技术定量分析绿茶中的主要呈味物质[J].中国食品学报,2008,8(6):109-115.

[19] WU T, ZHOU L, ZHAO Y, et al. Applications of deep learning in tea quality monitoring: a review [J]. Artificial Intelligence Review, 2025, 58(11): 342.

[20] 高钰敏,张雨茹,丁小莉,等.近红外光谱技术在茶叶品质评价中的研究[J].茶叶通讯,2025,52(2):239-244.

[21] XIA H, CHEN W, HU D, et al. Rapid discrimination of quality grade of black tea based on near-infrared spectroscopy (NIRS), electronic nose (E-nose) and data fusion [J]. Food Chemistry, 2024, 440: 138242.

[22] 孙晓荣,胡毅然,刘翠玲,等.基于近红外光谱技术的茶叶新旧鉴别及产地溯源研究[J].食品安全质量检测学报,2023,14(5):53-59.

[23] 汪小中.基于近红外光谱和计算机视觉技术的茶叶拼配与品质评价方法研究[D].合肥:安徽农业大学,2025.

[24] 李航.机器学习方法:第 2 版.[M].北京:清华大学出版社,2025.

[25] 张叶.面向茶叶拼配过程的机器学习算法研究[D].合肥:安徽农业大学,2023.

[26] 宁井铭,李姝寰,王玉洁,等.基于高光谱成像技术的工夫红茶数字化拼配[J].食品科学,2019,40(4):318-323.

[27] 宋彦,汪小中,赵磊,等.基于近红外光谱技术的眉茶拼配比例预测方法[J].农业工程学报,2022,38(2):307-315.

[28] XIA Z J, ZHOU Q Y, YANG S E, et al. Optimization strategy for black tea digital blending by fusing image and spectral information [J]. Food Research International, 2025, 202: 115870.

[29] SCHULZ E, SPEEKENBRINK M, KRAUSE A. A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions [J]. Journal of Mathematical Psychology, 2018, 85: 1-16.

[30] ZHAO J Y, CHEN Z G, LIU S, et al. Research on rapid non-destructive detection of tannin and protein content in sorghum based on multi-output Gaussian process[J]. Journal of Food Composition and Analysis, 2025, 141: 107326.

[31] LING C J, HUANG L Z, BIAN Y, et al. Taste evaluation by NIR spectroscopy and blending strategy of black tea based on GDE3 algorithm [J]. Infrared Physics & Technology, 2025, 146: 105739.

[32] REVILLA I, HERNÁNDEZ JIMÉNEZ M, MART ÍNEZ-MART Ín I, et al. The potential use of near infrared spectroscopy (NIRS) to determine the heavy metals and the percentage of blends in tea [J]. Foods, 2024, 13(3): 450.

[33] ZHANG W K, SANAEIFAR A, JI X S, et al. Data-driven optimization of nitrogen fertilization and quality sensing across tea bud varieties using near-infrared spectroscopy and deep learning [J]. Computers and Electronics in Agriculture, 2024, 224: 109071.

[34] YIN H, FANG E S. Clustering consumer adoption behavior with respect to innovative tea products in the Chinese market [J]. Beverages, 2024, 10(3): 72.

[35] 李卫军,覃鸿,于丽娜,等.近红外光谱定性分析原理、技术及应用[M].北京:科学出版社,2020.