Supplementary Information (SI) for Analytical Methods.
This journal is © The Royal Society of Chemistry 2025

Supplementary materials

Rapid Detection of Colorants in Black Tea Using Mid- and Short-wave Near Infrared Spectroscopy

Zhengfei Luo,[abc] *Yunfeng Chai,[b] Guohua Zhao,[d] Dongling Qiao,[a] Fayin Ye,[a] Lin Lei,[a] and *Jia Chen[a]

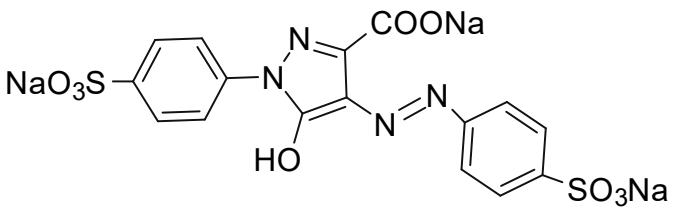
[a.] *College of Food Science / Chongqing Key Laboratory of Speciality Food Co-Built by Sichuan and Chongqing, Southwest University, Chongqing 400715, China. E-mail: swaumyh@swu.edu.cn*

[b.] *Tea Research Institute, Chinese Academy of Agricultural Sciences, Key Laboratory of Tea Quality and Safety Control, Ministry of Agriculture and Rural Affairs, Hangzhou 310008, China. E-mail: chaiyunfeng@tricaas.com*

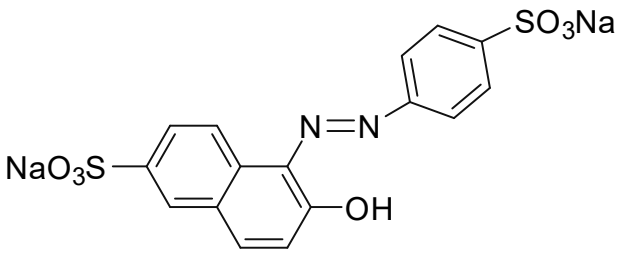
[c.] *School of Biotechnology and Engineering, West Yunnan University, Lincang 677000, China. E-mail: luopyh@126.com*

[d.] *College of Life Science, Sichuan Normal University, Chengdu 610101, China. E-mail: zhaoguohua1971@163.com*

(a)



(b)



(c)

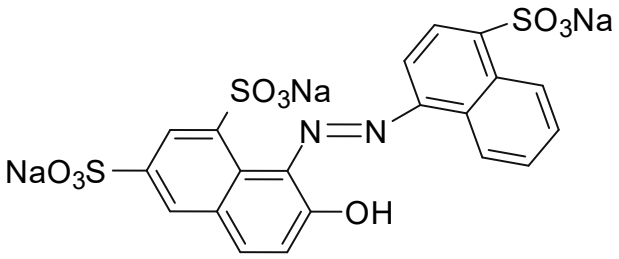


Figure S1. Chemical structure of (a) Tartrazine, (b) Sunset yellow and (c) Ponceau 4R.

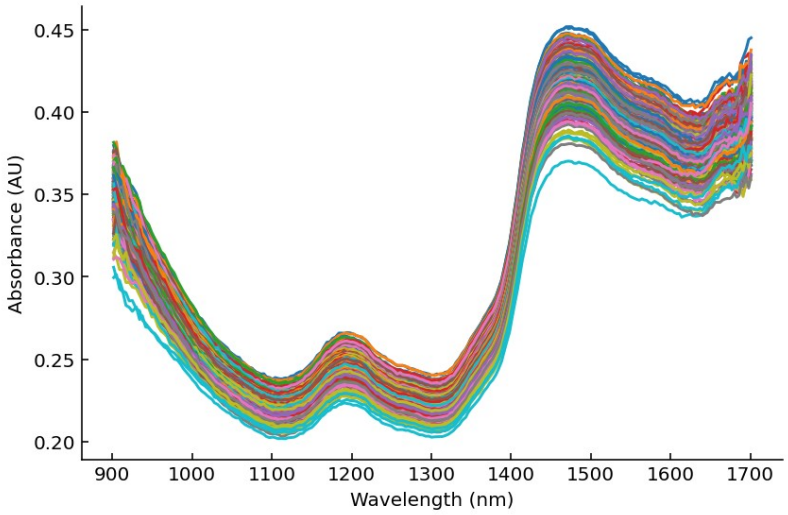


Figure S2. The raw spectra of all samples

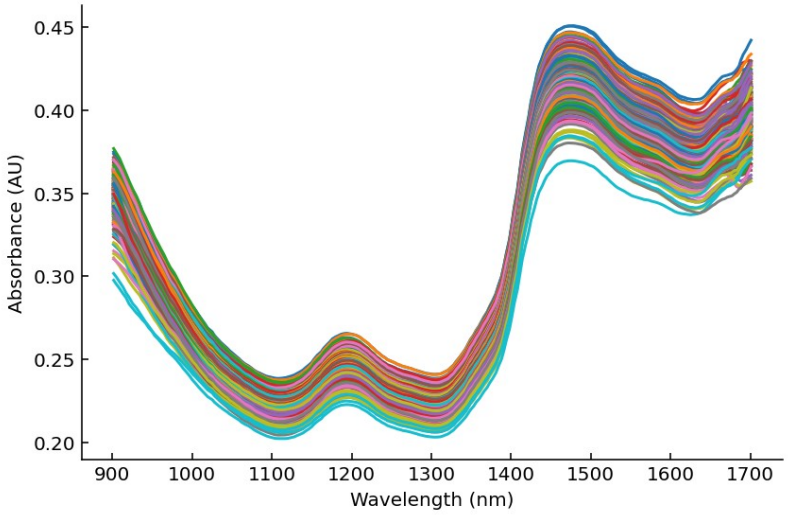


Figure S3. The spectra preprocessed by MA

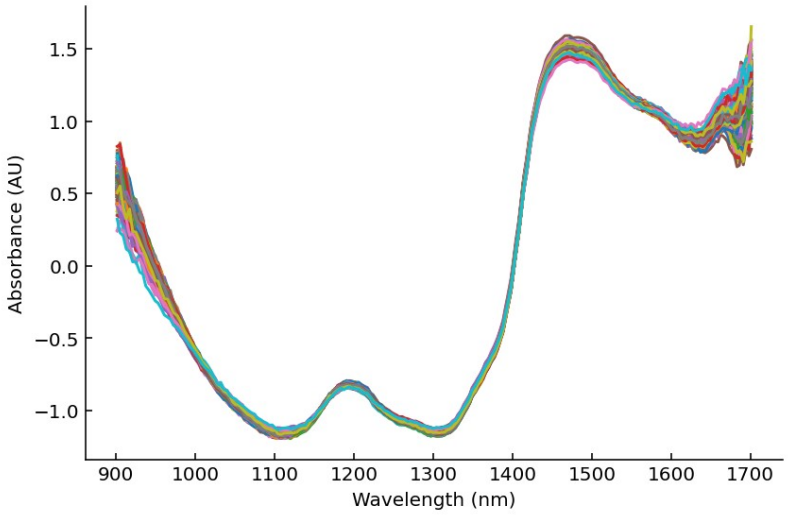


Figure S4. The spectra preprocessed by SNV

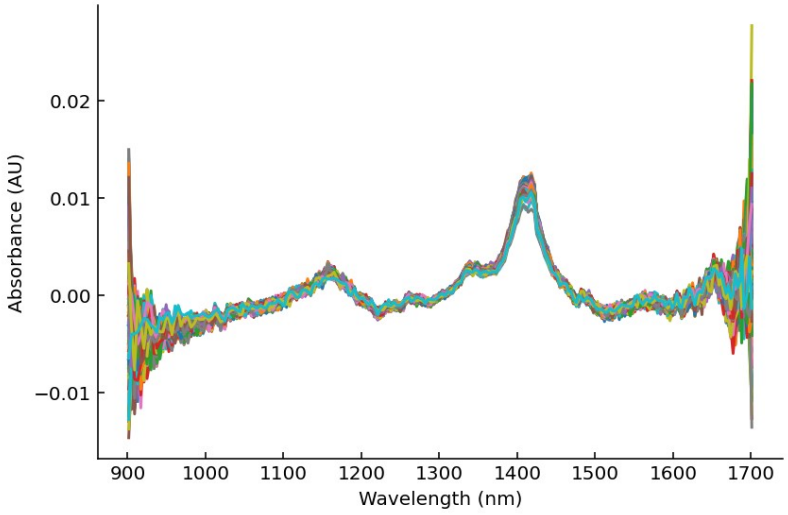


Figure S5. The spectra preprocessed by 1st Der

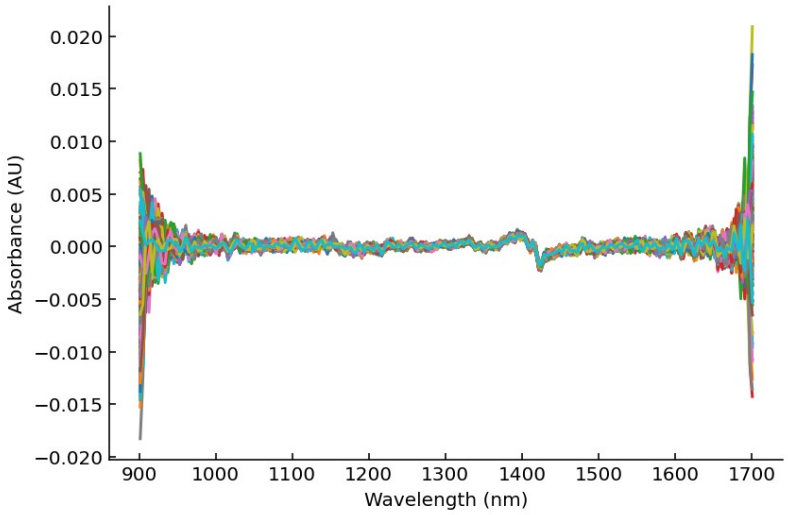


Figure S6. The spectra preprocessed by 2nd Der

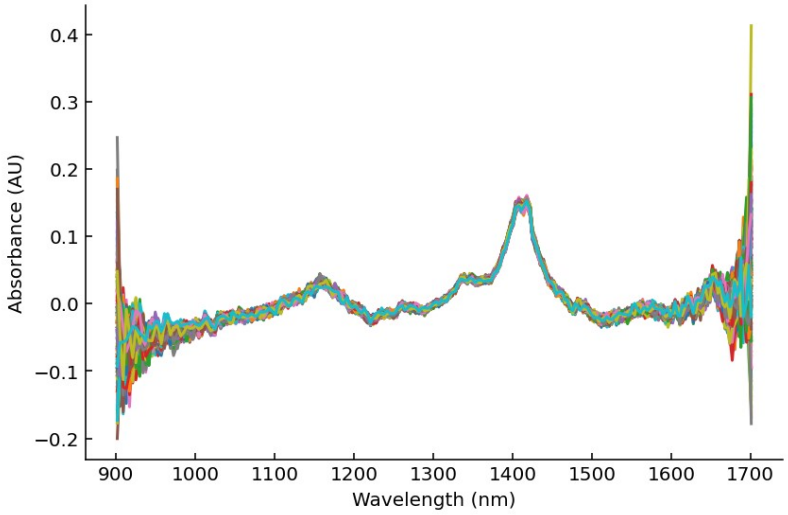


Figure S7. The spectra preprocessed by SNV+1st Der

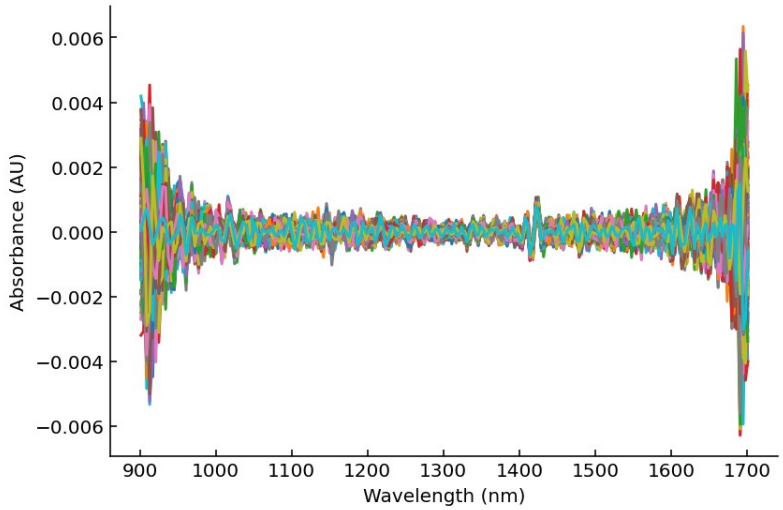
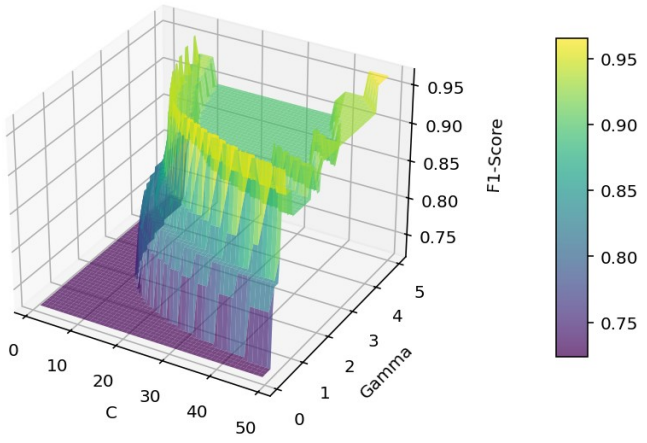


Figure S8. The spectra preprocessed by SNV+2st Der



Best C=9, gamma=4.55

Figure S9. Grid search results for SVM discriminative model of tartrazine

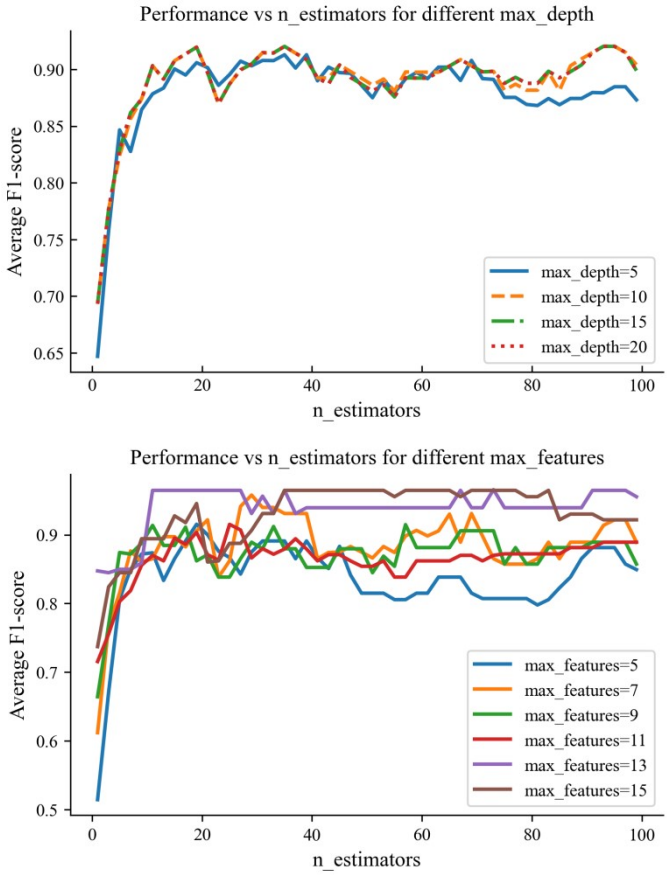


Figure S10. Hyperparameter tuning for the RF discriminative model of tartrazine

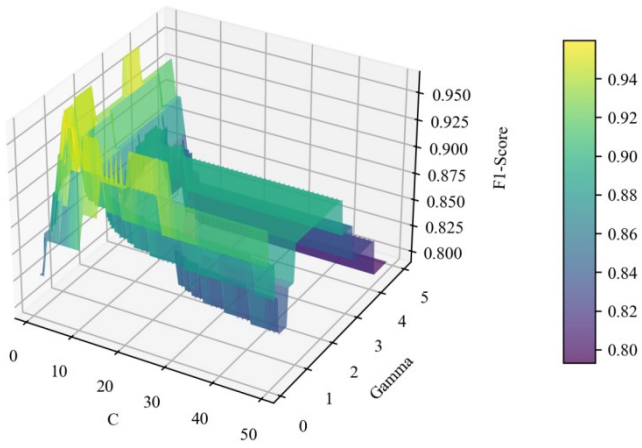


Figure S11. Grid search results for SVM discriminative model of sunset yellow

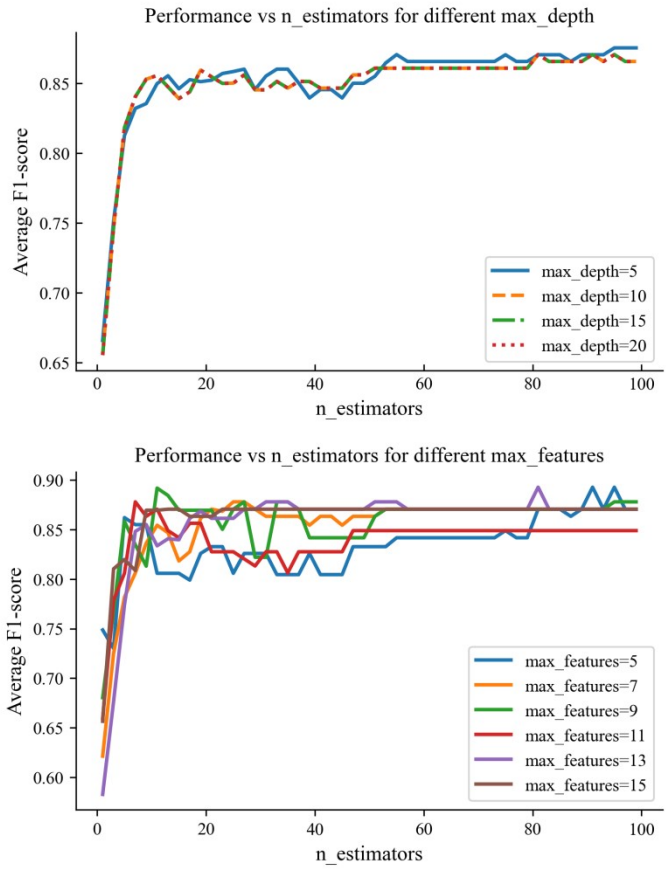


Figure S12. Hyperparameter tuning for the RF discriminative model of sunset yellow

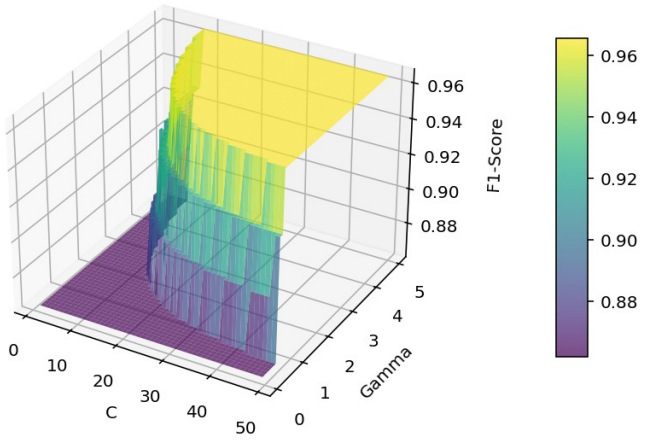


Figure S13. Grid search results for SVM discriminative model of ponceau 4R

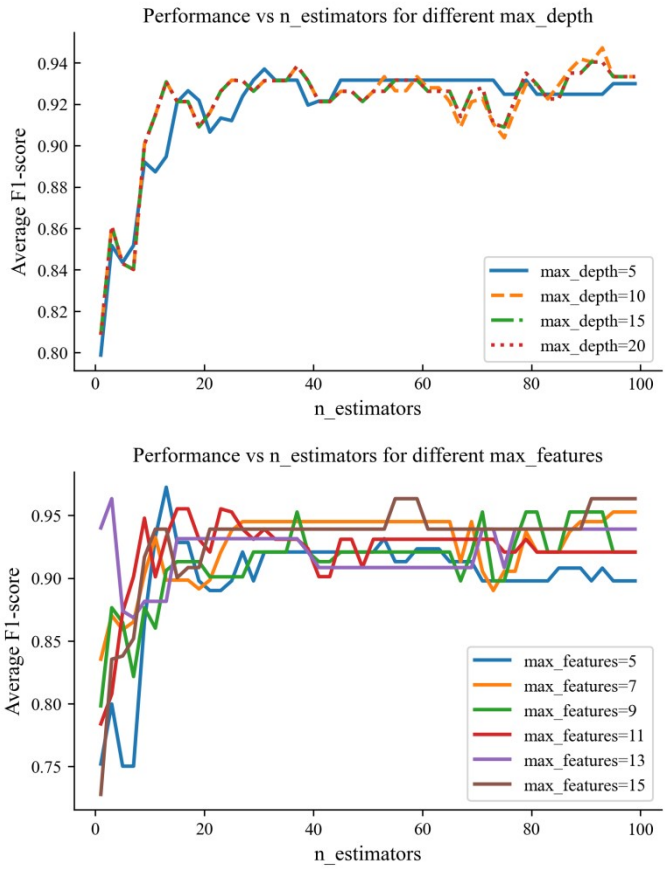


Figure S14. Hyperparameter tuning for the RF discriminative model of ponceau 4R

Table S1. The concentrations of colorants in the adulterated tea samples

Sample set	Sample size	Amount of colorant added (g/kg)			
		Min	Max	Mean	Std
Black tea (tartrazine added)	125	0.052	0.523	0.287	0.151
Black tea (sunset yellow added)	125	0.053	0.531	0.293	0.153
Black tea (ponceau 4R added)	125	0.053	0.535	0.295	0.155

Table S2 Validation results of the quantitative model

Colorants	Preprocessing method	SVM			RF		
		R^2	RMSEP	RPD	R^2	RMSEP	RPD
Tartrazine	Raw	0.4728	0.1349	1.14	0.4141	0.1421	1.08
	SNV	0.2674	0.1591	0.97	0.4662	0.1356	1.13
	MA	0.4252	0.1407	1.09	0.4619	0.1360	1.13
	1st Der	0.1538	0.1709	0.90	0.0093	0.1865	0.82
	2nd Der	0.1632	0.2002	0.77	0.1011	0.1949	0.79
	SNV+1st Der	0.0555	0.1806	0.85	0.0329	0.1887	0.81
	SNV+2nd Der	0.0997	0.1947	0.79	0.0907	0.1939	0.79
Sunset yellow	Raw	0.6149	0.1196	1.31	0.5821	0.1249	1.25
	SNV	0.3580	0.1546	1.01	0.4878	0.1382	1.13
	MA	0.6741	0.1100	1.42	0.6187	0.1192	1.31
	1st Der	0.3665	0.1536	1.02	0.4079	0.1487	1.05
	2nd Der	0.0455	0.1887	0.83	0.0015	0.1929	0.81
	SNV+1st Der	0.2962	0.1619	0.97	0.3832	0.1517	1.03
	SNV+2nd Der	0.0302	0.1900	0.82	0.0138	0.1916	0.82
Ponceau 4R	Raw	0.5287	0.1396	1.13	0.3503	0.1637	0.96
	SNV	0.4314	0.1533	1.03	0.5030	0.1432	1.10
	MA	0.5679	0.1334	1.18	0.3958	0.1578	1.00
	1st Der	0.3200	0.1676	0.94	0.1646	0.1857	0.85
	2nd Der	0.2671	0.2287	0.69	0.0145	0.2017	0.78
	SNV+1st Der	0.2233	0.1792	0.88	0.1308	0.1895	0.83
	SNV+2nd Der	0.0376	0.2069	0.76	0.0224	0.2010	0.79

*A model with RPD < 1.5 cannot be used for prediction, while a model with 1.5 < RPD < 2 is only suitable for rough evaluation, and a model with RPD > 2 is considered reliable.

SNV: standard normal variate

MA: moving average

1st Der: first derivative

2nd Der: second derivative

R^2: coefficient of determination for prediction set.

RMSEP: root mean square error for prediction set

RPD: ratio of prediction to deviation

$$R^2 = 1 - \frac{\sum(\hat{y}_i - y_i)}{\sum(\hat{y}_i - \bar{y})}$$

$$RMSEP = \sqrt{\frac{\sum(\hat{y}_i - y_i)^2}{n}}$$

$$RPD = \frac{SD}{RMSEP}$$

where $\hat{y}_i$ denotes the predicted value of the i-th sample, y_i denotes its corresponding actual value, and $\bar{y}$ denotes the average of the actual values of all samples in the validation set. In addition, n refers to the number of samples in the validation set.