

Table S1 Dyes the construction of CSA.

Name	Abbreviation	Code Name
2,3,7,8,12,13,17,18-Octaethyl-21H,23H-porphine	OEP	TPP1
5,10,15,20-Tetrakis(4-methoxyphenyl)-21H,23H-porphine iron(III) chloride	CH ₃ OTPPFeCl	TPP2
5,10,15,20-Tetrakis(4-methoxyphenyl)-21H,23H-porphine	CH ₃ OTPP	TPP3
5,10,15,20-Tetrakis(4-methoxyphenyl)-21H,23H-porphine cobalt(II)	CH ₃ OTPPCo	TPP4
5,10,15,20-Tetraphenyl-21H,23H-porphine	TPP	TPP5
5,10,15,20-Tetraphenyl-21H,23H-porphine zinc(II)	TPPZn	TPP6
5,10,15,20-Tetraphenyl-21H,23H-porphine copper(II)	TPPCu	TPP7
5,10,15,20-Tetraphenyl-21H,23H-porphine iron(III) chloride	TPPFeCl	TPP8
Bromocresol green	NA	pH1
Bromothymol blue	NA	pH2
Basic red	NA	pH3
Cresol red	NA	pH4

Table S2 Definition of the 50 color components.

No.	Feature name	Description
1	R	<i>Mean of R</i>
2	G	<i>Mean of G</i>
3	B	<i>Mean of B</i>
4	ExG	$2G-R-B$
5	ExR	$1.4R-G$
6	ExB	$1.4B-G$
7	ExG_min_ExR	$3G-2.4R-B$
8	CIVE	$0.441R-0.811G+0.385B+18.78745$
9	NDGR	$(R-G)/(R+G)$
10	CBDI	$R+G-2B$
11	NGRDI	$(G-R)/(G+R)$
12	GLI	$(2G-R-B)/(2G+R+B)$
13	NGBDI	$(G-B)/(G+R)$
14	WI	$(G-B)/(R-G)$
15	GRRI	G/R
16	GBRI	G/B
17	ri	$R/(0.2989R+0.5870G+0.1140B)$

18	gi	$G/(0.2989R+0.5870G+0.1140B)$
19	bi	$B/(0.2989R+0.5870G+0.1140B)$
20	R_nor	$R/(R+G+B)$
21	G_nor	$G/(R+G+B)$
22	B_nor	$B/(R+G+B)$
23	R_min_G_nor	$(R-G)^2/[(R-G)^2+(R-B)^2+(B-G)^2]$
24	R_min_B_nor	$(R-B)^2/[(R-G)^2+(R-B)^2+(B-G)^2]$
25	B_min_G_nor	$(B-G)^2/[(R-G)^2+(R-B)^2+(B-G)^2]$
26	RGB_dis	$\text{sqrt}(R^2+G^2+B^2)$
27	R_dis_nor	$R/\text{sqrt}(R^2+G^2+B^2)$
28	G_dis_nor	$G/\text{sqrt}(R^2+G^2+B^2)$
29	B_dis_nor	$B/\text{sqrt}(R^2+G^2+B^2)$
30	G_min_R	$G-R$
31	R_min_B	$R-B$
32	G_min_B	$G-B$
33	ERI	$(R-G)(R-B)$
34	EGI	$(G-R)(G-B)$
35	EBI	$(B-G)(B-R)$
36	H	<i>Mean of H</i>
37	S	<i>Mean of S</i>
38	V	<i>Mean of V</i>
39	L	<i>Mean of L</i>
40	a	<i>Mean of a</i>
41	b	<i>Mean of b</i>
42	chroma_C	$\text{sqrt}(a^2+b^2)$
43	hue	$\text{arctan}(b/a)$
44	hue_mod	$\arccos \left[\frac{2R-G-B}{2\text{sqrt}(R^2+G^2+B^2-RG-GB-RB)} \right]$
45	C	<i>Mean of C</i>
46	M	<i>Mean of M</i>
47	Y	<i>Mean of Y</i>
48	Y_1	<i>Mean of Y in NTSC YIQ</i>
49	I	<i>Mean of I in NTSC YIQ</i>
50	Q	<i>Mean of Q in NTSC YIQ</i>

S2.5.1 Outlier rejection and dataset division

The principal components score matrix, squared Mahalanobis distance were shown by the following expressions.

$$T_{n \times f} = X_{n \times m} \times P_{m \times f} \quad (S1)$$

$$D^2 = (T_i - T)M^{-1}(T_i - T)' \quad (S2)$$

Among them, where $X_{n \times m}$ was the CSA color components matrix, $P_{m \times f}$ was the loading matrix, n was the number of samples, m was the number of variables, f was the number of principal components, M-1 was the covariance matrix of the score matrix, T_i was the scores vector of sample i, T was the average color components of all samples.

S2.8 Establishment and evaluation of the pH value quantitative model

RBF and SVR mathematical model were shown by the following expressions.

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x_i, x)$$

$$s.t \begin{cases} 0 \leq \alpha_i \leq C \\ 0 \leq \alpha_i^* \leq C \\ \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \end{cases} \quad i = 1, 2, \dots, n \quad (S3)$$

$$K(x_i, x) = \exp(-g ||x_i - x||^2) \quad (S4)$$

Among them, α_i and α_i^* were relaxation variables, b was undetermined parameter, C was the penalty parameter, g was the kernel parameter, x was the center of the kernel function.

MSECV, R_c^2 , RMSEC, R_p^2 , RMSEP, and RPD were shown by the following expressions.

$$MSECV_k = \frac{1}{n} \sum_{i=1}^n \left(y_{fit} - y_{test} \right)^2 \quad (S5)$$

$$MSECV = \frac{1}{k} \sum_{i=1}^k MSECV_k$$

$$R_c^2 = 1 - \frac{\sum_{i=1}^{n_c} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n_c} (y_i - \bar{y}_c)^2} \quad (S6)$$

$$RMSEC = \sqrt{\frac{1}{n_c} \sum_{i=1}^{n_c} (y_i - \hat{y}_i)^2} \quad (S7)$$

$$R_p^2 = 1 - \frac{\sum_{i=1}^{n_p} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n_p} (y_i - \bar{y}_p)^2} \quad (S8)$$

$$RMSEP = \sqrt{\frac{1}{n_p} \sum_{i=1}^{n_p} (y_i - \hat{y}_i)^2} \quad (S9)$$

$$RPD = \frac{SD}{RMSEP} \quad (S10)$$

Among them, where n was the number of remaining 1-fold samples, y_{fit} was the predicted value of the i th sample, y_{test} was the measured value of the i th sample, k was the fold number of cross validation, n_c and n_p were the number of samples in the calibration and prediction sets, y_i was the measured value of the i th sample in the calibration and prediction sets, \hat{y}_i was the predicted value of the i th sample in the calibration and prediction sets, \bar{y}_c was the average of the measured values of all samples in the calibration and prediction sets, SD was the standard deviation of the prediction set.