A machine learning-based colorimetric sensor array for high-precision pathogen identification in household refrigerators

Yu Zhang, Gong-Xiang Qi, Yong-Liang Yu, Meng-Xian Liu* and Shuai Chen*

Research Center for Analytical Sciences, Department of Chemistry, College of Sciences, Northeastern University, Box 332, Shenyang 110819, China

Corresponding authors.

*E-mail: mengxian.liu.d5@tohoku.ac.jp (M.-X. Liu); chenshuai@mail.neu.edu.cn (S. Chen)

Tel: +86 24 83688944; Fax: +86 24 83676698

Electronic Supplementary Information

Chemicals, materials and instruments

2,4-dinitrophenylhydrazine, 4,4'-azo diphenylamine, P-toluenesulfonic acid, pararosaniline, tetraphenylcobalt porphyrin, tetraphenylporphyrin zinc, tetraphenyl, Benedict's reagent, and Ellman's reagent were purchased from Macklin Reagent Co., Ltd. Silver nitrate was purchased from Sinopharm Chemical Reagent Co., Ltd. The organic dyes were dissolved in deionized water and mixed with a predetermined volume of sulfuric acid and Para-Toluene Sulfonate acid (Acros). Sodium hydroxide acid, hydrochloric acid, and ammonia were used as auxiliary reagents in inorganic chromogenic analysis. The carrier substrate of dye spots is agar gel (Kermel Co., LTD), and the container for the agar and dye is Biosharp. For representative VOCs, 50% glutaraldehyde solution was used as an aldehyde, and 2-nonanone (Acros) was used as a representative ketone, both diluted asneeded. Commercial reagents are used as received. The detailed composition ofeach CSA dye spot is shown in Table S1.The eggs were bought at the local grocery store (campus supermarket, Shenyang, Liaoning, China). The *E.coli* O157:H7 (EHEC) CICC10907, *Staphylococcus aureus* (*S. aureus*) CMCC26003, *Salmonella* CCUG60567, and *Listeria monocytogenes* (*L. mono*) ATCC19111 were purchased from BeNa Culture Collection (BNCC, Beijing, China). Brain heart extract broth (BHI) was purchased from Solarbio (Beijing, China). Phosphate buffered saline (PBS) was purchased from Biosharp (Guangzhou, China).

The rheological characteristics of the agar gels were obtained by means of a rheometer (Mars 40, Thermo Fisher Scientific, USA). The porosity of the agar gels was obtained by a mercury-pressure instrument (Autopore V9620, Micromeritics, USA). Scanning electron microscopy (SEM) images were collected on a SU8010 field-emission electron microscope (Hitachi, Tokyo, Japan).

Fabrication of colorimetric sensor array (CSA)

Table S1 shows the detailed composition of each chromogenic dye spot and its corresponding analyte. 10 μL of each dye mixture shown in Table S1 was added to a microstrip containing 100 μL of agar gel (agar: H_2O = 3: 200, w/w). Then cool and

solidify in vacuum for 5 min before use.

Humidity tolerance experiment of CSA

The agar gel-based CSAs of different agar: H_2O ratios (1: 200, 3: 200, and 1: 40, w/w) were placed in environments with different RH (35%, 55%, 75%) at 4℃. Recorded the color patterns on the first day and the seventh day, and then evaluated the humidity tolerance of the CSA based on the color differences.

Frost resistance experiment of agar gels

Agar gels were made in the ratio of 1: 200, 3: 200, 1: 40 (agar: H_2O, w/w) and stored at 4℃ and RH = 55%. Recorded the shape on the first day and the seventh day, and then evaluated the frost resistance of the CSA based on the shape changes.

Response of CSA to VOCs in refrigerated environments

In this study, the response of the CSA to VOCs from different pathogens in a single culture was evaluated. EHEC (CICC 10907), *S. aureus* (CMCC 26003), *Salmonella* (CCUG 60567), and *L. mono* (ATCC 19111) were selected as the target foodborne pathogens, including Gram-positive, Gram-negative, psychrophilic, and neutral bacteria, all of which are of great public health relevance. Standard bacterial cultures were resuscitated from -80℃ by night-growing BHI broth (Solarbio Co.). Single colonies of the four bacteria species were obtained by plating on plate-counting agar (Kermel Co., LTD). Single colonies of each strain were inoculated into BHI broth and cultured at 37℃ overnight. The pathogen population of each strain was counted in plate agar. To determine the interactions between VOCs and CSA in liquid culture, an appropriate volume of overnight bacterial culture with an initial bacterial species colony density of ~ 3 log CFU/mL was inoculated into a conical flask containing 20 mL of BHI broth. After incubated at 37℃ for 12 h, CSA was placed in the headspace of the conical vial and set in a refrigerated environment (4℃, RH = 55%) for 7 days. During

this period, the conical flask was opened at fixed times and the bacterial numbers and color patterns were recorded.

Response of the CSA to a real sample in refrigerated environments

The strains used were the same as those mentioned above. *S. aureus* (CICC 10907) and *L. mono* (ATCC 19111) strains were collected by centrifugation at 8000 rpm for 5 min, then washed twice with sterile phosphate-buffered saline (1 × PBS) and resuspended in PBS. Eggs were purchased from a local grocery store and the shells were reserved. The segmented eggshells (1.5 ± 0.5 g) were placed in a Petri dish and each dot was seeded with cell suspension (~ 5 log CFU/mL). Then packaged the Petri dish, sealed with a sealing membrane, attached the CSA to the dish and stored for 7 days (4℃, RH = 55%). The images were then photographed to extract RGB values, and the pathogen species on the eggshells were analyzed based on the acquired images.

Get color pattern diagram and RGB value extraction

Machine learning is a powerful tool for extracting relevant information from large and diverse datasets. Therefore, we developed and trained a multi-layer neural network (NN) to provide a highly automated and reusable solution. The CSA was photographed by a smartphone using CAMS (Shanghai INTSIG Information Technology Development Co., Ltd.). The obtained images were adjusted by rotation and noise reduction, after which the average R/G/B value of each dye point was read using Snipaste software.

Principal component analysis (PCA)

PCA of the original data matrix was performed using Origin 2021 software. The RGB information of the 16 stained points in the obtained CSA was mapped to a completely new orthogonal feature, the PC, by the algorithm in Origin 2021 (Table S6). A set of mutually orthogonal axes is sequentially found from the original space. The first new axis is chosen to be the direction with the largest variance in the original data, the second new axis is chosen to be the plane orthogonal to the first axis that makes the

largest variance, and the third axis is the plane orthogonal to the 1st and 2nd axes that makes the largest variance. By analogy, the new axes are obtained in this way. This is equivalent to keeping only the dimensional features that contain most of the variance and ignoring the dimensional features that contain almost zero variance to realize the dimensionality reduction of the data features.

Linear discriminant analysis (LDA)

For LDA, the raw data matrix consisting of RGBs of the 16 dye points in the CSA obtained was analyzed using IBM SPSS Statistics 27 software (Table S7). The raw RGB response patterns were converted to standard scores by the description algorithm in IBM SPSS Statistics 27, and all observations were grouped under the condition that the ratio of between-class variance to within-class variance was maximized according to the pre-assigned groups. LDA builds two scatter matrices: (1) an in-between-class matrix that computes the distance between the mean of each class, and (2) a within-class matrix that returns the distance between the mean of each class and the data within that class. The remaining process is similar to PCA, where the new features, the Factor, are obtained.

Machine learning and pattern recognition

Each set of 16 digitized R/G/B color information was paired with a category label prior to NN training and testing. A key advantage of the multilayer NN structure is that the accuracy of the training data set increases with the number of repetitions for each category. A database of 120 CSA's data for training, 24 CSA's data for testing, and an additional 24 CSA's data for differentiating pathogen on eggshells were applied to a refrigerated environment to validate its potential for application in a real-world setting. The results of all combinations were combined to report the accuracy of the predictions. The NN algorithm also matches the assigned labels with the independently validated labels and reports and predicts classification accuracy for both the BHI experiment and the eggshell experiment with the trained model. The assigned and independently validated labels must be matched to the correct classification (identification) of the samples.

NN trained on Spyder Python 3.7 (MIT, Cambridge, MA) using Tensorflow (Google Brain, Mountain View, CA), 1000-step (iterative) neural network implemented on a five-layer neural network (one layer of input nodes, one layer of output nodes, and three layers of hidden nodes) optimization and maximum test accuracy with 256 neurons per layer of hidden nodes. The last layer of the multilayer neural network contains a threshold function to discretize the results of the neural network and assigns a discrete label to each set of 48 R/G/B input values in the experiment. Other optimized training parameters include a learning rate of 0.02 and a learning rate of 2×10^{-6} used in the Adam Optimizer algorithm to minimize cross-entropy.

Limit of detection (LOD)

Defining LOD for the CSA-NN method is not as straightforward as traditional microbial detection. When the CSA and sample are enclosed in a chamber, the change in dye pattern corresponds to the time integral of the VOC concentration, which is closely related to factors such as the number of microorganisms, growth rate, growth phase, and time. the CSA exposure time is particularly important for establishing LODs. Therefore, if the exposure time is long enough or the growth rate is high enough, the LOD can be very low. As can be seen in Fig. S11 and S12, at a microbial load of ~5 log CFU/mL, CSA developed a significant pattern shift within 12 h At a microbial load of 3-4 log CFU/mL, CSA developed a significant pattern shift within 24 h. Considering that the application of CSA in household refrigerators usually requires extended exposure times (e.g., long-term food storage), the time-integration capability of our method implies that the LOD of CSA, in this case, would be reduced to levels much less than 3-4 log CFU/mL.

Read out time

The normal pathogen population level of the food product in retail sale is approximately 3-5 log CFU/mL.[1,2] Furthermore, in a refrigerated environment, pathogen growth is slow when the pathogen concentrations are in the range of 3-5 logs CFU/mL, e.g. at a concentration of 5 log CFU/mL, the concentration of EHEC, *S.aureus*, and *Salmonella* remains almost constant, with a slight increase in *L.mono* (~10%) after 24 h' growth (Fig. S12). Therefore, in order to simulate the actual application environment, we

examined the changes in the colour pattern of CSA with increasing exposure time at a pathogen concentration of 5 log CFU/mL in a refrigerated environment to examine the read out time. In this experiment, the read out time was defined as the time taken for a significant change (≥3 staining spots) or a color change visible to the naked eye in the CSA colour pattern in a refrigerated environment (4°C, RH = 55%) with a microbial load of ~5 log CFU/mL.[3,4]As shown in Fig. S11, the colour patterns of the CSA changed significantly after 4 h of exposure to pathogen-produced VOCs, e.g. dye spots 5, 8, and 14-16. Therefore, the readout time was set to 4 h.

Color stay time

The RGB images of the CSA at room temperature and refrigerated conditions for 7 days were examined to assess its color stay time. As shown in Fig. S10, the color change of the CSA was negligible after 7 days of storage at room temperature and refrigerated conditions, respectively, indicating that the color of the CSA can be maintained for at least 7 days.

Specificity response

As shown in Fig. S12, it is obvious that the color difference maps of CSA after being exposed to blank controls and BHI medium show negligible changes, while those of CSA exposed to different species of pathogens display fingerprint-type changes, indicating that the CSA responds specifically to the VOCs of the pathogens.[5]
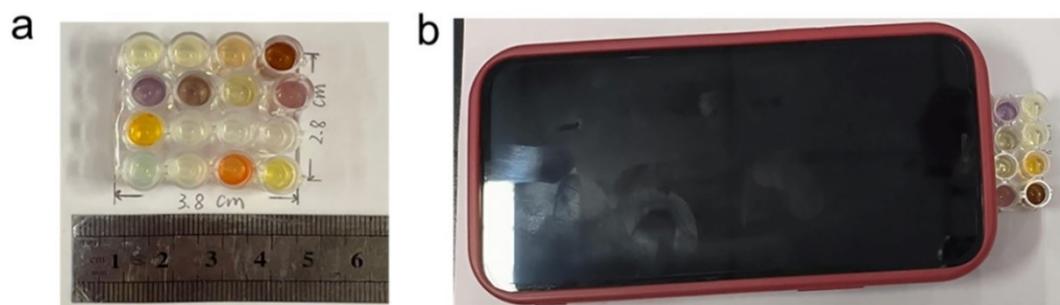
Fig. S1. (a) Real picture of the colorimetric set-up. (b) Real picture of smartphone and the colorimetric set-up.
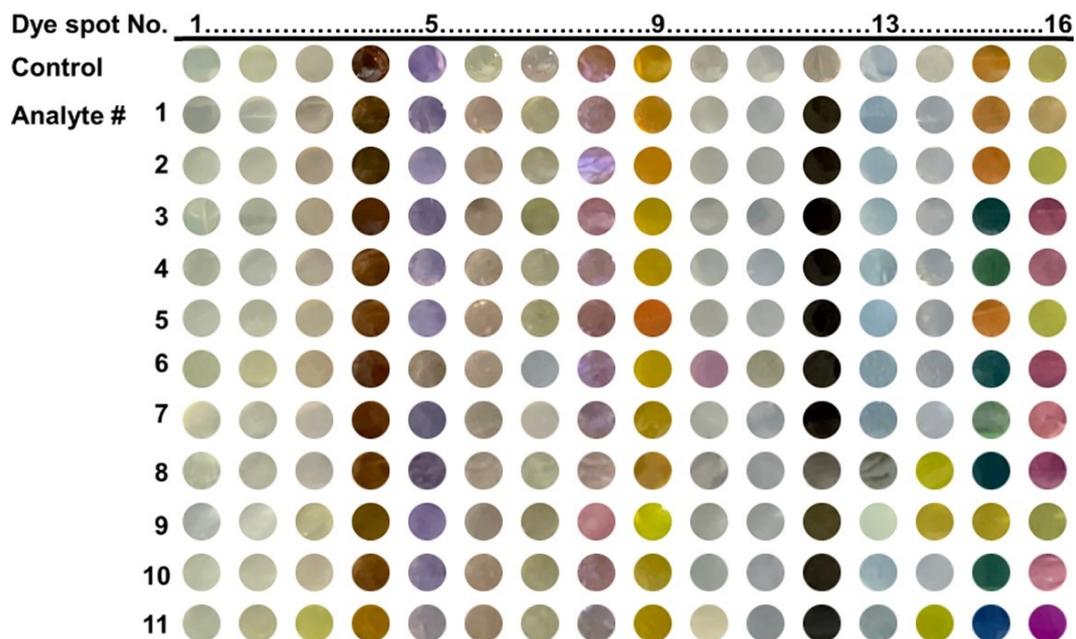
Fig. S2. CSA standard chromogenic response to analyte-VOCs. The compositions of dye spots 1-16 and VOCs are listed in Table S2-S4, respectively.
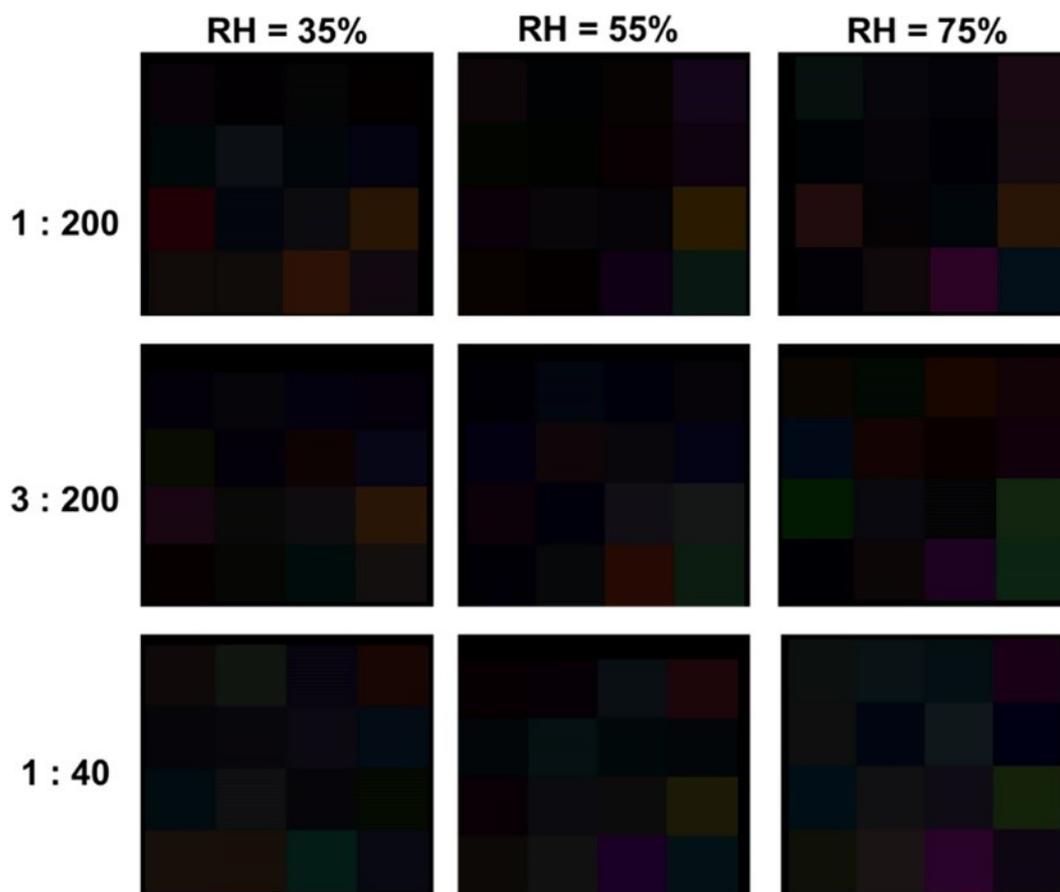
Fig. S3. Humidity tolerance of CSA assembled using agar gels made in different ratios of agar to water. The 7 days color difference maps of CSA in RH = 35%, RH = 55%, and RH = 75%, respectively. (Agar: H_2O = 1: 200, 3: 200, and 1: 40, w/w)
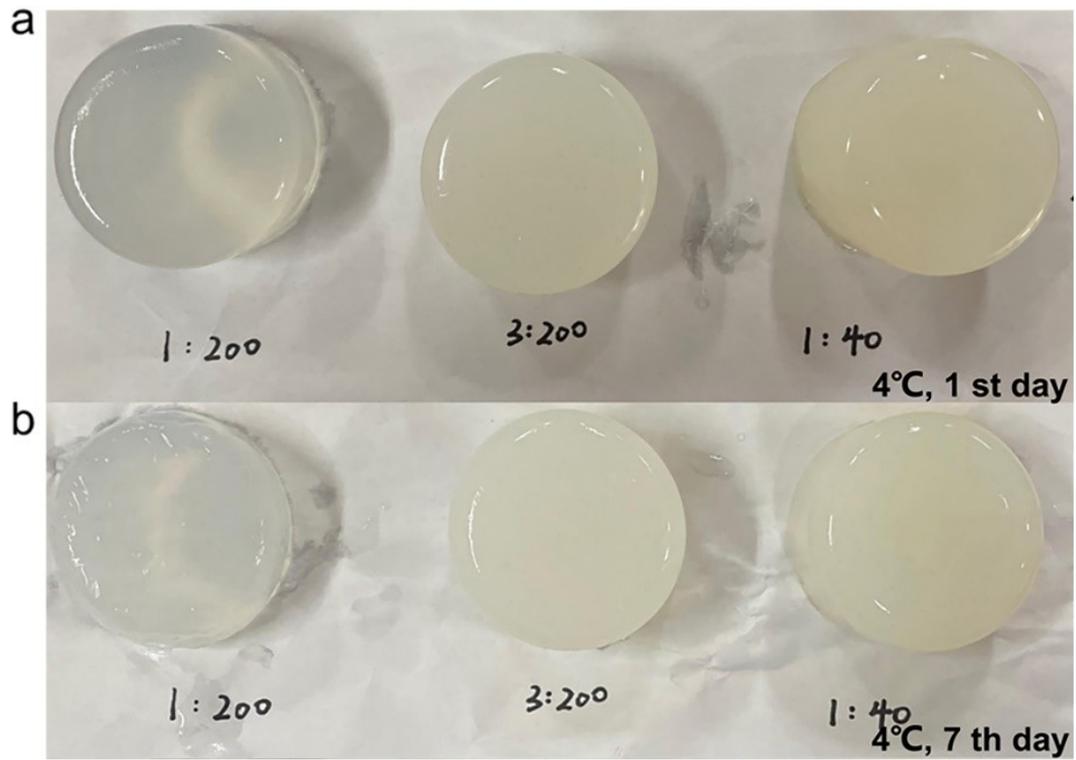
Fig. S4. Frost resistance of agar gels. Images of the prepared agar gel on the (a) first and (b) seventh day in a refrigerated environment (4°C, RH = 55%), respectively.
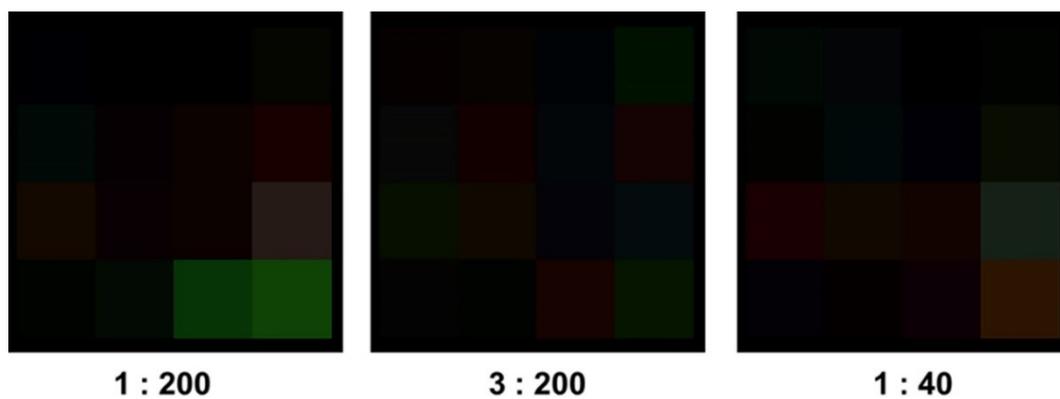
Fig. S5. Color difference maps the agar gels prepared at different ratios of agar to water after 7 days of storage in a refrigerated environment. (Agar: H_2O = 1: 200, 3: 200, and 1: 40, w/w)
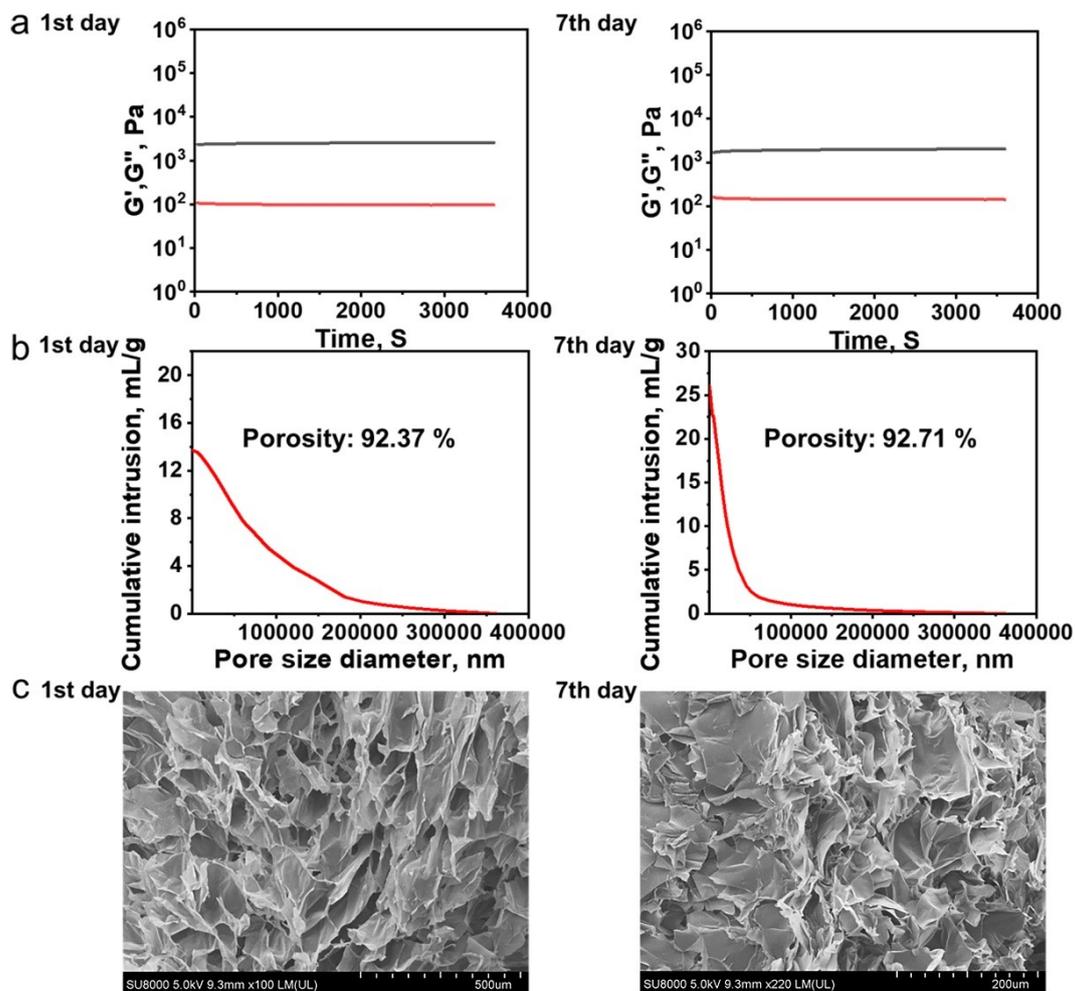
Fig. S6. (a) Rheometer measurements of agar gels before and after 7 days of storage in a refrigerated environment. (b) Pore size distribution and (c) SEM images of freeze-drying agar gels before and after 7 days of storage in a refrigerated environment. (Agar: H_2O = 3: 200, w/w)
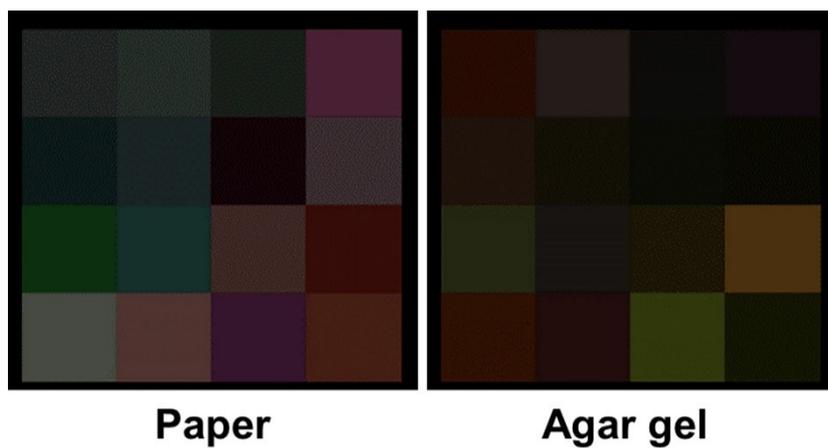
Paper   **Agar gel**

Fig. S7. Color difference maps of paper- and agar gel-based CSAs before and after 7 days of storage in a refrigerated environment.
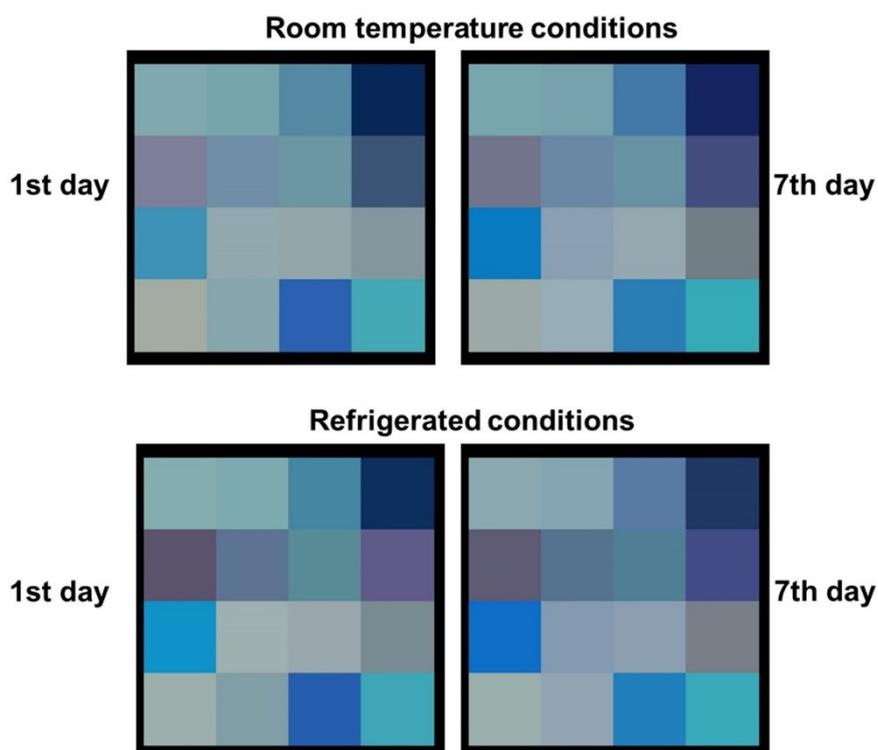
Fig. S8. RGB images of the CSA before and after 7 days of storage at room temperature and refrigerated conditions. (Room temperature conditions: 20°C, RH = 50%; refrigerated conditions: 4°C, RH = 55%)
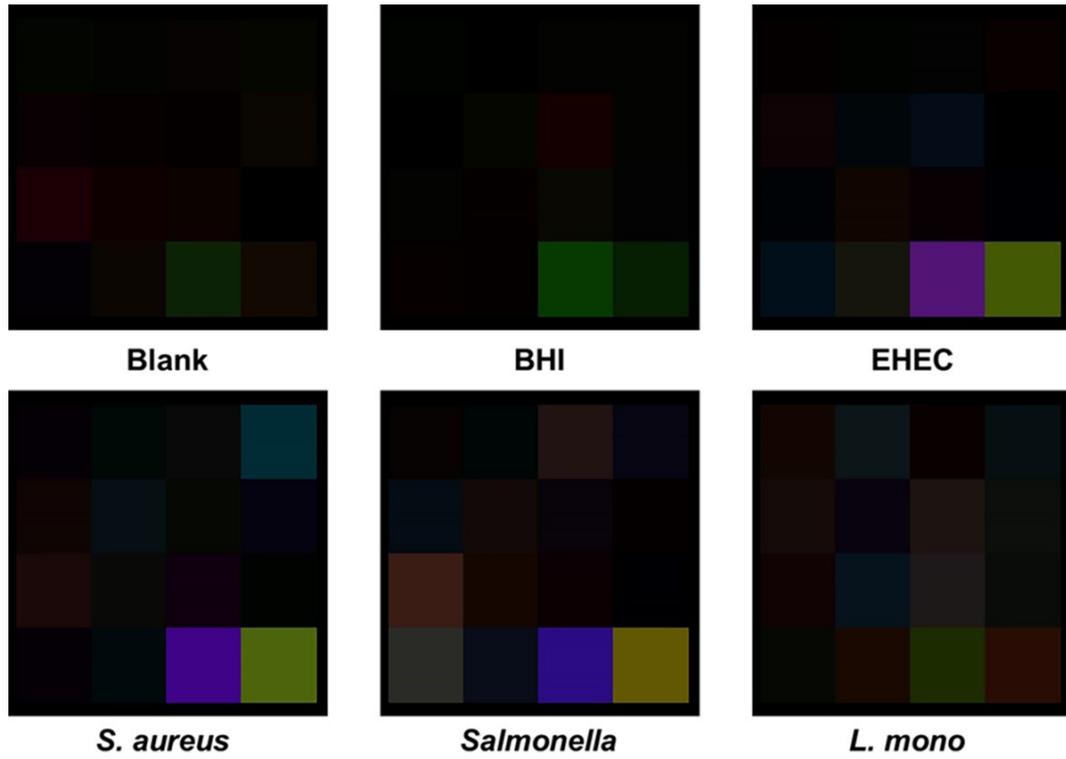
Fig. S9. Specific response of the CSA. Color difference maps of the CSA after exposed to blank, BHI medium, EHEC, *S. aureus*, *Salmonella*, and *L. mono* for 24 h. (Pathogen concentration: ~5 log CFU/mL)
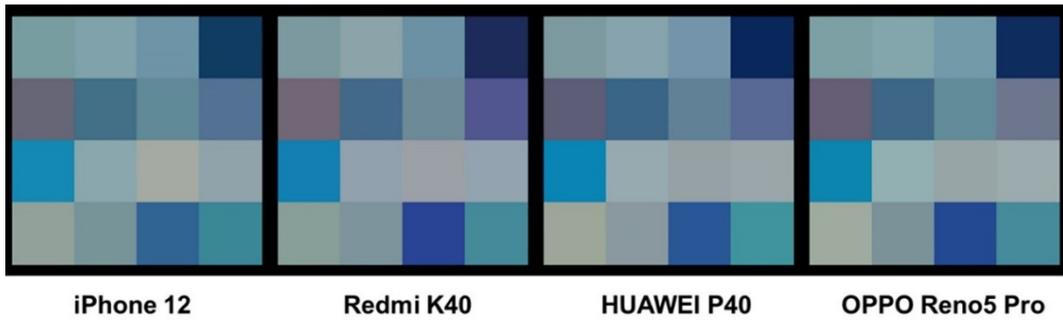
Fig. S10. RGB images obtained by shooting the same CSA with different models and brands of smartphones.

Fig. S11. Colour patterns of CSA incubated with EHEC, *S. aureus*, *Salmonella*, and *L. mono* over different time. (Pathogen concentration: ~5 log CFU/mL)

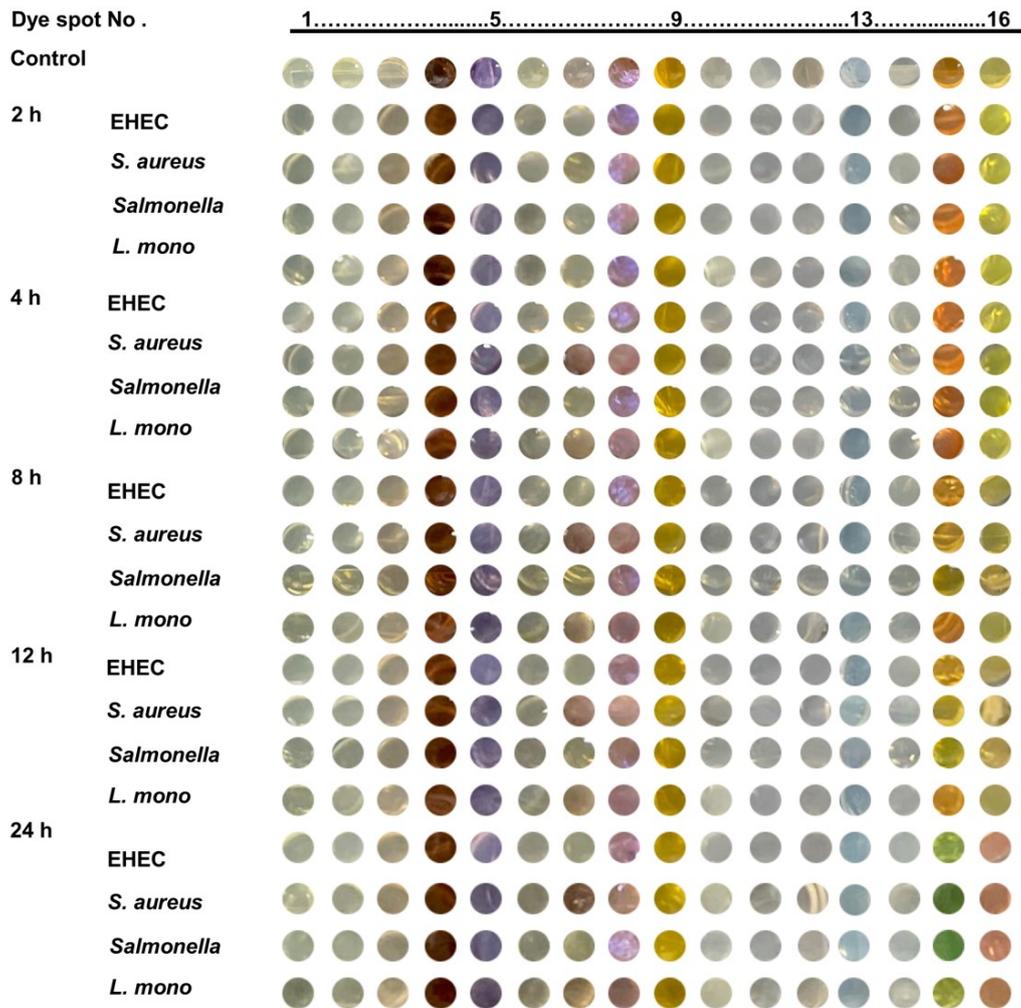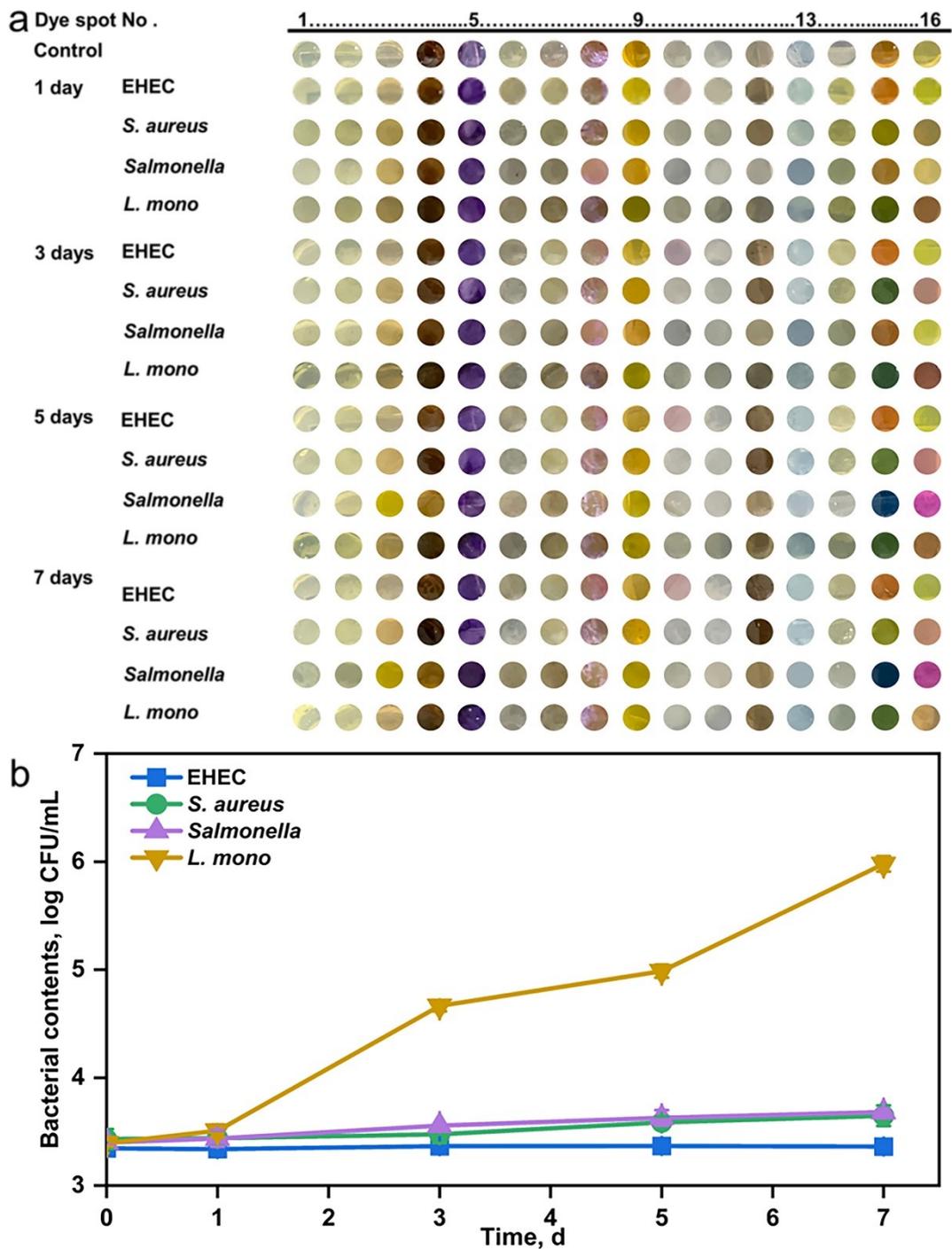Fig. S12. (a) CSA color patterns from VOC testing for EHEC, *S.aureus*, *Salmonella*, and *L.mono* in BHI broth at 4°C, RH = 55% for 7 days. (b) Growth curves of EHEC, *S. aureus*, *Salmonella*, and *L. mono*.

Table S1. Conventional culturing and commercial rapid detection methods of pathogens.

Method	Multiplex without Customization	Machine Learning	Use in the refrigerator directly	Reference
Plate counts	No	No	No	6
NGS[a]	Yes	No	No	7
DNA/PCR[a]	No	No	No	8
RNA/RT-PCR[a]	No	No	No	9
Lateral flow	No	No	No	10
Immunoassay	No	No	No	11
Colorimetric sensor array (CSA)	Yes	Yes	Yes	This work

[a] Abbreviations: NGS (next generation sequencing); PCR (polymerase chain reaction); RT-PCR: (reverse transcription polymerase chain reaction).

[b] Red indicates advantages.

Table S2. Detecting targets and dyes used in this study.

Target VOCs	Dye Composition (s)
aldehydes, ketones[12]	2,4-dinitrophenylhydrazine + H_2SO_4
aldehydes, ketones	2,4-dinitrophenylhydrazine + TsOH
ketones	4,4'-azodianiline + H_2SO_4
ketones	4,4'-azodianiline + TsOH
ketones	Pararosaniline + H_2SO_4
amines[13]	CoTPP
amines	Fe(TPP)Cl
amines	ZnTPP
pH[14]	Methyl orange
indole	4-dimethylaminobenzaldehyde
sulfur compounds[15]	Chloranil
aldehydes	Tollen's reagent
reducing sugar	Benedict's reagent
sulfhydryl group	Ellman's reagent
pH[16]	Methyl red + Bromothymol blue
pH	Phenol red

Table S3. List of VOCs used in the CSA. Group A: aldehydes and ketones; Group B: acids, aromatic compounds, sulfur compounds, and alcohols; Group C: amines.

VOC groups	VOC#	VOCs	Preparation
Control	No VOC (4℃, RH = 55%)		
Group A (aldehydes and ketones)	1	5% glutaraldehyde	5% v/v in DI H_2O
	2	10% glutaraldehyde	10% v/v in DI H_2O
	3	5% 2-nonanone	5% w/w in dimethyl sulfoxide
	4	5% 2-nonanone	5% w/w in DI H_2O
Group B (common microbial metabolites)	5	5% acetic acid	5% w/w in DI H_2O
	6	1% indole	10 mg/mL in 10 x diluted DMSO
	7	1,4-butanediol vapor	10 mL in a closed 50 mL tube (100 wt%)
	8	hydrogen sulfide vapor	2 mg/mL Na_2S in DI H_2O, 10 mL
	9	ethanol vapor	10 mL (100 wt%)
Group C (amines)	10	Tris buffer, pH 7.4	0.05 mol/L
	11	10% trimethylamine	10% v/v in DI H_2O

Table S4. Preparation of dye spots.

Spot #	Preparation
1	1 mg/mL 2,4-dinitrophenylhydrazine in MOE + 1 M H_2SO_4; 80: 1 v/v
2	1 mg/mL 2,4-dinitrophenylhydrazine in MOE + 1 M TsOH; 80: 1 v/v
3	1 mg/mL 4,4'-azodianiline in MOE + 1 M H_2SO_4
4	2 mg/mL 4,4'-azodianiline in MOE + 1 M TsOH; 80: 1, v/v
5	1 mg/mL pararosaniline in MOE + 1 M H_2SO_4
6	6 mg/mL in MOE
7	6 mg/mL in MOE
8	6 mg/mL in MOE
9	0.3% methyl orange in ethanol
10	5 g 4-dimethylaminobenzaldehyde+75 mL ethanol+25 mL HCl (12 mol)
11	0.25 g tetra-chloro-benzoquinone + 100 mL N, N-dimethylformamide
12	2 mL of 0.2 M $AgNO_3$ in a tube rinsed with 3 M NaOH was mixed with one drop of 3 M NaOH; 2.8% ammonia solution was then added drop by drop until no precipitate of silver oxide was observed.
13	Used as received.
14	Used as received.
15	0.3% (w/v) methyl red in aqueous ethanol (1: 1 v/v) + 0.3% (w/v) bromothymol blue in aqueous ethanol (1: 1 v/v); 3: 2, v/v
16	0.04% phenol red (wt%) in DI H_2O

Table S5. Variation of the volume of agar gels prepared at different ratios of agar to

Agar gels (agar/H_2O, W/W)	1: 200 (cm³)	3: 200 (cm³)	1: 40 (cm³)
1st day	4	7	8
7th day	4.7	7.5	10
Difference	0.7	0.5	2

water before and after 7 days of storage in a refrigerated environment.

Table S6. Correlation coefficient of PCs of PCA in pathogens discrimination.

Principal Component Number	PC 1 (31.7%)	PC 2 (26.0%)
1	0.61756	4.76259
2	0.13407	5.33623
3	0.04261	5.24877
4	−1.70063	5.6268
5	−1.1978	5.51697
6	−7.96891	−1.29026
7	−6.7743	−0.94856
8	−3.63348	−0.401
9	−3.54795	−1.08554
10	−3.24054	−1.33127
11	4.06405	−0.33452
12	3.27663	−0.34149
13	2.35412	−1.12026
14	4.84469	0.23322
15	7.44994	0.81323
16	−1.82039	−4.95027
17	−0.26954	−4.97705
18	1.2777	−3.91849
19	1.77055	−3.17659
20	4.32162	−3.66252

Table S7. Canonical discriminant function coefficients of LDA in pathogens discrimination.

VAR #	Function 1	Function 2	Function 3
VAR00002	–0.404	–0.100	–0.037
VAR00007	–0.175	–0.113	0.056
VAR00049	–0.105	–0.065	–0.034
VAR00044	0.207	0.026	0.202
VAR00048	0.198	–0.060	–0.076
VAR00047	–0.428	0.058	0.001
VAR00046	0.294	–0.195	–0.048
VAR00034	0.167	–0.164	0.165
VAR00045	–0.450	0.113	–0.223
VAR00042	0.822	0.553	0.686
VAR00043	–0.071	0.122	–0.061
VAR00041	–0.106	–0.549	–0.684
VAR00036	0.138	0.076	0.035
VAR00040	0.134	0.176	0.272
VAR00038	0.085	–0.333	0.081
VAR00031	0.031	0.252	–0.066
(Constants)	–39.634	15.190	–49.981

Supplementary References

1. R. Elson, F. Burgess, C. L. Little, R. T. Mitchell, S. Local Authorities Co-Ordinators of Regulatory and A. the Health Protection, *J. Appl. Microbiol.*, 2004, **96**, 499-509.

2. J. K. Andersen, B. Nørrung, S. da Costa Alves Machado and M. Nauta, *Food Control.*, 2015, **58**, 29-32.

3. Z. Jia, Y. Luo, D. Wang, Q. N. Dinh, S. Lin, A. Sharma, E. M. Block, M. Yang, T. Gu, A. J. Pearlstein, H. Yu and B. Zhang, *Biosens. Bioelectron.*, 2021, **183**, 113209.

4. Z. Jin, Y. Mantri, M. Retout, Y. Cheng, J. Zhou, A. Jorns, P. Fajtova, W. Yim, C. Moore, M. Xu, M. N. Creyer, R. M. Borum, J. Zhou, Z. Wu, T. He, W. F. Penny, A. J. O'Donoghue and J. V. Jokerst, *Angew. Chem. Int. Ed.*, 2022, **61**, e202112995.

5. M. Yang, X. Liu, Y. Luo, A. J. Pearlstein, S. Wang, H. Dillow, K. Reed, Z. Jia, A. Sharma, B. Zhou, D. Pearlstein, H. Yu and B. Zhang, *Nat. Food.*, 2021, **2**, 110-117.

6. J. P. Gorsuch, Z. Jones, D. L. Saint and C. L. Kitts, *J. Microbiol. Methods.*, 2019, **164**, 105682.

7. K. Horiba, J. I. Kawada, Y. Okuno, N. Tetsuka, T. Suzuki, S. Ando, Y. Kamiya, Y. Torii, T. Yagi, Y. Takahashi and Y. Ito, *Sci. Rep.*, 2018, **8**, 3784.

8. T. Dreo, M. Pirc, Z. Ramsak, J. Pavsic, M. Milavec, J. Zel and K. Gruden, *Anal. Bioanal. Chem.*, 2014, **406**, 6513-6528.

9. J. M. Williams, M. Trope, D. J. Caplan and D. C. Shugars, *J. Endod.*, 2006, **32**, 715-721.

10. S. Shan, W. Lai, Y. Xiong, H. Wei and H. Xu, *J. Agric. Food. Chem.*, 2015, **63**, 745-753.

11. M. Magliulo, P. Simoni, M. Guardigli, E. Michelini, M. Luciani, R. Lelli and A. Roda, *J. Agric. Food. Chem.*, 2007, **55**, 4933-4939.

12. Z. Li, M. Fang, M. K. LaGasse, J. R. Askim and K. S. Suslick, *Angew. Chem. Int. Ed.*, 2017, **56**, 9860-9863.

13. N. A. Rakow, A. Sen, M. C. Janzen, J. B. Ponder and K. S. Suslick, *Angew. Chem. Int. Ed.*, 2005, **44**, 4528-4532.

14. Y. Zhang and L.-T. Lim, *Sens. Actuat, B-Chem.*, 2018, **255**, 3216-3226.

15. K. Liu, L. He, X. He, Y. Guo, S. Shao and S. Jiang, *Tetrahedron Lett.*, 2007, **48**, 4275-4279.

16. C. Rukchon, A. Nopwinyuwong, S. Trevanich, T. Jinkarn and P. Suppakul, *Talanta*, 2014, **130**, 547-554.