ORIGINAL PAPER

Optimization of electronic nose measurements and discrimination of apple juices in combination with supervised pattern recognition

Hao Wu[1,2] · Xuxin Gao[1] · Jiaru Shan[1] · Xiangxue Jia[1] · Yue Wang[1]

Abstract

A common theme of food authentication studies is the requirement to which the raw apple juice samples can be compared to establish its authenticity. In order to discriminate eight varieties of apple juices using electronic nose (EN), we tried to extract the most relevant information from EN response signals. Experiment parameters were optimized to ensure the response curve fully characterizing sample information. The optimal conditions were 10 mL of volume and 90 min equilibration. Sensor optimization was conducted to eliminate redundant information. Sensors W1C, W5C, W3S, W2S, W5S and W1W were chosen for pattern recognition. This process improved the PCA-based pattern performance. Best discrimination performance was obtained utilizing response signals of stationary phase according to multivariate analysis. Linear discriminant analysis (LDA) and Support Vector Machine (SVM) were carried out to develop discrimination models. Both LDA and SVM achieved satisfactory variety-based classification performance, with both 100% accuracy classification rates in terms of recognition ability and prediction ability. The perfect performance indicated that EN can be successfully applied in apple juices discrimination. And the results obtained will play a positive role in successful commercialization application of EN to apple juice industry.

Keywords Apple juice · Electronic nose · Optimization · Feature extraction · Discrimination · SVM

Introduction

Fresh apple juice is one of the most frequently consumed fresh juices all over the world, due to natural viridian, pleasant taste, good smell, established nutritional and economical value [1]. Fresh apple juice is usually processed directly to consumers or developed apple juice products. Aroma is one of the most important factors to evaluate character and quality of apple juices and final apple juice products [2]. There are more than 300 volatiles produced by apples [3]. Some compositions have been proved to be "character impact" compounds. In particular, some components in very low concentrations contribute typical characteristics to apple aroma [4].

Gas chromatography (GC) techniques coupled with GC data is the most often employed to determine aromatic profiles of fruits and beverages [5]. But solely chemical analysis is not possible to assess apple juice quality, due to the complex synergism, antagonism and combination of aroma compounds. Sensory evaluation techniques are of additional importance by refining the products' sensory properties [6]. However, these methods are difficult for common producers and consumers due to technique equipment and trained expert panelist requirements. Therefore, it is necessary to develop a reliable and simple method to certify the apple juice aroma, which is highlighting and in an attempt to ultimately appeal to producers and consumers.

The electronic nose (EN), an innovative analytical technique, recognizes samples via olfaction, offering a rapid, sensitive, non-invasive and easy-to-handle system to assess food quality [7]. Just like the nose of the human being, EN identifies the whole volatile information without individual ones. EN consists of a sampling system, an array of chemical sensors with different selectivity, an instrumentation system collecting electrochemical signals, and a pattern recognition system [8]. It is widely used in headspace analysis of liquid

✉ Hao Wu
beizai.nanfang@163.com

1   Heze Branch, Qilu University of Technology(Shandong Academy of Sciences), Heze, China

2   Biological Engineering Technology Innovation Center of Shandong Province, Heze, China

⚛ Springer

or solid food samples [9]. EN has been widely applied in apple juice sector, including quality assessment, control and assurance [10, 11], apple cultivars discrimination [12], differentiation of apple flavors and essences [13], prediction of Penicillium expansum spoilage and patulin concentration in apples and juice production [14], rapid screening of Alicyclobacillus acidoterrestris spoilage in fruit juices [15].

Although EN has such a wide range of applications, it still poses some problems, such as sensor drift, unsatisfactory repeatability and reproducibility due to sensitivity to operational conditions, or poor gas sensitivity and selectivity. The way to overcome these difficulties is to extract the most suitable information from the sensor array [16]. Generally, the measurement process of sensor array system consists of sensing, feature extraction, feature selection and pattern classification [17].

Headspace sampling is the critical point during the sensing phase. Sufficient volatiles generation can provide comprehensive information of samples. Experimental parameters influencing this process, such as water vapor, temperature, concentration of sample gas and equilibration time, have been identified. In the feature extraction phase, extracting the most relevant features from the output of sensor array maximizes available information for the follow-up analysis [18]. While applying EN to certain applications, not all the sensors are equally important. Each sensor responds to exposed substances in a different way. Broad spectrum response characteristic to compounds and cross-reactivity make sensors contain a number of irrelevant redundant features. Therefore, sensor array optimization is necessary, in order to eliminate noisy, irrelevant redundant information and improve identification performance [19]. Some sensors can be sequentially selected by weighting their discriminant ability to samples. The sensor optimization simplifies the sensor array. The classification accuracy rate is the most important standard to evaluate the performance of sensor set. Available feature selection can improve the output of the classifier and thus enhance the success recognition and prediction rate in classification pattern.

The aim of this work was to quantify the influence of various experimental parameters on EN measurements, extract the most suitable information from response curve, optimize sensor array and build satisfactory classification patterns.

Materials and methods

Samples preparation

Apple fruits including eight different varieties (Fuji, Golden Delicious, Qinguan Delicious, Pink Lady, Jonan Golden, Ralls Genet, Starkrimson and Gala) were collected in the late September in Baishui, Shaanxi, China. Carefully washed with tap water, fruits (3 kg apples per varieties) were freshly squeezed with a centrifugal juicer (Midea, JP351, China). Apple juices were obtained. The samples were selected in random sequences for each measurement to avoid chained analysis. All juice samples were immediately measured by EN to minimize the quality change.

Electronic nose data acquisition

Analyses were conducted with a portable electronic nose device PEN 3 (Airsense Analytics GmbH, Schwerin, Germany). The device consists of a sampling apparatus, a sensor chamber and pattern recognition software for analysis and data recording. The PEN 3 has an array of 10 different metal oxide semiconductor type chemical sensors positioned into small chambers (V = 1.8 mL). Table 1 lists all used sensors, their selectivity towards volatile compound and detection limitations.

Apple juices were placed into headspace vials with a volume of 40 mL. The headspace vials were then closed and the headspace inside it was equilibrated. During the

Table 1 Sensors used and their main applications in PEN3

Number in Array	Sensor-name	General description	Reference, mL/m³(ppm)
1	W1C	Aromatic compounds	Toluene, 10
2	W5S	Very sensitive, broad range sensitivity, react on nitrogen oxides, very sensitive with negative signal	NO2, 1
3	W3C	Ammonia, used as sensor for aromatic compounds	Benzene, 10
4	W6S	Mainly hydrogen, selectively (breath gases)	H2, 0.1
5	W5C	Alkenes, aromatic compounds, less polar compounds	Propane, 1
6	W1S	Sensitive to methane broad range	CH3, 100
7	W1W	Reacts on sulfur compounds, sensitive to many terpenes and sulfur organic compounds, which are important for smell, limonene, pyridine	H2S, 1
8	W2S	Detects alcohols, partially aromatic compounds, broad range	CO, 100
9	W2W	Aromatics compounds, sulfur organic compounds	H2S, 1
10	W3S	Reacts on high concentrations, sometime very selective (methane)	CH3, 100

measurement process, one luer-lock needle connected to a Teflon-tubing was used to penetrate the septum of the vial and to absorb headspace gas inside it. The sample gas was sucked into sensor chambers through the inlet at a rate of 400 mL/min. Zero gas was pumped from its port at the backside of the instrument into the sample gas path. The measurement phase lasted 55s, time enough for sensors to reach stable values. The response values were recorded by a computer every second. When a measurement was completed, a cleaning phase lasting 300s was initiated to clean the circuit and return sensors to their baseline.

Preliminary experiments showed that too much water vapor damaged the sensors. We enriched juice samples under the condition of normal temperature to reduce the bad effects of water vapor. Meanwhile, active carbon was placed in the gas inlet of the Teflon-tubing to extract water contents from sample gas. In addition, the mixture of substances in the headspace is dependent on temperature. Huge differences in temperature will change the composition of compounds within the gas sample and thus lead to a different pattern. By keeping the temperature constant, the variations in sample preparation are minimized. EN were held at temperature of 20 ± 1 °C and 50–60% RH during all experiments.

To explore the influence of experimental parameters on the responses of sensor array, different concentrations of volatile compounds and equilibration times were analyzed. All juice samples were separately sealed in vials of 40 ml. The volumes of apple juices was 1mL, 2.5 mL, 5 mL, 7.5 mL, 10 mL, 12.5 mL and 15 mL, aims to enhance gas concentration. The equilibration times was 5 min, 30 min, 60 min, 90 min, 120 min, 180 min, 240 min, 300 min. The one parameter varies simultaneously, while the other one is kept constant. The relevant variation interval must be determined for each factor, provided that the analyzed outputs vary monotonously within this range.

Data analysis

A one-way analysis of variance(ANOVA)compared with multiple comparisons (Tukey) was conducted to evaluate whether significant influences existed on EN response signals of different variances. The impact of controllable factors on the results was assessed by analyzing the contribution of individual factors' variation to the total variation using ANOVA. By applying the arcsine square root transformation, categorical outcomes can be effectively analyzed.

The Principal Component Analysis (PCA) is a crucial method for extracting features in pattern recognition. It enables the extraction of essential information from datasets by reducing the original data dimensionality. This transformation aims to maximize variance differences while minimizing dataset correlations, facilitating the visualization of similar samples and outliers, and displaying the exploratory overview of primary data distribution.

The supervised method of linear discriminant analysis (LDA) was employed to establish classification models, which has been widely and successfully utilized in the field of food quality control. LDA calculation utilizes class information to maximize the between-class variance ratio and minimize the within-class variance ratio. The leave-one-out method was applied as a cross-validation procedure to assess the performance of the LDA discriminant model. In this approach, the sample set was randomly divided into two datasets, with 2/3 of the samples allocated to the training set and the remaining 1/3 assigned to the testing set. The analysis focused on evaluating the accuracy rate of sample classification.

The Support Vector Machine (SVM), recognized as a highly effective classification method, is applicable to both linear and nonlinear data sets. SVM employs the theory of structural risk minimization, which results in the optimization of the error function and exceptional generalization performance. Not only does SVM accurately separate distinct sample groups, but it also exhibits the widest class intervals. In the approach of SVM, 30 juice samples of each variety were randomly selected for the training set and 9 samples for the testing set. There are $30 \times 8 = 240$ samples in the training set and $9 \times 8 = 72$ samples in the testing set. SVM pattern classifier achieves high classification accuracy by providing an interpretation of the learned model.

The EN dataset was recorded by its own pattern recognition software (WinMusterv.1.6.2) for multivariate data analysis. Supervised method Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM) were performed to construct classification models. ANOVA, PCA and LDA were carried out by SPSS 20.0 software (SPSS Inc., Chicago, IL, USA). SVM was conducted by MATLAB R2013a software (Mathworks, USA).

Results and discussion

Electronic nose response to apple juice

Typical response of 10 sensors for apple juice was shown in Fig. 1. G/G_0 represents the sensors' ratio of conductance, where G and G_0 express conductivities of sample gas and zero gas, respectively. Each curve represents a different sensor transient. The conductivity changes with the presence of volatile compounds, so sensor are able to detect changes of different concentrations of samples. As shown in Fig. 1, G/G0 increased sharply in the initial period, reached peak at 7–11 s, fell slowly and then stabilized after 40s.
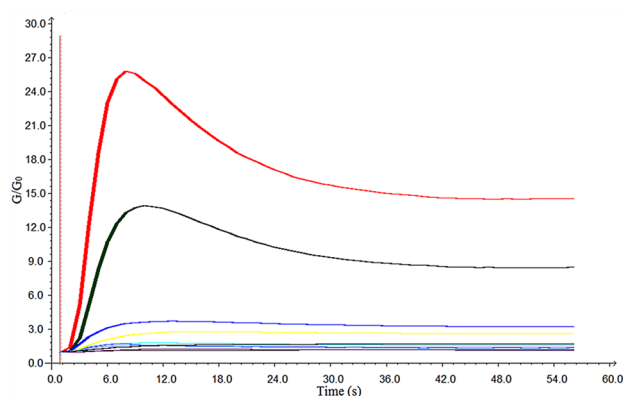
Fig. 1 Typical response curve of ten sensors for apple juice

Optimization of electronic nose measurements and signal processing

In order to find the optimal operative conditions, a set of experiments were performed modifying the following parameters: concentration of volatile compounds and equilibration time. ANOVA was performed to evaluate whether significant influences existed on the response signals of EN analysis of different variances. The effects of the juice volumes (V), equilibration time (T), sensor array (R) and collection time (CT) were evaluated. The results were shown in Table 2.

The equilibration time, sensor array and the collection time had significant effect on responses. The F-values indicated the relative importance of the variances [20]. However, the juice volume had no significant influence on sensor signals. This result showed that the range of juice volumes was enough to reach sample saturation condition. Following experiments were performed at volume of 10 mL.

Optimization of equilibration time

The influence of equilibration times on the response of EN was analyzed by considering the repeatability of the sensors. The response values at 48s (steady state) were selected as the datasets.

Figure 2 (a) showed the mean response values of EN at steady state. As shown, sensor responses at 30 min, 90 min,

Table 2 ANOVA of four variances

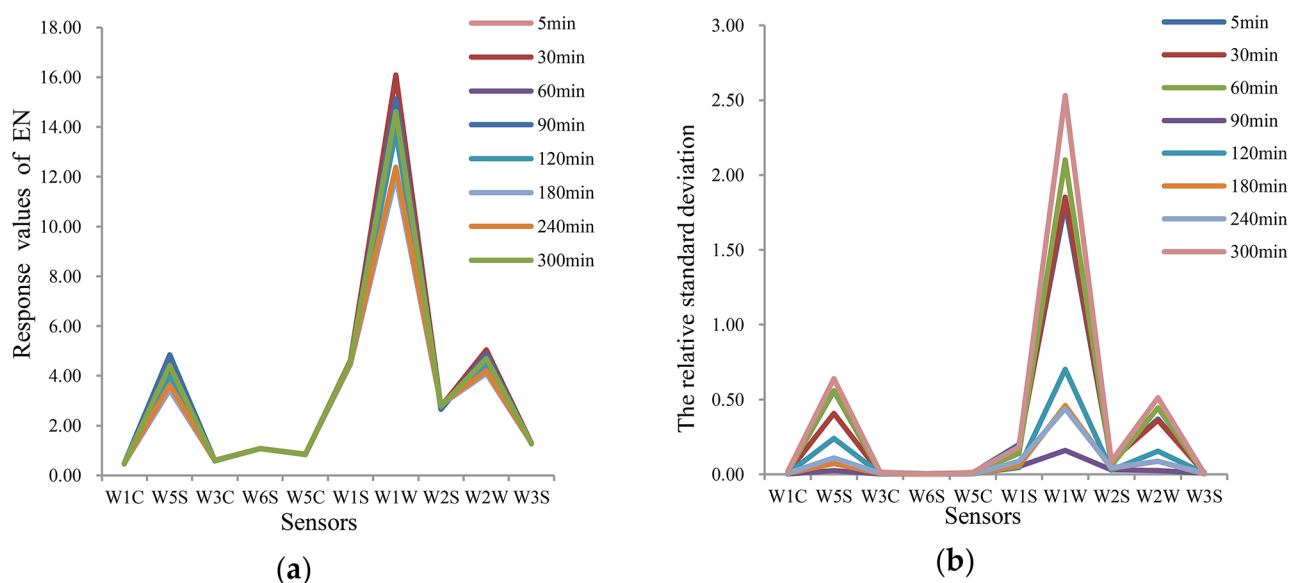| Source of variance	Type III Sum of Squares	df	Mean Square	F	Sig. $p \leqq 0.05$
Corrected Model	8079.846[a]	188	42.978	229.103	.000
Intercept	1449.585	1	1449.585	7727.338	.000
V	0.402	6	0.067	0.357	0.905
T	38.619	7	5.517	29.409	.000
R	1182.873	9	131.430	700.619	.000
CT	2037.649	49	41.585	221.677	.000
Error	35.830	191	0.188		
Total	16075.944	380			
Corrected Total	8115.676	379			

(a)

(b)

Fig. 2 (**a**): Response values of EN at steady state; (**b**): The relative standard deviation of different sensors

Table 3 Multiple comparisons of ten sensors

Sensors	Mean	5% level of significance	1% level of significance
W1C	0.559	a	A
W3C	0.644	a	A
W5C	0.839	ab	AB
W6S	1.090	ab	AB
W3S	1.331	b	B
W2S	2.836	c	C
W2W	3.801	d	D
W1S	3.919	d	D
W5S	4.217	d	D
W1W	8.499	e	E

Sensors in same letters are not remarkable; sensors in different letters are remarkable

60 min and 120 min present relatively higher values than other equilibration times. The response changes as the targeted analytic concentration changes. Meanwhile, the sensitivity of a sensor indicates how much the sensor response [17]. It can be imply that the higher response values the more information of volatile compounds the sensors sensing. Figure 2 (b) showed the relative standard deviation (RSD) values of responses from 10 sensors. The smaller the RSD value, the higher the repeatability of the sensor. As shown, the RSD of response values at 90 min was the smallest, 240 min, 180 min and 120 min followed. Based on the above results, 90 min was selected as the optimal equilibration time. Following experiments were performed at volume of 10 ml and equilibrated 90 min.

Optimization of sensor array

As shown in Table 2, the sensor array had significantly influence on response values of ten sensors according to variety.

Multiple comparisons (Tukey) were carried out to explore whether significant differences existed among ten sensors. The results were shown in Table 3.

The ten sensors are with different selectivity, which are not equally important for certain applications. Due to the natural complexity of volatile compounds of apple juices as well as the broad spectrum response characteristic of sensors, the redundant or irrelevant information was generated inevitably which will greatly affect the discriminant performance. In order to improve the discriminant ability of the sensor array, sensors which are no remarkable in the array may be switched-off or ignored in the analysis. An optimized sensor array with the maximum diversity can be obtained by comparing the correlation and similarity between sensors. Integrated the sensitivities of sensors shown in Table 1 with the result of multiple comparisons, sensors W1C, W5C, W3S, W2S, W5S and W1W were chosen to carry out pattern recognition.

To further validate the optimization results, PCA was employed before and after sensor optimization. These results were compared according to original sensor array and optimized sensor array, separately. PCA plot based on the data of original sensor array was shown in Fig. 3 (a). The first two principal components PC1 and PC2 explain 64.43% and 21.74% of the total variance with value of 86.17%. Two clusters were defined according to the original data matrix. The Qinguan Delicious was distinguished completely, but other varieties overlap with each other. Figure 3 (b) showed the result of PCA which was performed on the data of the optimized sensor array. The PC1 explains 64.29% of the total variance with value of 91.18%. The PC2 explains 26.89% of the total variance. The varieties of Qinguan Delicious, Jonan Golden, Golden Delicious and Fuji
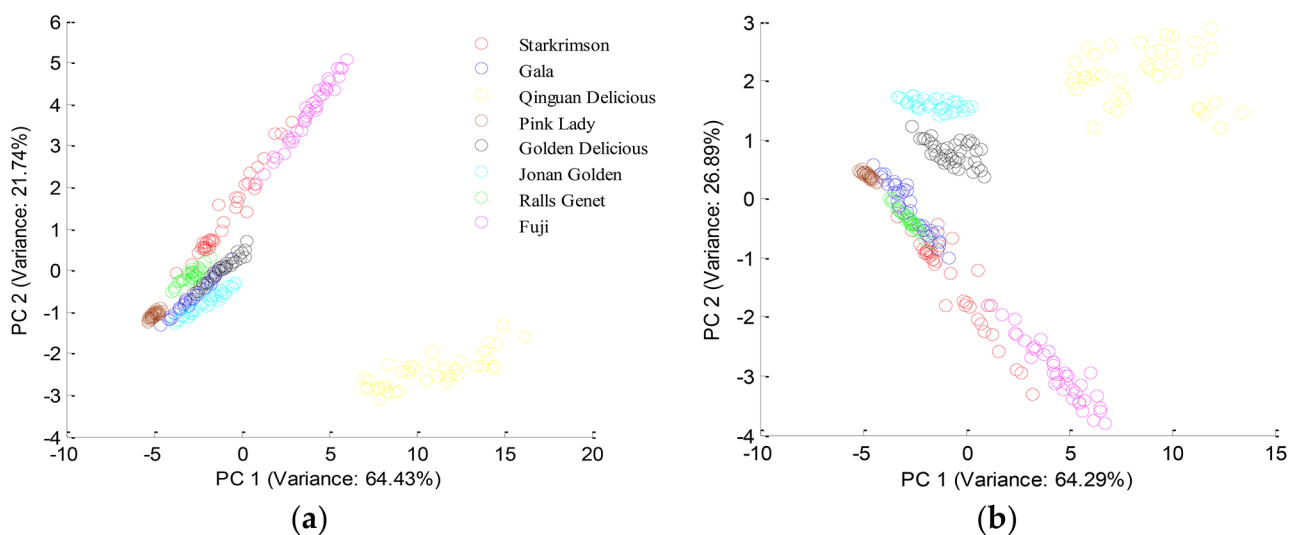
Fig. 3 (**a**): PCA for response of original sensor array; (**b**): PCA for response of optimized sensor array set (W1C, W5C, W3S, W2S, W5S and W1W)

were distinguished well. Although there were still four varieties overlap with each other, the discriminant performance was much better than the result obtained by the original data matrix. This indicated that the sensor optimization indeed improved the extraction of feature information. Therefore, sensor array optimization is necessary to improve pattern recognition performance and eliminate irrelevant information. Meanwhile, the simpler the sensor array, the less time consuming in the application of large number of samples. However, the recognition pattern is not competent for the discrimination of all eight varieties in this study. More accurate feature information needed to be extracted.

Feature extraction from the response curve

The dynamic responses of the whole curve were selected as test signals. The response signals varied with the exposure of analyzed gas compositions. The sensor transient of every 1 s was utilized to explore whether significant influence existed among collection times on discrimination. The corresponding result was displayed in Table 4.

As shown, except 4 s, 5 s… and 18 s, other collection times all have significant influence on EN response signals for the eight varieties. The sizes of the F values indicate the relative importance of the factors [20]. The F values were

increased over time and reached highest at 50 s. To explore whether signals at 50 s perform best on the sample discrimination, PCA was carried out based on data matrix of 1 s, 2 s, 3 s, 19 s, 20 s… and 50s (results not shown). The PCA results demonstrated that the signals at 50 s present relatively strong discriminant capacity. But there was no improvement in comparison with the result shown in Fig. 3 (b), which was obtained by utilizing the EN signals at 48 s. Both F-values and PCA results indicated that the discrimination performance was better utilizing the stationary response signals.

To take all advantages of the response curve and extract more feature information, the response curve was divided into several stages according to the trend showed in Fig. 1. The rising phase, the peak phase, the descent phase and the stationary phase were at 0–7 s, 7–11 s, 12–35 s and 36–50 s, respectively. The PCA results obtained based on datasets of different phases were displayed in Fig. 4.

It can be obviously observed that the discriminant performances of different phases were significantly different. No cluster of samples was observed on the basis of the rising phase dataset. For the peak phase, three clusters were obtained. Both former two systems had not enough resolution to separate sample clearly. Five varieties were clearly discriminated according to the descent phase dataset. But there were still three varieties overlap with each

Table 4 Result of multivariate analysis for the influence of collection times

CT (s)	Wilks' lambda	F	Sig.	CT (s)	Wilks' lambda	F	Sig.
1	0.933	2.386 *	0.022	26	0.925	2.705 * *	0.010
2	0.933	2.367 *	0.024	27	0.922	2.788 * *	0.008
3	0.925	2.694 *	0.011	28	0.920	2.884 * *	0.007
4	0.944	1.950	0.063	29	0.917	2.983 * *	0.005
5	0.958	1.457	0.184	30	0.915	3.094 * *	0.004
6	0.964	1.25	0.276	31	0.911	3.223 * *	0.003
7	0.966	1.181	0.314	32	0.908	3.362 * *	0.002
8	0.966	1.18	0.315	33	0.904	3.521 * *	0.001
9	0.965	1.21	0.298	34	0.900	3.666 * *	0.001
10	0.963	1.264	0.269	35	0.897	3.787 * *	0.001
11	0.961	1.338	0.233	36	0.895	3.908 * *	0.000
12	0.959	1.412	0.201	37	0.892	4.029 * *	0.000
13	0.957	1.491	0.171	38	0.889	4.148 * *	0.000
14	0.955	1.576	0.143	39	0.886	4.265 * *	0.000
15	0.952	1.668	0.118	40	0.883	4.382 * *	0.000
16	0.950	1.762	0.096	41	0.881	4.492 * *	0.000
17	0.947	1.86	0.077	42	0.878	4.593 * *	0.000
18	0.944	1.967	0.060	43	0.876	4.686 * *	0.000
19	0.941	2.073 *	0.047	44	0.874	4.771 * *	0.000
20	0.938	2.173 *	0.037	45	0.872	4.850 * *	0.000
21	0.936	2.279 *	0.029	46	0.871	4.913 * *	0.000
22	0.933	2.379 *	0.023	47	0.869	4.980 * *	0.000
23	0.931	2.466 *	0.019	48	0.868	5.022 * *	0.000
24	0.929	2.551 *	0.015	49	0.867	5.065 * *	0.000
25	0.926	2.630 *	0.012	50	0.865	5.192 * *	0.000

*Significant at level of P ≤ 0.05; **Significant at level of P ≤ 0.01
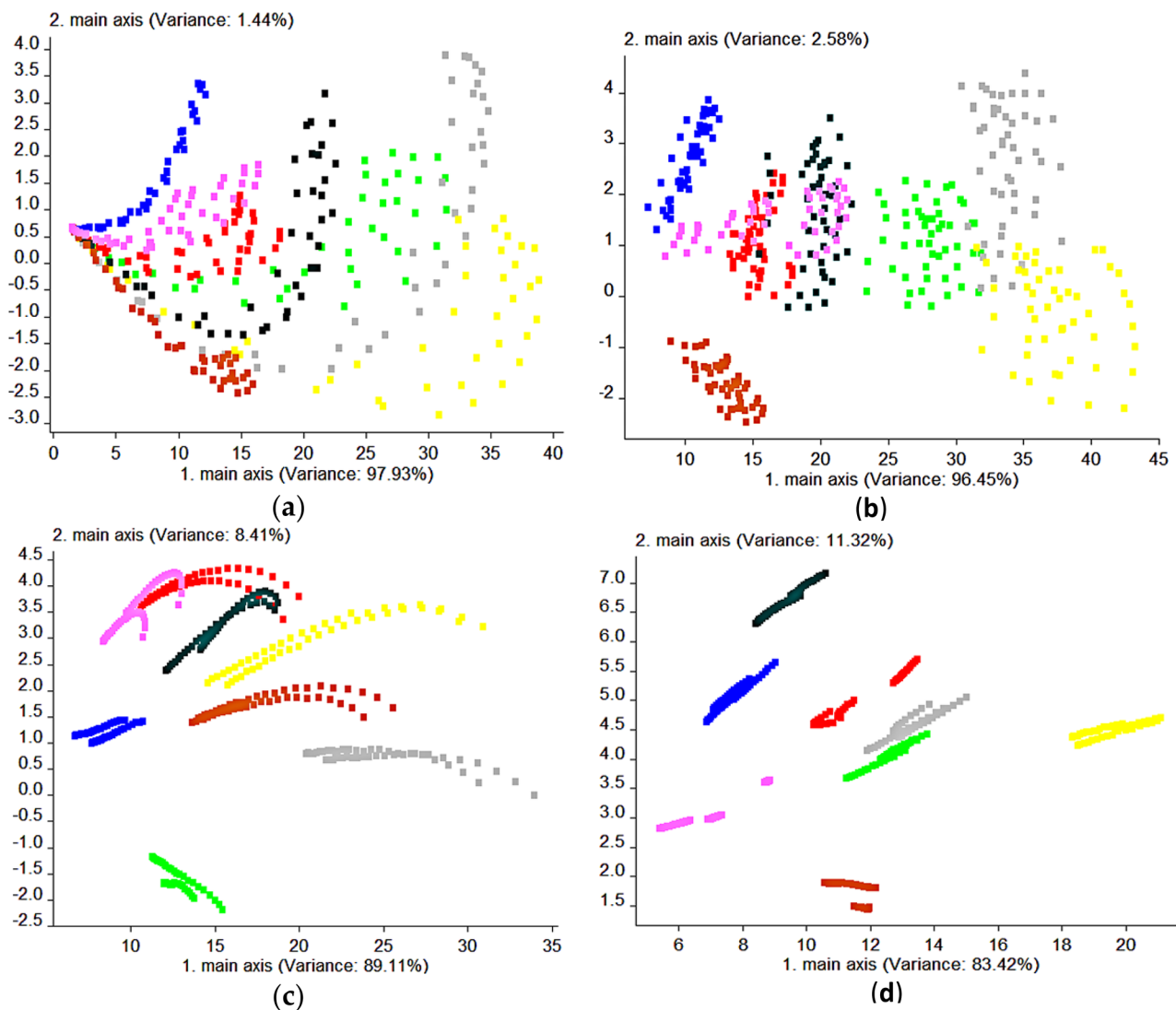
Fig. 4 PCA plots of different phases on the response curve (**a**): The rising phase; (**b**): The peak phase; (**c**): The descent phase; (**d**): The stationary phase

other. Clear separation of the samples into eight clusters was found according to the stationary phase dataset. All samples belonging to eight varieties were distinguished completely. Each cluster had strong convergence. These results indicated that features extracted from the stationary phase present strongest discriminant ability. Compared with the result shown in Fig. 3 (b), the discriminant performance was better. It can imply that better recognition can be obtained using phase feature information than just only one point.

Pattern recognition

The exploratory overview by PCA has already presented in the previous section. PCA pattern had not enough resolution to distinguish all samples based on original dataset. Appropriate selection of sensors and suitable feature extraction improved the performance of PCA recognize pattern, and all samples were discriminated. The discriminant performance was widely used for validation of whether the most relevant features have been extracted. However, the discriminant ability was not only affected by feature information but also the recognition pattern [17]. Not all recognition patterns have equally discriminant ability in certain applications. Thus, in order to improve the discriminant ability of the EN system, we built and compared different patterns.

Classification of LDA

LDA was carried out based on the original dataset. The corresponding results were shown in Table 5. 100% recognition ability and 100% prediction ability were obtained. The pattern showed a very satisfactory discriminant performance.

Table 5 LDA recognition and prediction abilities for juice samples

Variety	Fuji	Golden Delicious	Qin Guan Delicious	Jonah Golden	Pink Lady	Starkrimson	Gala	Ralls Genet
Recognition	100%	100%	100%	100%	100%	100%	100%	100%
Leave-one-out CV prediction	100%	100%	100%	100%	100%	100%	100%	100%

Table 6 Coefficients of sensor variables in fisher linear discriminant functions for sample discrimination according to variety

| LD Function	1	2	3	4
Eigenvalues	455.426	53.533	23.838	20.674
Explained variances (%)	81.1	9.5	4.2	3.7
Canonical correlations	0.999	0.991	0.980	0.977
W1C	-2.104	-2.587	-4.165	9.782
W5S	1.126	-1.254	1.283	-1.270
W3C	1.359	5.351	6.575	-11.046
W5C	2.000	-2.158	-1.680	2.286
W1S	-1.953	0.346	4.292	3.928
W1W	0.851	1.344	2.512	-0.030
W2S	1.657	-0.703	-3.052	-3.495
W2W	-0.495	0.207	-4.079	1.254
W3S	0.272	0.258	-0.114	0.008

LDA extracted information from all sensors, and could maximize the variance between groups and minimize the variance within groups [21]. Therefore, LDA could improve the resolution of recognition pattern and present stronger discriminant ability than utilizing PCA. Meanwhile, the LDA progress also demonstrated the conclusion that better discriminant ability can be obtained based on more information from response curve rather than single point.

The eigenvalues, explained variances, canonical correlations and coefficients of the sensor variables for the first four fisher linear discriminant functions were shown in Table 6.

The coefficients revealed which sensors had greater influence on discrimination. In particular, remarkable importance of the sensors W1C, W5C and W3C was highlighted. As shown in Table 1, all the three sensors are sensitive to aromatic compounds. This implied that aromatic compounds as the most important volatile compounds in apple juice may play a key role for juices variety-based discrimination. Further studies are needed to verify this inference.

SVM for classification

The radial basis function (RBF) was used as the kernel function, and ten-fold cross-validation was applied to get the optimal classification performance. The classification accuracy of SVM is dependent on the magnitude of the parameters C and g [22]. C determines the trade-off between the training error minimization and the model complexity minimization. g was the bandwidth parameter [23]. To improve the classification ability, the key parameters C and g were optimized by a grid search method. The optimizing process was shown in Fig. 5. The optimal combination of C and g was found at the value of $C = 5.6569$ and $g = 2.8284$, respectively.
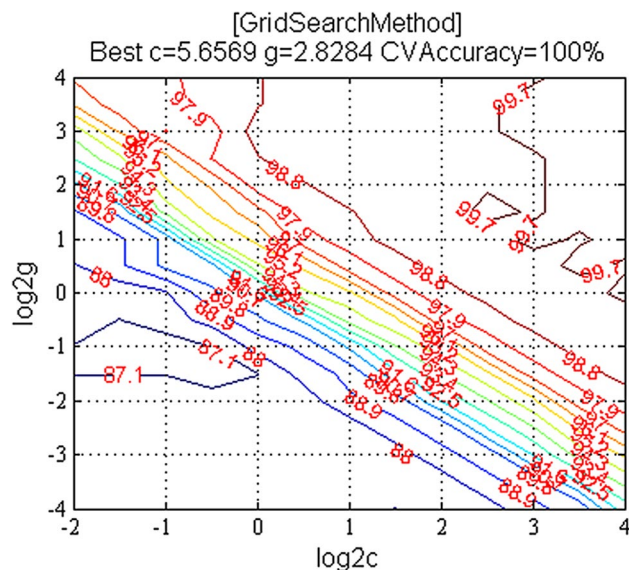
Fig. 5 Grid search on optimal C and g for the SVM model
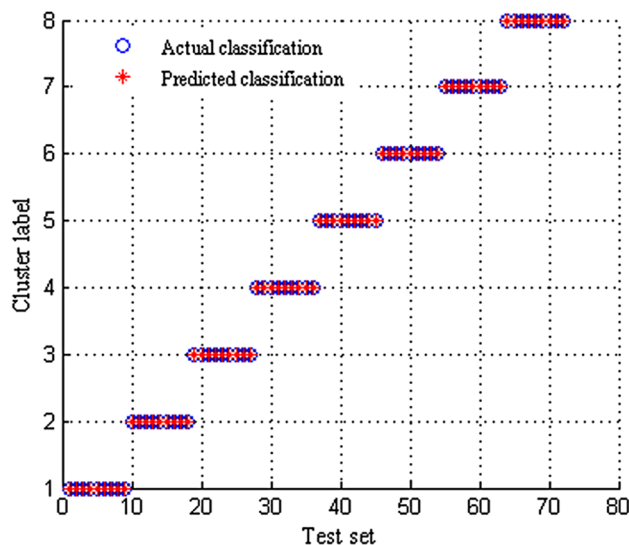
Fig. 6 Actual classification and predicted classification results of the testing set

The accuracy classification rates of training set and testing set were both 100%. The performance of SVM on testing set was shown in Fig. 6. All these results indicated that the SVM model presented satisfactory classification performance on samples according to variety. The SVM based on selected sensor array was also carried out. There was no difference in comparison with the original dataset. But the simpler the model, the less time consume. Applying to large size samples, the sensor

selection will show the great advantage. It can be imply that feature selection can not only extract available information from the multi-sensor array, but also simplify models.

Conclusions

ANOVA was performed and the conclusion that equilibration times, sensor array and collection times had significant effect on signal responses was obtained. The equilibration time was optimized and the optimal experiment parameters were obtained as the volume of 10 ml and 90 min equilibration time. Sensors W1C, W5C, W3S, W2S, W5S and W1W were chosen as optimized sensor array to carry out PCA-based pattern recognition. Compared with original sensor array, sensor optimization indeed improved the recognition performance of PCA. The sensor transient of every 1 s and trend stages signals were utilized to explore the influence of collection times on discrimination. The strongest discriminant ability was presented by utilizing the stationary phase signals. PCA-based pattern had not enough resolution to distinguish all samples. LDA was carried out and 100% recognition ability and 100% prediction ability were obtained. Remarkable important sensors W1C, W5C and W3C were highlighted in the LDA model. SVM model obtained both 100% accuracy classification rates of training set and testing set. Both LDA and SVM pattern recognitions presented satisfactory classification performance.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

1. Y. Sun, L. Zhong, L. Cao, W. Lin, X. Ye, Sonication inhibited browning but decreased polyphenols contents and antioxidant activity of fresh apple (malus pumila mill, cv. Red Fuji) juice. J. Food Sci. Technol. **52**(12), 1–7 (2015)
2. M.P. Nikfardjam, D. Maier, Development of a headspace trap HRGC/MS method for the assessment of the relevance of certain aroma compounds on the sensorial characteristics of commercial apple juice. Food Chem. **126**(4), 1926–1933 (2011)
3. M. Vanoli, C. Visai, A. Rizzolo, The influence of harvest date on the volatile composition of 'Starkspur Golden' apples. Postharvest Biology Technol. **6**(6), 225–234 (1995)
4. J. Dixon, E.W. Hewett, Factors affecting apple aroma/flavour volatile concentration: a review. New. Z. J. Crop Hortic. Sci. **28**(3), 155–173 (2000)
5. C.M. Braga, A.A.F. Zielinski, F.K.F. de Silva KMd, G.A.M. Pietrowski, M. Couto, D. Granato, G. Wosiacki, A. Nogueira, Classification of juices and fermented beverages made from unripe, ripe and senescent apples based on the aromatic profile using chemometrics. Food Chem. **141**(2), 967–974 (2013)
6. G.V. Civille, K.N. Oftedal, Sensory evaluation techniques — make good for you taste good. Physiol. Behav. **107**(4), 598–605 (2012)
7. X. Hong, J. Wang, Use of Electronic nose and tongue to Track freshness of Cherry tomatoes squeezed for Juice Consumption: comparison of different Sensor Fusion approaches. Food Bioprocess. Technol. **8**(1), 158–170 (2014)
8. T. Arroyo, J. Lozano, J.M. Cabellos, M. Gil-Diaz, J.P. Santos, C. Horrillo, Evaluation of wine aromatic compounds by a sensory human panel and an electronic nose. J. Agricultural Food Chem. **57**(24), 11543–11549 (2009)
9. S. Buratti, S. Benedetti, M. Scampicchio, E.C. Pangerod, Characterization and classification of Italian Barbera wines by using an electronic nose and an amperometric electronic tongue. Anal. Chim. Acta. **525**(1), 133–139 (2004)
10. A.M. Spanier, J.C. Beaulieu, K.L. Bett, K. Gross, Use of Electronic Nose Technology to Examine Apple Quality *; 1999
11. R.N. Bleibaum, H. Stone, T. Tan, S. Labreche, E. Saint-Martin, S. Isz, Comparison of sensory and consumer results with electronic nose and tongue sensors for apple juices. Food Qual. Preference. **13**(6), 409–422 (2002)
12. P.H. Heinemann, R.E. Crassweller, E. Leblanc, W.N. Marrazzo, ELECTRONIC NOSE CHEMICAL SENSOR FEASIBILITY STUDY FOR THE DIFFERENTIATION OF APPLE CULTIVARS, Trans. Asae. **48**(5), p–gs (2005)
13. R.A. Saftner, R.E. Crassweller, E. Leblanc, W.N. Marrazzo, P.H. Heinemann, Electronic nose chemical sensor versus gas chromatography: a feasibility study for the differentiation of apple flavors and essences. Trans. Asae. **48**(5), p–gs (2005)
14. K. Karlshøj, P.V.N. And, T.O. Larsen, Prediction of Penicillium Expansum Spoilage and Patulin Concentration in apples used for Apple Juice Production by electronic nose analysis. J. Agricultural Food Chem. **55**(11), 4289–4298 (2007)
15. S. Cagnasso, M. Falasconi, M.P. Previdi, B. Franceschini, C. Cavalieri, V. Sberveglieri, P. Rovere, Rapid Screening of Alicyclobacillus acidoterrestris Spoilage of Fruit Juices by Electronic nose: a confirmation study. J. Sens. **2010**(1687–725X), 23–59 (2010)
16. S. Roussel, G. Forsberg, P. Grenier, V. Bellon-Maurel, Optimisation of electronic nose measurements. Part II: influence of experimental parameters. J. Food Eng. **39**(1), 9–15 (1999)
17. Z. Xu, X. Shi, S. Lu, Integrated sensor array optimization with statistical evaluation. Sens. Actuators B Chem. **149**(1), 239–244 (2010)
18. S. Roussel, G. Forsberg, V. Steinmetz, P. Grenier, V. Bellonmaurel, Optimisation of electronic nose measurements. I. Methodology of output feature selection. J. Food Eng. **37**(2), 207–222 (1998)
19. J.W. Gardner, P. Boilot, E.L. Hines, Enhancing electronic nose performance by sensor selection using a new integer-based genetic algorithm approach. Sens. Actuators B **106**(1), 114–121 (2005)
20. H. Yu, J. Wang, Discrimination of LongJing green-tea grade by electronic nose. Sens. Actuators B Chem. **122**(1), 134–140 (2007)
21. A.A.F. Zielinski, C.W.I. Haminiuk, C.A. Nunes, E. Schnitzler, S.M. Ruth, D. Granato, Chemical composition, sensory properties, Provenance, and Bioactivity of Fruit Juices as assessed by Chemometrics: a critical review and Guideline. Compr. Rev. Food Sci. Food Saf. **13**(3), 300–316 (2014)
22. G.M. Foody, A. Mathur, Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification. Remote Sens. Environ. **93**(1–2), 107–117 (2004)
23. J. Dong, W. Guo, Z. Wang, D. Liu, F. Zhao, Nondestructive determination of soluble solids content of 'Fuji' apples produced in different areas and bagged with different materials during ripening. Food. Anal. Methods. **9**(5), 1087–1095 (2015)