Mathematical Biosciences and Engineering

MBE, 19(9): 9079–9097.
DOI: 10.3934/mbe.2022422
Received: 07 March 2022
Revised: 09 June 2022
Accepted: 13 June 2022
Published: 22 June 2022

http://www.aimspress.com/journal/MBE

Research article

Discrimination and chemical composition quantitative model of Raw Moutan Cortex and Moutan Cortex Carbon based on electronic nose and machine learning

Sujuan Zhou[1,2,†], Huajian Lin[3,†] and Jiang Meng[3,*]

[1] College of Medical Information Engineering, Guangdong Pharmaceutical University, Guangzhou 510006, China
[2] Department of Automation, Guangdong University of Technology, China
[3] School of Traditional Chinese Medicine, Guangdong Pharmaceutical University /Key Laboratory of Digital Quality Evaluation of Chinese Materia Medica, State Administration of Traditional Chinese Medicine (TCM) /Engineering Technology Research Center for Chinese Materia Medica Quality of Universities in Guangdong Province, Guangdong 510006, China

† These authors contributed equally to this work

*** Correspondence:** Email: jiangmeng666@126.com; Tel: +8602039352169; Fax: +86 02039352174.

Abstract: Raw Moutan Cortex (RMC) is a traditional medicinal material commonly used in China. Moutan Cortex Carbon (MCC) is a processed product of RMC by stir-frying. As raw and processed products of the same Chinese herb pieces, they have different effects. RMC has the effects of clearing heat and cooling blood, promoting blood circulation and removing blood stasis, but MCC has the contrary effect of cooling blood and hemostasis. Therefore, it is necessary to distinguish them effectively. The traditional quality evaluation method of RMC and MCC still adopts character identification, and mainly relies on the working experience and sensory judgment of employees with experience. This will lead to strong subjectivity and poor repeatability. And the final evaluation result may cause inevitable errors and the processed products with different processing degrees in actual production，which affects the clinical efficacy. In this study, the electronic nose technology was introduced to objectively digitize the odor of RMC and MCC. And the discrimination model of RMC and MCC was constructed in order to establish a rapid, objective and stable quality evaluation method of RMC and MCC. According to the correlation analysis, the experiment found the content of gallic

acid, 5-hydroxymethylfurfural (5-HMF), paeoniflorin and paeonol determined by high performance liquid chromatography (HPLC) had a certain correlation with their odor characteristics. Thus, partial least squares regression (PLSR) and support vector machine regression (SVR) were compared and established the chemical composition quantitative model. Results showed that the quantitative data of RMC and MCC odor could be used to predict the contents of the chemical components. It can be used for quality control of RCM and MCC.

Keywords: Moutan Cortex; Moutan Cortex Carbon; discrimination model; composition quantitative model; electronic nose; support vector machine regression; HPLC

1. Introduction

Raw Moutan Cortex (RMC) is the dry root bark of Paeonia suffruticosa Andr [1]. It was first published in shennong herbs classic [2,3]. It is a traditional medicine commonly used in China. It tastes bitter, spicy and slightly cold, and has the effect of clearing heat and cooling blood, promoting blood circulation and removing stasis. Moutan Cortex charcoal is the processed product of RMC. It has the function of cooling blood and stopping bleeding. In the actual processing process of MCC, due to the different production conditions, equipment and people's subjective judgment and other reasons, the processing results are often too excessive or not reach the standard, so as to obtain the different processed product, such as light MCC (LMCC), standard MCC (MCC), and heavy MCC (HMCC). Underprocessing or overprocessing will affect the effectiveness of drugs, and then affect the clinical efficacy.

At present, the quality control of MCC mainly adopts the traditional quality evaluation method-feature recognition, namely [4] "Quality evaluation based on feature recognition". Chinese Pharmacopoeia and local Chinese medicine treatment specifications [5] describe MCC as "dark brown on the outer surface, brown on the inner surface, with a burnt aroma, slightly bitter and astringent taste". However, feature recognition often depends on the experience of practitioners and people's sensory judgment, which is often easily affected by subjective feelings. The final evaluation may cause inevitable error and poor repeatability, and the processed products with different processing degrees in actual production. Now, many research have studied about the Chemical changes in RMC and its different products, these results showed that most compounds were decreased with the deepening of processing degree and the increase of temperature, such as catechin, paeonol, quercetin, Kaempferide, isorhamnetin and tannin, While the content of gallic acid and 5-HMF were firstly increased with the extension of processing time and then began to decline [6–8]. And at the same time the pharmacodynamics studied showed that tannins such as catechin have astringent and hemostatic effects [9–11]. Paeonol had the effect of promoting blood circulation and removing blood stasis [12,13]. 5-HMF was an aldehyde produced by dehydration of glucose and other monosaccharide compounds under high temperature or weak acid conditions. As the temperature continues to rise, 5-HMF is easily decomposed into levulinic acid and formic acid, which was a marker of heating process [14,15]. Chemical components are the material basis of pharmacodynamic effects. Different chemical components in different processing degrees will certainly cause the change of effects, and then affect the clinical efficacy. Now, high performance liquid chromatography (HPLC), gas chromatography-mass spectrometry (GC-MS) [16], thin layer chromatography (TLC) and other analytical methods [17,18]

have also used for quantitative quality identification of MCC. Although these methods have great advantages in detection accuracy, they also have some disadvantages, such as sample destruction, large amount of chemical reagents consumption, long analysis time, and difficulty in obtaining sample quality evaluation results quickly in the production process. Therefore, it is necessary to establish a fast, nondestructive and sensitive method to evaluate and control the quality of RMC and MCC.

Electronic nose is an electronic sensing instrument. It simulates human's sense of smell to obtain sample odor data, and performs objective digital processing of odor [19,20]. It is mainly composed of gas sensor array, signal processing unit and pattern recognition [21]. It has the advantages of simple sample pretreatment, convenient operation and fast reaction speed. The quality, authenticity and processing degree of Traditional Chinese medicine determine the characteristics and intensity of odor to a certain extent. On the other hand, the odor of traditional Chinese medicine is directly related to its internal chemical composition, which can reflect the internal nature and become the correlation point between external quality performance and internal material basis. For example, MOS Type Sensor-Array and machine learning were used to classify and identify the Potato Cultivars [22]. MAU-9 electronic-nose MOS sensor array components and ANN classification were used to discriminate the herb and fruit essential oils [23]. Opto-electronic nose coupled to a Silicon Micro Pre-Concentrator Device were used to select sensing of flavored waters [24].

In this study, electronic nose and machine learning were used to discriminate and quantitative analysis chemical composition of RMC and MCC. Firstly, HPLC was used to determine the contents of gallic acid, 5-hydroxymethylfurfural, paeoniflorin and paeonol in different processing levels of MCC. Secondly, Electronic nose was used to determine the smell information of different processing degrees of MCC. Then PCA, SVM and other methods were used for qualitative identification. Finally, the PLSR and SVR quantitative models were compared and analysis the content of galic acid, 5-hydroxymethylfurfural, paeoniflorin and paeonol in RMC and MCC. It provided a rapid, simple and non-invasive monitoring method to quality evaluation the RMC and MCC.

2. Materials and methods

2.1. Officinal material

27 batches of RMC pieces were collected and purchased from the pharmaceutical companies all over the country. It was identified by Associate Professor Liu Jizhu, School of traditional Chinese medicine, Guangdong Pharmaceutical University. Voucher specimens were deposited at the Herbarium Centre, Guangdong Pharmaceutical University. Part of each batch was processed at 180°C for 3–5, 6–8 and 9–11 min respectively, and different processing degrees products were achieved, including Light Carbon (LMCC), Standard Carbon (SMCC) and Heavy Carbon (HMCC).[6]. All samples were crushed by a high-speed multifunctional grinder (JP-150A, Jiupin Industry and Trade Co., LTD., Yongkang, China) and then passed through an 80-mesh sieve, and then dried at 45℃ and sealed for preservation. All the samples were 108 batches and were summarized in Table 1.

The reference standards Gallic acid (Batch No: CHB180114), 5-HMF (Batch No: CHB180118) and Paeoniflorin (Batch No: CHB190124) (purity≥98% for each) were purchased from Chengdu Chroma-Biotechnology Co., Ltd. (Sichuan, China). Chromatographic grade methanol was from Oceanpak Alexative Chemical., Ltd. (Sweden). Ultrapure water was used in the whole experiment (Watson, China).

Table 1. Sample information of RMC and MCC.

Sample number				batch number	producing areas
RMC1	LMCC1	SMCC1	HMCC1	YPA9A0001	Anhui
RMC2	LMCC2	SMCC2	HMCC2	YPA8J0001	Anhui
RMC3	LMCC3	SMCC3	HMCC3	YPA7H0001	Sichuan
RMC4	LMCC4	SMCC4	HMCC4	YPA8H0001	Sichuan
RMC5	LMCC5	SMCC5	HMCC5	YPA9C0001	Hebei
RMC6	LMCC6	SMCC6	HMCC6	181100019	Anhui
RMC7	LMCC7	SMCC7	HMCC7	170901	Anhui
RMC8	LMCC8	SMCC8	HMCC8	181201	Anhui
RMC9	LMCC9	SMCC9	HMCC9	190101	Anhui
RMC10	LMCC10	SMCC10	HMCC10	180600159	Anhui
RMC11	LMCC11	SMCC11	HMCC11	20190515	Anhui
RMC12	LMCC12	SMCC12	HMCC12	180600029	Anhui
RMC13	LMCC13	SMCC13	HMCC13	1990101	Anhui
RMC14	LMCC14	SMCC14	HMCC14	181201	Sichuan
RMC15	LMCC15	SMCC15	HMCC15	190401	Anhui
RMC16	LMCC16	SMCC16	HMCC16	20180701	Anhui
RMC17	LMCC17	SMCC17	HMCC17	190409	Anhui
RMC18	LMCC18	SMCC18	HMCC18	190303	Anhui
RMC19	LMCC19	SMCC19	HMCC19	190415	Anhui
RMC20	LMCC20	SMCC20	HMCC20	190235	Anhui
RMC21	LMCC21	SMCC21	HMCC21	180811	Anhui
RMC22	LMCC22	SMCC22	HMCC22	180928	Anhui
RMC23	LMCC23	SMCC23	HMCC23	181117	Anhui
RMC24	LMCC24	SMCC24	HMCC24	HX19K01	Anhui
RMC25	LMCC25	SMCC25	HMCC25	190439	Anhui
RMC26	LMCC26	SMCC26	HMCC26	184902	Anhui
RMC27	LMCC27	SMCC27	HMCC27	201904	Hebei

2.2. Instrument for chemical composition content determination

The contents of gallic acid, 5-HMF, paeoniflorin and paeonol in RMC and different processed MCC were determined by HPLC. The instruments used include: Shimadzu high performance liquid chromatograph, equipped with LC-20AT binary pump, SPD-M20A detector, SIL-20A injector (Shimadzu, Japan),1/100000 electronic balance (sartorius, Germany), KQ-300DE ultrasonic cleaning instrument (Shanghai Lingke).

2.3. E-Nose equipment and measurements

Samples were detected by portable e-nose PEN3 (Airsense Analytics, Schwerin, Germany), which, with the built-in sensor array, sampling and cleaning channels, and data acquisition system, is characterized by its automatic adjustment, calibration, and system enrichment functions [25].

The sensor array is composed of ten MOS sensors sensitive to different compounds, the sensitive characteristics of each sensor was listed in Table 2. The sensor response value was relative resistivity G / G0 (G represented the resistance of the sensor after the action of volatile gas in the sample to be tested, G0 represented the resistance of the sensor after the action of reference gas filtered by standard active carbon). The electronic nose device is shown in Figure 1.

Figure 1. Experimental instrument of PEN-3.

Table 2. Gas sensor array of PEN-3.

Number	Name of sensor	Detection of chemical components
S1	W1C	Aromatic
S2	W5S	Nitrogen Oxides
S3	W3C	Ammonia, aromatic
S4	W6S	hydrogen
S5	W5C	Alkanes, aromatic ingredients
S6	W1S	Methane
S7	W1W	Sulfide
S8	W2S	Ethanol
S9	W2W	Aromatic ingredients, organic sulfur compounds
S10	W3S	Alkanes

2.4. Determination method of chemical composition

2.4.1. Preparation of test solution

Accurately weigh 0.5 g of RMC and MCC with different processing degrees (passing 80 mesh sieve), place it in a conical flask with a stopper, accurately add 25 mL of 50% methanol, weigh and extract with ultrasound (power 100W) for 30 min, cool to room temperature, weigh, then make up the weight loss with 50% methanol, filter, and take the continuous filtrate as the test solution.

2.4.2. Preparation of reference solution

The standard substances of gallic acid, 5-HMF, paeoniflorin and paeonol were accurately weighed and dissolved with 50% methanol to obtain the stock solutions at the concentrations of 0.7000, 1.000 and 2.000 mg/mL respectively. The working standard solutions, after prepared by mixing and diluting the stock solutions with methanol, were filtered through a 0.22 μm PTEE filter. The stock solutions and working solutions were stored at 4 °C for further use.

2.4.3. Chromatographic conditions

The separation was performed on an Ultimate TM XB-C 18 analytical column (250 mm × 4.6 mm, 5 μm) at 30 °C. The mobile phase consisted of a mixture of 0.1% phosphoric acid in water (A)-acetonitrile (B). A gradient program was set as follows: 0–15 min, 5–10% B; 15–25 min, 10–24% B; 25–50 min, 24–39% B; 50–65 min, 39–50% B; 65–90 min, 50–53% B; 90–95 min, 53–95% B. The flow rate was 1.0 mL/min and the detection wavelength was 230 nm, 10 μL of the working solution or the sample solution was injected for HPLC analysis.

2.4.4. Methodology validation.

The linearity of the HPLC method for each analyte was evaluated by calibration curves. Each analyte at a series of different concentrations was analyzed in triplicates. The linearity of the calibration curve was constructed by plotting the peak area ratios vs. the concentration of four components. The precision of the HPLC method was determined by intraday and interday measurements. The working standard solution was analyzed in six replicates on the same day to obtain the intraday precision while the interday precision was obtained by analyzing the working standard solution daily (six replicates) for three successive days. Meanwhile, the stability was assessed by analyzing the same sample solution (LMCC4) at 0, 3, 6, 9, 12, and 24 h, respectively. Besides, recovery tests (LMCC4) were performed according to Chinese pharmacopeia to investigate the accuracy of the developed HPLC method. Mixed standard solutions at the uniform concentration level (100%) were added into 0.5 g of the known real samples, and each solution was done three copies in parallel according to the proposed HPLC method. The results were expressed as relative standard deviation (RSD %) of the measurements.

2.5. Data acquisition of electronic nose

2.5.1. Experimental pretreatment

The different batch of RMC and MCC were firstly put them in the quartz container separately, and then sealed the quartz container with double-layer fresh-keeping film. Before each test, let the samples stand for 30 min to fill the whole quartz container with volatile smell; Warmed up the machine and flushed the metal sensor of electronic nose for 300 s before detection.

2.5.2. Detecting parameters

The electronic nose was connected to the computer, and the corresponding curve of the sample sensor was obtained in real time by Winmaster workstation. After the sample standed for 30 min, insert the injection needle of the electronic nose was inserted into the fresh-keeping film and fixed it, and

inhaled the sample gas to be tested at a flow rate of 150 mL/min. The intake air was passing through activated carbon. Each sampling time is 120 s, the sampling interval is 1 s, and the sensor cleaning time is 120 s. when measuring the sample, the ambient temperature and humidity are controlled at about 25°C and 30% respectively. Each batch of samples were measured three times in parallel, and the average response curve was taken as the test data of the samples, and the odor response value matrix of 108 batches of samples was obtained.

2.6. Method validation

Precision of the method was determined by intraday and interday measurements. The sample Powder (LMCC14) was analyzed in six replicates on the same day to obtain intraday precision, and they were analyzed daily (six replicates) for three successive days to obtain the interday results. The stability was assessed by analyzing the same sample powder (LMCC14) at 0, 2, 4, 6, 8 h, respectively.

2.7. Data analyzing and Statistical tests

Data analysis was completed in the MATLAB 2020 environment, qualitative analysis using Classification toolbox 5.2. PCA was performed with SIMCA-P + 12.0 software. The significance test was carried out by two-tailed test in this paper.

3. Experimental results and analysis

3.1. Determination results of chemical composition

3.1.1. Methodology validation

The results of the methodology validation for HPLC analysis were shown in Table 3 and Figure 2. The calibration curves of each analyte displayed good linearity over the range ($R^2 > 0.9997$) of different concentrations. The RSD values of the precision test were 0.10–2.74% for intraday assays and 0.52–1.64% for interday assays. The RSD values of stability tests were 0.14–2.79%. The recoveries of the HPLC method were above 96.94%, and the RSD values were less than 3.0%. The results demonstrated that the developed HPLC method was capable of accurately determining the contents of the twelve chemical ingredients in different RMC and MCC samples.

3.1.2 Sample analysis

The developed HPLC method was applied to simultaneously determine the contents of the chemical ingredients in RMC and MCC. The results were shown in Table 4. There was a significant difference in the contents of the four chemical ingredients between RMC and different products of MCC. The contents of paeoniflorin and paeonol in RMC were higher than the different degree process product of MCC. And their contents were declined with the extension of processing time. While the content of gallic acid and 5-HMF were the highest among the RMC and MCC, at the same time they were firstly increased and then declined with the processing time. This results was consisted with previous studies [6].

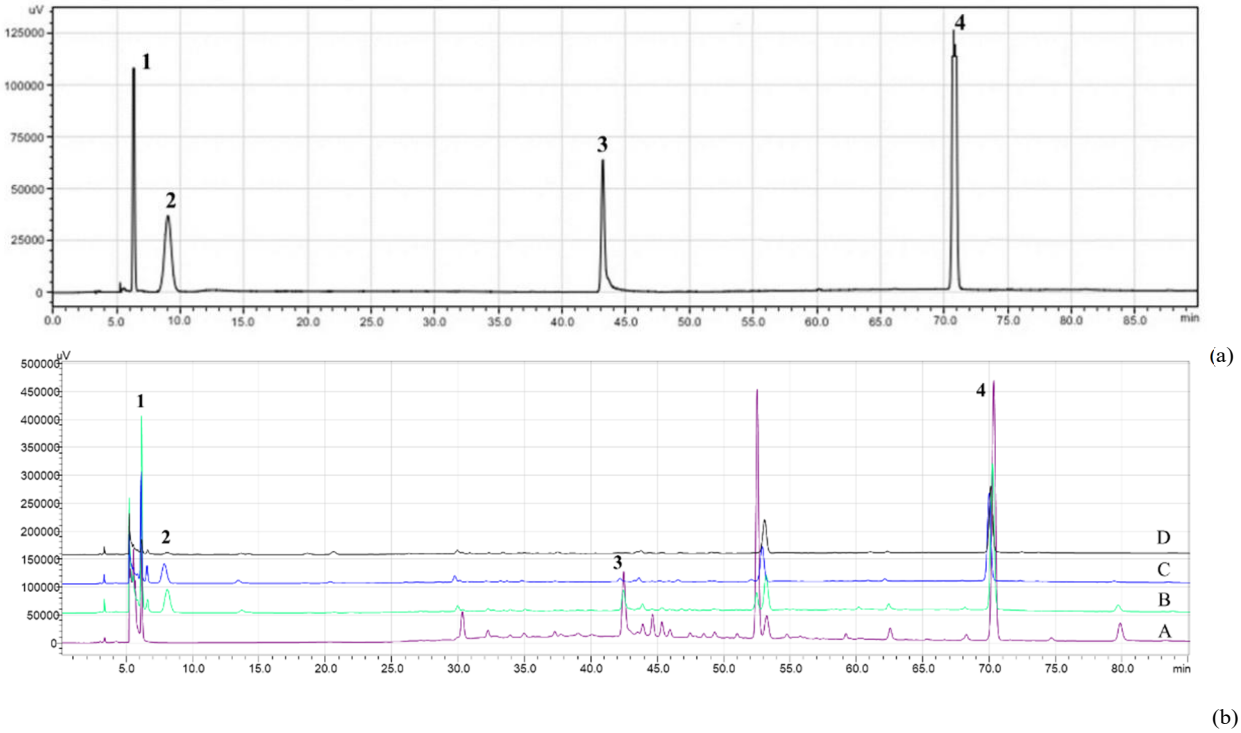

Figure 2. HPLC chromatogram.

**Note*: a. HPLC Chromatogram of reference substance; b. HPLC Chromatogram of samples;

1. gallic acid; 2. 5-HMF; 3. paeoniflorin; 4. paeonol; A. RMC; B. LMCC; C. SMCC; D. HMCC

Table 3. The results of methodology validation for HPLC analysis.

composition	regression equation/ R^2	linear rang (μg/mL)	Recovery (%)		Precision/ RSD (%)		Repeatability RSD (%)
			mean	RSD	Intra-day	Inter-day	
gallic acid	$Y = 26797X + 40702$ R^2=0.9999	1.75~134.75	96.94	0.02	2.20	2.22	2.19
5-HMF	$Y = 11417X + 6233.3$ R^2=0.9999	3.00~125.00	100.55	0.55	0.36	1.04	2.22
paeoniflorin	$Y = 13372X - 20305$ R^2=0.9997	0.50~190.5 0	100.47	1.15	2.74	2.79	0.52
paeonol	$Y = 31918X - 49375$ R^2=0.9999	20.00~400.00	98.83	1.85	0.10	0.14	1.59

3.2. Results of electronic nose data processing

3.2.1. Methodological investigation

The RSD values of precision test were 0.44–2.51% for intraday assays and 1.84–3.14% for

interday assays. The RSD values of stability test were 0.68–4.21% (Table 5). The system was considered suitable for analysis of RMC and MCC.

Table 4. Average content of chemical components in RMC and MCC (N = 27).

	gallic acid (mg/g)	5-HMF (mg/g)	paeoniflorin (mg/g)	Paeonol (mg/g)
RMC	2.13 ± 0.48	0	8.31 ± 1.28	12.09 ± 3.28
LMCC	4.41 ± 0.78	3.86 ± 1.00	2.78 ± 0.97	9.26 ± 1.51
SMCC	2.60 ± 0.89	2.97 ± 1.10	0.55 ± 0.36	6.75 ± 1.39
HMCC	0.29 ± 0.19	0.38 ± 0.18	0.16 ± 0.02	2.78 ± 0.71

Table 5. Investigation on the precision of electronic nose sensor.

	RSD (%)				*RSD* (%)		
sensor	Intraday (n = 6)	Interday (n = 6)	sensor	*Stability RSD (%)*	Intraday (n = 6)	Interday (n = 6)	*Stability RSD (%)*
W1C	0.84	2.13	W1S	0.96	2.34	2.13	4.21
W5S	1.11	3.14	W1W	3.42	0.46	2.01	3.18
W3C	1.06	1.98	W2S	0.83	2.51	2.09	4.01
W6S	0.44	1.84	W2W	0.68	0.67	1.87	4.18
W5C	0.67	2.76	W3S	0.81	1.09	2.01	0.93

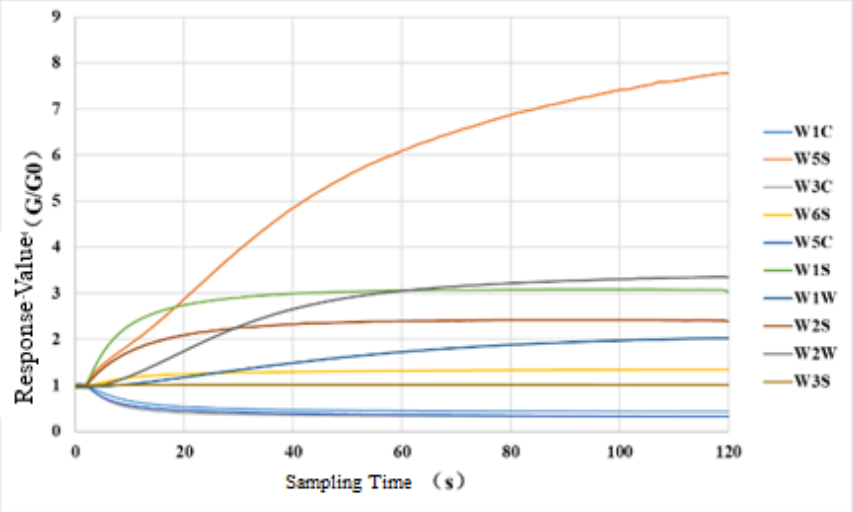

Figure 3. Response curve of electronic nose sensor.

3.2.2. Response curve of odor sensor

Figure 3 showed the odor response curve of 10 sensors, using electronic nose on one sample within 120 seconds. It could be seen from the figure that the change of the sensor response value increased gradually and then tended to be gentle. This was because during headspace injection, the

concentration of volatile substances in the sample entering the sensor channel increases continuously and finally reaches dynamic equilibrium.

3.2.3. Response value of odor sensor

The radar diagram of the sample odor sensor using the response value of each sensor were construct when it reached equilibrium (Figure 4). From the radar diagram, it could be seen that the strongest sensor of RMC was W5S (nitrogen oxide), followed by W1S (methyl) and W2W (organic sulfide), indicating that the volatile gas substances of RMC were mainly nitrogen oxide, methyl and organic sulfide.

The response values of sensors W1C (aromatic components), W3C (ammonia) and W5C (aromatic alkanes) changed little, while the response values of sensors W5S (nitrogen oxides), W1S (methyl), W1W (sulfide), W2S (alcohols) and W2W (organic sulfide) changed greatly, which revealed that nitrogen oxides, methyl, sulfide, alcohols and organic sulfide were the differential compounds of odor between RMC and MCC. On the whole, the difference of sensor response values between carbon products with different processing degrees was small. It was difficult to distinguish carbon products with different degrees by radar map alone, and other discrimination methods needed to be used for further analysis.

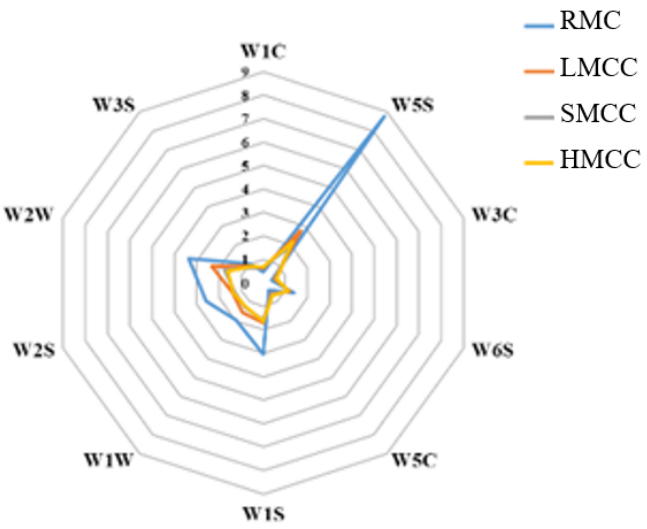

Figure 4. Radar diagram of odor sensor of RMC and MCC.

3.2.4. Optimization of sensor array

The sensitivity of sensors to gas was partially crossed and relatively nonspecific, so collinearity and other problems may occur between some sensors. In order to reduce the miscellaneous information between sensor arrays and the complexity of high-dimensional data on the model, Pearson correlation analysis was used to calculate the correlation coefficient between the two gas sensors by taking the response values of 10 sensors as variables. When the correlation coefficient value of the two sensors was larger, it proved that the correlation of the two sensors was better, and the consistency of the

information obtained was closer. So, the two sensors could replace each other, and one of them could be considered to eliminated.

Table 6. Correlation analysis results of electronic nose sensors in RMC and MCC

	W1C	W5S	W3C	W6S	W5C	W1S	W1W	W2S	W2W	W3S
W1C	1	-0.842	**0.984**	-0.827	0.934	-0.921	-0.900	-0.924	-0.919	-0.805
W5S		1	-0.822	0.752	-0.786	0.855	0.898	0.875	0.887	0.801
W3C			1	-0.782	**0.971**	-0.867	-0.892	-0.884	-0.91	-0.734
W6S				1	-0.702	0.851	0.800	0.842	0.830	0.904
W5C					1	-0.776	-0.886	-0.816	-0.899	-0.621
W1S						1	0.810	**0.980**	0.830	0.884
W1W							1	0.813	**0.995**	0.762
W2S								1	0.833	0.875
W2W									1	0.780
W3S										1

From Table 6, it could be seen that the correlation coefficient of W1C and W3C, W3C and W5C, W1S and W2S, W1W and W2W were large, namely 0.984, 0.971, 0.980 and 0.995 respectively. Therefore, W1W, W5C, W1W and W2S sensors were eliminated in the subsequent analysis.

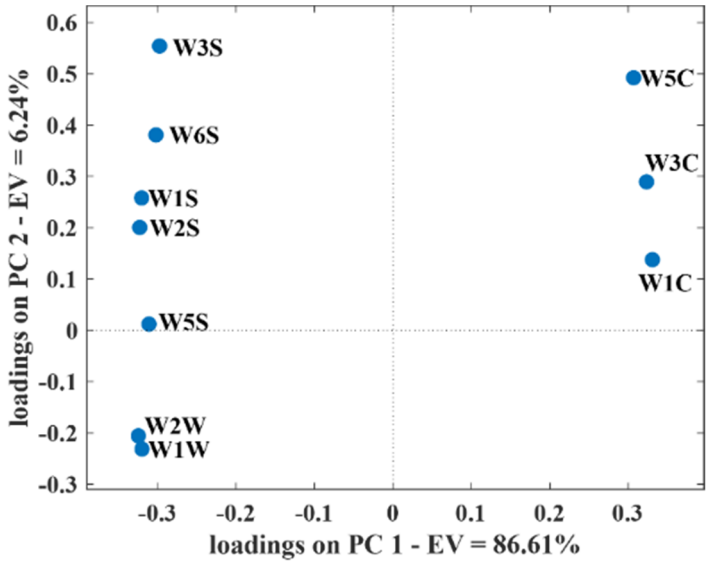

Figure 5. Load analysis diagram of electronic nose odor sensor.

Load analysis diagram of electronic nose odor sensor in RMC and MCC was listed in Figure 5. It could be seen that the variance contribution rates of the 10 sensors on the first principal component were basically the same, and the factor loads of sensors W1S and W2S, W1W and W2W were very close, which indicated that sensors W1S and W2S, W1W and W2W were similar to each other, and one of them can be eliminated to optimize the sensor array. In the second principal component, the contribution rate of W5S was the smallest, so the W5S sensor was removed. According to the results of correlation analysis and load analysis, sensors W3C, W6S, W1S, W2W and W3S were selected to

form a new sensor array for subsequent analysis.

3.3. Rapid discrimination of RMC and MCC

3.3.1. Principal component analysis

In this experiment, the sample odor response value was used as the input variable, and the unsupervised identification of RMC and MCC samples was carried out by principal component analysis (PCA) [26]. The effects of sensor array optimization of the models were compared (Figure 6). before the sensor array optimized, the cumulative interpretation of nine principal component reached 99.98% (Figure 6a), among of the first two principal components reached 92.85% (PC1 = 86.61%, PC2 = 6.24%). After optimized, the accumulation interpretation of six principal components reached 100% (Figure 6b), and the cumulative interpretation of the first 2 principal component score reaches 92.76% (PC1 = 87.17%, PC2 = 5.59%). The results indicated that the first two principal components can represent above 92% of all odor information characteristics of the sample, and the extracted information was well representative.

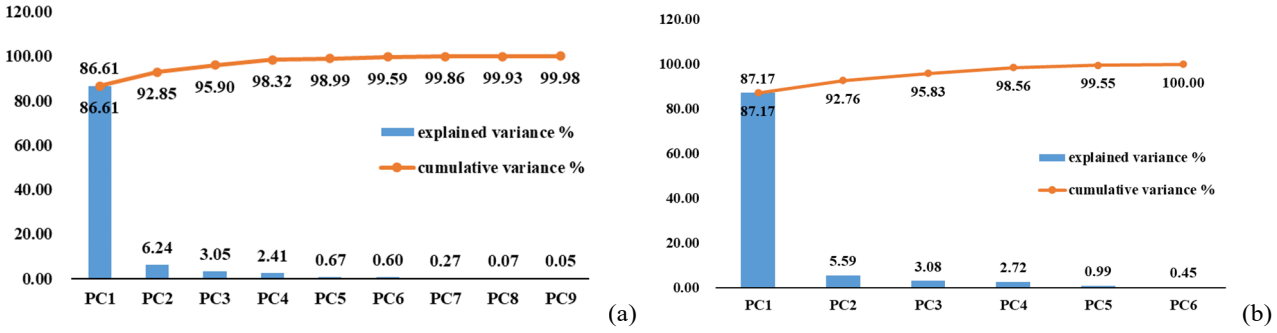

Figure 6. The Sensor array optimization results of PCA model explanatory variables and cumulative explanatory variable. (a) before optimization (b) after optimization

The scores of the first two principal components of the PCA model before and after the optimization of the sensor array were shown in Figure 7. It could be seen from the figure that the RMC had a large spatial distribution range, indicating that there were large differences in odor information among raw products, which may be related to the origin and production date of them, but there was no spatial overlap between raw products and carbon products, indicating that the odor of MCC samples has changed significantly after processed. The smell information of MCC with different processing degrees overlaps seriously in space. At the same time, after the optimization of sensor array, the spatial distribution range of MCC smell information was further reduced, and the clustering effect of the same category was better. So, the unsupervised recognition method of PCA could not effectively distinguish MCC with different processing degrees, the supervised pattern recognition method with better training effect needs to be adopted.

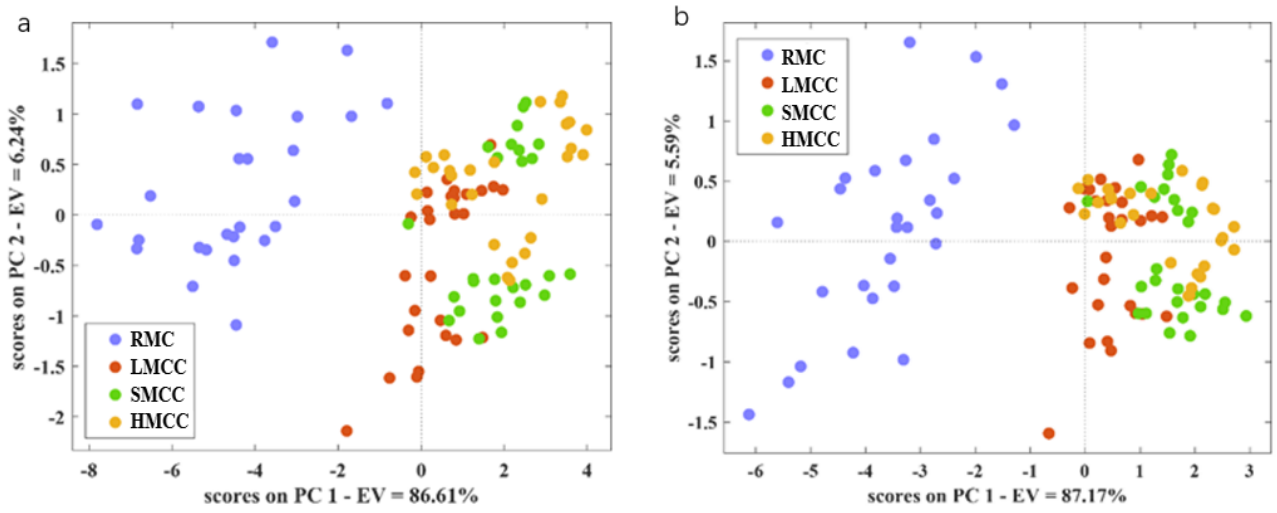

Figure 7. Scores of the PCA model before (a) and after the optimization of the sensor array model (b).

3.3.2. Supervised pattern recognition

In this experiment, taking the collected sample odor quantitative data as the independent variable and the sample category as the dependent variable, the discriminant models of RMC and MCC samples with different processing degrees were established by using linear discriminant (LDA) [27], partial least squares-discriminant analysis (PLS-DA) [28,29] and support vector machine (SVM) [30,31]. The performance of the model was evaluated by 10-fold cross validation and external validation. The samples were divided into training set and verification set according to the ratio of 2:1, which included 72 batches of training set and 36 batches of verification set. The training set was used to train the model and optimize the best parameters of the model; The validation set was used to test the application effect of the model. The identification results of LDA, PLS-DA and SVM discrimination models based on the response value of electronic nose odor sensor were shown in Table 7. Comparing the results of the three models, it could be seen that the positive judgment rates of cross validation and external validation of SVM models were higher than 90.00%, indicating that this method could accurately complete the rapid discrimination of RMC raw products and MCC with different processing degrees.

Table 7. Supervised pattern recognition results**.**

Model	Parameter	Training set					Validation set				
		RMC	LMC	SMC	HMC	Correct-judgment	RMC	LMC	SMC	HMC	Correct-judgment
LDA	-	18/18	16/18	13/18	12/18	81.94%	9/9	9/9	6/9	6/9	83.33%
PLS-DA	Components = 3	12/18	9/18	6/18	11/18	50.00%	6/9	4/9	2/9	1/9	36.11%
SVM	**RBF kernel function C = 2, g = 0.0052**	**18/18**	**17/18**	**17/18**	**14/18**	**91.67%**	**9/9**	**9/9**	**7/9**	**8/9**	**91.67%**

3.4. Quantitative analysis of internal components of RMC and MCC based on electronic nose

3.4.1. Correlation analysis

In this experiment, the odor quantitative data of RMC and MCC were correlated with the content of internal components, and the Pearson correlation analysis was carried out by SPSS 23.0 software. The Pearson correlation analysis results were shown in Table 8. From the correlation coefficient, it could be seen that the correlation between gallic acid content and each sensor was low, and the correlation coefficients are lower than 0.3; The contents of 5-HMF, paeoniflorin and paeonol were significantly correlated with the sensors W3C, W6S, W1S, W2W and W3S (significance less than 0.01). The contents of 5-HMF were positively correlated with the sensors W3C and negatively correlated with W6S, W1S, W2W and W3S, while the contents of paeoniflorin and paeonol were negatively correlated with the sensors W3C and positively correlated with W6S, W1S, W2W and W3S. To a certain extent, the higher the W3C response value of the sensor, the lower the W6S, W1S, W2W and W3S of the sensor, the higher the content of 5-HMF, and the opposite was true for paeoniflorin and paeonol.

Table 8. Correlation analysis results between odor characteristics and internal component content of RMC and MCC (number of cases = 108).

		W3C	W6S	W1S	W2W	W3S
gallic acid	Pearson correlation	-0.183	0.143	-0.045	0.238*	-0.061
	Significance (2-tailed)	0.057	0.141	0.644	0.013	0.530
5-HMF	Pearson correlation	0.265**	-0.362***	-0.488***	-0.268**	-0.527***
	Significance (2-tailed)	0.005	0.000	0.000	0.005	0.000
paeoniflorin	Pearson correlation	-0.761***	0.889***	0.777***	0.890***	0.831***
	Significance (2-tailed)	0.000	0.000	0.000	0.000	0.000
paeonol	Pearson correlation	-0.662***	0.682***	0.564***	0.784***	0.557***
	Significance (2-tailed)	0.000	0.000	0.000	0.000	0.000

Note: *: Significance (2-tailed) < 0.05; **: Significance (2-tailed) < 0.01; ***: Significance (2-tailed) < 0.001.

3.4.2. Regression model for compounds content of RMC and MCC based on odor quantification

According to the results of "3.4.1" correlation analysis, the content of gallic acid,5-HMF, paeoniflorin and paeonol had a certain correlation with their odor characteristics. In this experiment, the quantitative data of odor were taken as independent variables, gallic acid, 5-HMF, paeoniflorin and paeonol were taken as dependent variables, and partial least squares regression (PLSR) and support vector machine regression (SVR) were used to establish the component content regression model. The performance of the model was evaluated by 10-fold cross validation and external validation. The determination coefficient (R^2), root mean square error (RMSE) and relative analysis error (RPD) were used as the evaluation indexes of the regression model.

Table 9. Results of chemical composition quantitative model based on odor response value of electronic nose.

	Model	Parameter	R^2c	*RMSEc*	R^2p	*RMSEp*	*RPD*
gallic acid	PLSR	PCs = 5	0.3773	1.2839	0.2848	1.3492	1.1992
	SVR	**c = 8** **g = 8**	**0.8251**	**0.6600**	**0.8509**	**0.6161**	**2.6259**
5-HMF	PLSR	PCs = 4	0.5339	1.1621	0.4579	1.2702	1.3840
	SVR	**c = 724.0773** **g = 0.7071**	**0.7457**	**0.7364**	**0.6605**	**0.8705**	**2.0194**
paeoniflorin	PLSR	PCs = 3	0.8757	1.1215	0.8970	1.1792	3.1600
	SVR	**c = 32** **g = 1.4142**	**0.9585**	**0.6477**	**0.9229**	**1.0205**	**3.6510**
paeonol	PLSR	PCs = 5	0.6242	2.2585	0.7175	2.3485	1.9082
	SVR	**c = 2** **g = 9.7656e-04**	**0.8182**	**1.5706**	**0.8019**	**1.9669**	**2.2784**

The results of the quantitative model of compounds content of RMC and MCC based on odor quantitative data were shown in Table 9. According to the model evaluation indexes in the table, the fitting effect of the regression model established by SVM was better than that of PLSR model, the model correlation coefficients R^2c and R^2p were significantly improved, and RMSEc and RMSEp were smaller, such as gallic acid quantitative model, R^2c and R^2p increased from 0.3773 and 0.2848 to 0.8251 and 0.8509 respectively, RMSEc and RMSEp decreased from 1.2839 and 1.3492 to 0.6600 and 0.6161 respectively, and RPD increased from 1.1992 to 2.6259, indicating that SVM had high prediction effect in dealing with the quantitative problem of component content with low correlation with odor characteristics. This may be related to its working principle in dealing with nonlinear problems. The correlation results between the measured values and predicted values of each component were shown in Figure 8. It could be observed that the predicted results of 5-HMF were scattered on the regression line, and the correlation coefficient of the quantitative model was lower than 0.75, indicating that the performance of the model may be improved. The other compounds gallic acid, paeoniflorin, paeonol were concentrated on the regression line. From the above results revealed that SVM was able to identify the electronic nose relevant to the target compounds and accurately predict the contents of these compounds except for 5-HMF. At the same time with a good prediction, which enabled electronic nose to accurately analyze the influence of processed on MCC quality.

In fact, it is a very popular work to detect samples by electronic nose to obtain the effect of different other conditions on samples. For example, Sana Tatli et al. used electronic nose to detect the response difference of volatile organic compound emission in cucumber to track the effect of different urea fertilizers [32]； Robert Rusinek et al. analyzed the effects of fiber additives on vocs in bread using electronic nose technology [33]. Faraneh Khodamoradi et al. investigated the effects of different nitrogen fertilizer amounts on basil [34]. However, all of these studies were based on qualitative analysis, and this study combined qualitative and quantitative analysis to provide a more accurate monitoring of the degree of carbon frying in moutan cortex.

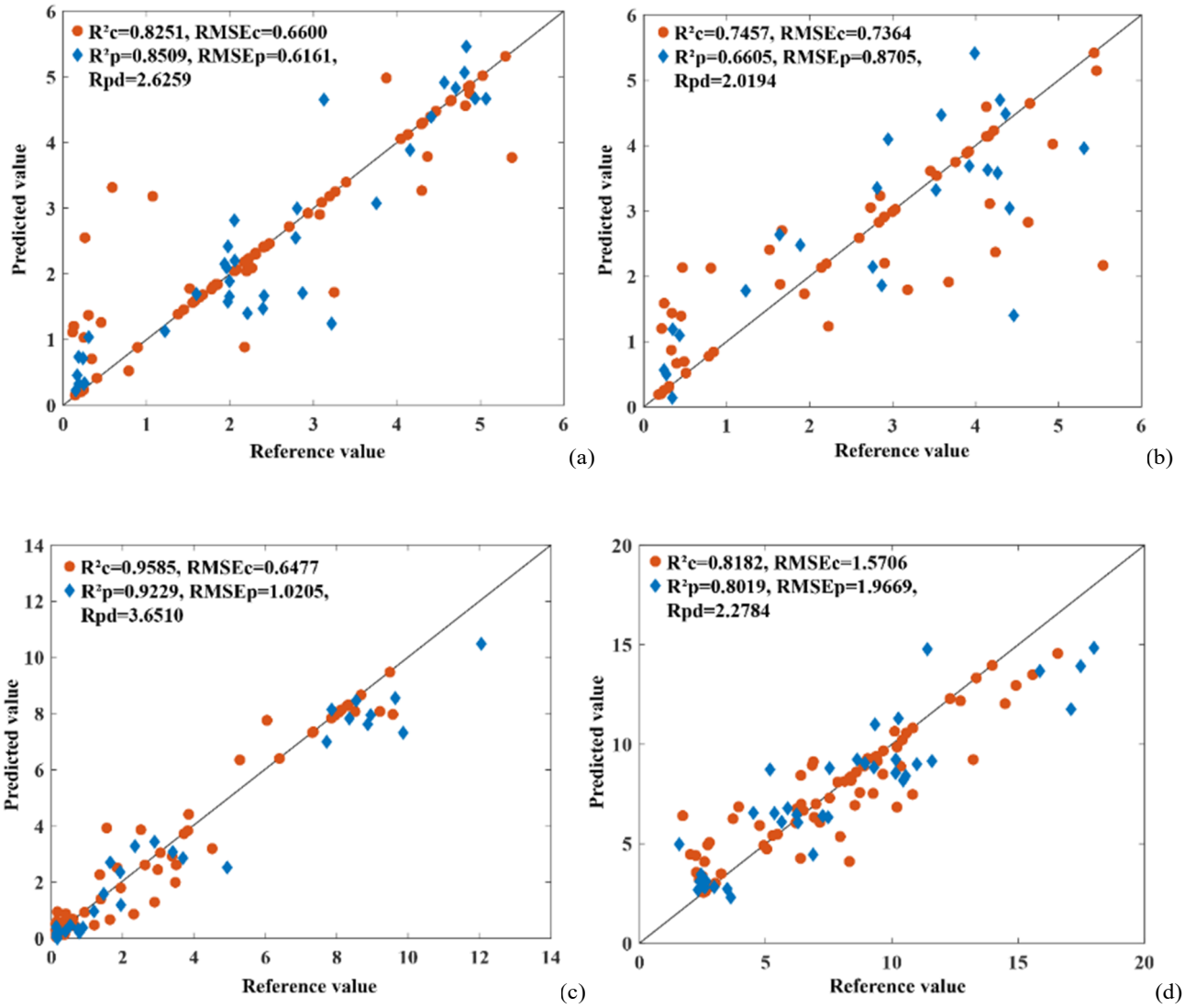

Figure 8. Predicted values and measured values of chemical composition. (a)—gallic acid; (b)—5-HMF; (c)—paeoniflorin; (d)—paeonol

4. Conclusions

The electronic nose combined with chemometrics was introduced to digitize the smell of RMC and MCC. The discrimination model and chemical composition quantitative model of RMC and MCC with different processing degrees were constructed. The experimental results showed that:

1) After the RMC was stir-fried, there was little difference in the odor response of MCC with different processing degrees, indicating that the volatile components did not change significantly with the deepening of processing degree; Combined with supervised SVM model, MCC with different processing degrees could be identified and predicted accurately, and the correct rate of sample discrimination was 91.67%.

2) Based on the odor digitization of RMC and MCC, combined with PLSR and SVM, the quantitative models of gallic acid, 5-HMF, paeoniflorin and paeonol in RMC and MCC were established. Except for 5-HMF, the determination coefficients (R^2) of the quantitative models of gallic

acid, paeoniflorin and paeonol were higher than 0.8. The results showed that the quantitative data of RMC and MCC odor could be used to predict the contents of three chemical components; The fitting effect of 5-HMF quantitative model based on odor response value was general, and the model could be optimized and improved by fusing the eigenvalues of other sensors.

In addition, this study established a reliable quantitative model, which is not available in most of the latest studies mentioned above. Quantitative research not only gives more accurate interpretation of samples, but also can be used to control the degree of processing more accurately, which is a more comprehensive perspective of analysis.

Acknowledgments

This work was financially supported by the Project of the National Natural Science Foundation of China (No. 82173973, 81473352) and Key projects of Guangdong Provincial Department of Education (No. 2018KZDXM040). We are also grateful to many of our colleagues for supporting this work.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. Q. L Ye, P. Cheng, D. H. Yan, Y. Sun, Y. Zhang, H. Cao, et al., Nine absorbed components pharmacokinetic of raw and processed Moutan Cortex in normal and blood-heat and hemorrhage syndrome model rats, *Biomed. Chromatogr.*, **34** (2020). https://doi.org/10.1002/bmc.4963
2. X. L. Zhao, A. W. Ding, L. Zhang, H. Zhang, A study on the historical changes of preparation of Cortex Moutan, *Chin. Arch. Tradit. Chin. Med.*, **26** (2008), 1907–1910. https://doi.org/10.13193/j.archtcm.2008.09.68.zhaoxl.061
3. G. G. Gu, *Shen Nong's Herbal Classic*, Lanzhou University Press, (2009), 113.
4. S. J. Zhou, J. Meng, Z. P. Huang, S. Z. Jiang, Y. Q. Tu, A method for discrimination of processed ginger based on image color feature and a support vector machine model, *Anal. Methods*, **8** (2016), 2201–2206. https://doi.org/10.1039/C5AY03171K
5. Administration GfaD, *Guangdong Standard for Processing Traditional Chinese Medicine*, Guangdong Science and Technology Press, 1984.
6. Q. Y. Huang, S. J. Zhou, Q. F. Chen, J. Meng, S. M. Wang, Study on the content changes of multiple components in different processing levels of Peduncularum mout, *Chin. Med. Mater.*, **39** (2016), 1024–1027. https://doi.org/10.13863/j.issn1001-4454.2016.05.017
7. X. Li, X. L. Wei, X. L. Zhao, A. W. Ding, Comparison of chemical composition changes of Mout bark before and after charcoal frying, *Chin. J. Exp. Formulae*, **17** (2011), 32–35. https://doi.org/10.3969/j.issn.1005-9903.2011.23.010
8. H. Zhang. A. W. Ding, L. Zhang, Optimization of processing technology of Danpi charcoal by orthogonal method, *China Pharm. Ind.*, **17** (2008), 25–26. https://doi.org/ 10.3969/j.issn.1006-4931.2008.03.019

9. Y. Wang, X. Yang, P. F. Xia, X. Ma, R. J. Yang, J. R. Hu, et al., Research progress on chemical constituents, pharmacological effects and predictive analysis of quality markers of rhubarb, *Chin. Herb. Med.*, **50** (2019), 4821–4837. https://doi.org/10.7501/j.issn.0253-2670.2019.19.033
10. Q. L. Ye, P. Cheng, D. H. Yan, Y. Sun, Y. Zhang, H. Cao, et al., Nine absorbed components pharmacokinetic of raw and processed Moutan Cortex in normal and blood-heat and hemorrhage syndrome model rats, *Biomed. Chromatogr.*, **34** (2020), 49–63. https://doi.org/10.1002/bmc.4963
11. X. Li, L. Zhang, A. W. Ding, Study on the effective site and mechanism of hemostatic action of Paederma charcoal, *Chin. Herb. Med.*, **40** (2009), 1278–1280. https://doi.org/10.3321/j.issn:0253-2670.2009.08.033
12. J. Y. Zhang, Y. K. Li, W. L. Wen, L. Zhao, Progress in Cardiovascular Activity of Paeonol, *New Chin. Med. Clin. Pharm.*, **27** (2016), 148–150. https://doi.org/10.3969/j.issn.1003-9783.2016.01.032
13. W. Li, Y. L. Wang, S. X. Cai, H. Y. Zhang, H. Y. Shi, F. L. Huang, et al., Comparison of the effects of paeonol and aspirin on hemorrheology in rats, *Chin. Herb. Med.*, (2000), 31–33. https://doi.org/10.3321/j.issn:0253-2670.2000.01.017
14. L. Zhang, L. L. Cao, X. L. Zhao, A. W. Ding, J. Chen, X. Li, Changes of 5-hydroxymethylfurfural content in Pedunculus mout charcoal with different processing levels, *Chin. Pharm.*, **20** (2009), 2624–2626.
15. D. H. Yan, *Study on Processing Mechanism of Moutan Bark Charcoal Based on Global Pharmacokinetics*, Guangdong Pharmaceutical University, 2017.
16. J. Z. Tan, W. L Kerr, Determining degree of roasting in cocoa beans by artificial neural network (ANN)-based electronic nose system and gas chromatography/mass spectrometry (GC/MS), *J. Sci. Food Agric.*, **98** (2018), 3851–3859. https://doi.org/10.1002/jsfa.8901
17. X. X. Chao, X. L. Ren, M. Wang, L. Z. Wang, Y. R. Deng, Research progress on specifications, grades and quality standards of traditional Chinese medicine and decoction pieces, *Tradit. Chin. Med.*, **44** (2021), 494–498. https://doi.org/10.13863/j.issn1001-4454.2021.02.044
18. X. Y. Xie, X. Y. Zhen, H. Yu, Z. W. Zhang, S. W. Si, Y. Wang, et al., Advances in technology of pharmaceutical analysis in 2018, *Chin. J. Pharm. Anal.*, **40** (2020), 18. https://doi.org/10.16155/j.0254-1793.2020.05.01
19. R. Q. Yang, *Rapid Quality Evaluation Of Moldy Traditional Chinese Medicine Based On Electronic Nose*, Beijing University of traditional Chinese Medicine, 2019.
20. J. Tan, J. Xu, Applications of electronic nose (e-nose) and electronic tongue (e-tongue) in food quality-related properties determination: A review, *Artif. Intell. Agric.*, **4** (2020), 104–115. https://doi.org/10.1016/j.aiia.2020.06.003
21. S. T. Kim, I. H. Choi, H. Li, Identification of multi-concentration aromatic fragrances with electronic nose technology using a support vector machine, *Anal. Methods*, **13** (2021), 4710–4717. https://doi.org/10.1039/d1ay00788b.
22. A. Khorramifar, M. Rasekh, H. Karami, U. Malaga-Toboła, M. Gancarz, A machine learning method for classification and identification of potato cultivars based on the reaction of MOS type sensor-array, *Sensors*, **21** (2021), 5836. https://doi.org/10.3390/s21175836
23. M. Rasekh, H. Karami, A. D. Wilson, M. Gancarz, Performance analysis of mau-9 electronic-nose mos sensor array components and an classification methods for discrimination of herb and fruit essential oils, *Chemosensors*, **9** (2021), 243. https://doi.org/10.3390/chemosensors9090243

24. S. Slimani, E. Bultel, T. Cubizolle, C. Herrier, T. Rousselle, T. Livache, Opto-electronic nose coupled to a silicon micro pre-concentrator device for selective sensing of flavored waters, *Chemosensors*, **8**(2020), 60. https://doi.org/10.3390/chemosensors8030060
25. Q. Tao, Study on odor detection of shellfish based on PEN3 type electronic nose sensor, *Adv. J. Food Sci. Technol.*, **10** (2016), 796–800. https://doi.org/10.19026/ajfst.10.2264
26. Q. Zhang, H. G. Li, M. Li, L. Ding, Feature extraction of face image based on LBP and 2-D Gabor wavelet transform, *Math. Biosci. Eng.*, **17** (2020), 1578–1592. https://doi.org/10.3934/mbe.2020082
27. P. Xanthopoulos, P. M. Pardalos, T. B. Trafalis, Linear Discriminant Analysis, in *Robust Data Mining*, Springer, New York, (2013), 27–33. https://doi.org/10.1007/978-1-4419-9878-1_4
28. M. Pérez-Enciso, M. Tenenhaus, Prediction of clinical outcome with microarray data: a partial least squares discriminant analysis (PLS-DA) approach, *Hum. Genet.*, **112** (2003), 581–592. https://doi.org/10.1007/s00439-003-0921-9.
29. D. Ballabio, V. Consonni, Classification tools in chemistry. Part 1: linear models. PLS-DA, *Anal. Methods*, **5** (2013), 3790–3798. https://doi.org/10.1039/c3ay40582f.
30. V. N. Vapnik, S. Mukherjee, Support vector method for multivariate density estimation, *Adv. Neural Inf. Process. Syst.*, **76** (2000), 659–665. https://doi.org/10.1.1.899.1861
31. S. Mohapatra, K. Ganesh, M. Punniyamoorthy, R. Susmitha, Developing a classification model using SVM, in *Service Quality in Indian Hospitals*, (2018), 63–66. https://doi.org/10.1007/978-3-319-67888-7_8
32. A. D. Wilson, Rapid detection of urea fertilizer effects on voc emissions from cucumber fruits using a mos e-nose sensor array, *Agronomy*, **12** (2021). https://doi.org/10.3390/agronomy12010035
33. R. Rusinek, M. Gawrysiak-Witulska, A. Siger, A. Oniszczuk, A. A. Ptaszyńska, J. Knaga, et al., Effect of supplementation of flour with fruit fiber on the volatile compound profile in bread, *Sensors*, **21** (2021), 2812. https://doi.org/10.3390/s21082812
34. F. Khodamoradi, E. Mirzaee-Ghaleh, M. J. Dalvand, R. Sharifi, Classification of basil plant based on the level of consumed nitrogen fertilizer using an olfactory machine, *Food Anal. Methods*, **14** (2021), 2617–2629. https://doi.org/10.1007/s12161-021-02089-y

AIMS Press