Tea quality evaluation by applying E-nose combined with chemometrics methods

Min Xu[1] · Jun Wang[1] · Luyi Zhu[1]

Abstract Tea is one of the most popular beverage with distinct flavor consumed worldwide. It is of significance to establish evaluation method for tea quality controlling. In this work, electronic nose (E-nose) was applied to assess tea quality grades by detecting the volatile components of tea leaves and tea infusion samples. The "35th s value", "70th s value" and "average differential value" were extracted as features from E-nose responding signals. Three data reduction methods including principle component analysis (PCA), multi-dimensional scaling (MDS) and linear discriminant analysis (LDA) were introduced to improve the efficiency of E-nose analysis. Logistic regression (LR) and support vector machine (SVM) were applied to set up qualitative classification models. The results indicated that LDA outperformed original data, PCA and MDS in both LR and SVM models. SVM had an advantage over LR in developing classification models. The classification accuracy of SVM based on the data processed by LDA for tea infusion samples was 100%. Quantitative analysis was conducted to predict the contents of volatile compounds in tea samples based on E-nose signals. The prediction results of SVM based on the data processed by LDA for linalool (training set: $R^2 = 0.9523$; testing set: $R^2 = 0.9343$), nonanal (training set: $R^2 = 0.9617$; testing set: $R^2 = 0.8980$) and geraniol (training set: $R^2 = 0.9576$; testing set: $R^2 = 0.9315$) were satisfactory. The research manifested the feasibility of E-nose for qualitatively and quantitatively analyzing tea quality grades.

Introduction

Tea is one of the most popular beverages consumed across the world with attractive flavor and abundant health-benefit compounds such as theanine and polyphenols (Krstic et al. 2015). The differences in tea quality grades possesses different external flavor quality and internal health-benefit components. Tea quality is affected by various factors, such as plucking season, soil, climate and processing techniques (Unachukwu et al. 2010). Usually, teas with superior quality were planted in good environment condition and plucked at certain time, as well as processed with specific technique. Therefore, the supply of superior tea was limited by these factors above. Meanwhile, higher grades of tea quality usually means higher prices in market. Hence, some merchants intentionally sell inferior tea products as superior ones for making benefits, which disrupts the stability of tea markets and impairs the interest of consumers. Moreover, it is difficult for ordinary consumer to discriminate the tea products with different quality grades. Hence, it is necessary to establish efficient method for precisely estimating tea quality.

Tea aroma is an important sensory attribute in evaluating tea quality. Human olfactory system consists of a large number of receptors cells that could respond to different types of volatile components, therefore tea aroma is usually assessed by human sensory evaluation (Chen et al. 2010). Professional evaluators are required to describe properties

✉ Jun Wang
  jwang@zju.edu.cn

[1] Department of Biosystems Engineering, Zhejiang University, 866 Yuhangtang Road, Hangzhou 310058, People's Republic of China

of tea aroma with terminology that hard for consumers to understand and then grade tea quality according to the characteristics of tea aroma. The professional evaluators take an important role in the whole evaluation process. Hence, the result is subjective and difficult to reproduce since it is easily influenced by physiological and psychological condition of evaluators or other personal issues. Physico-chemical analysis and modern instruments like gas chromatography-mass spectrometer (GC–MS) and high performance liquid chromatography (HPLC) are also applied in evaluating tea quality (Wang et al. 2014; Lin et al. 2012; Rawat et al. 2007; Qin et al. 2013). Although the results provided by these methods are objective, they are time-consuming and need complicated pretreatments for tea samples.

Electronic nose (E-nose) is developed to mimic human olfactory system for the detection of complex odors in the headspace of samples. The E-nose mainly consists of a sensor array and a pattern recognition system. Sensor array acts as receptors cells of human to receive and react with volatile components and then a characteristic fingerprint is generated from the reaction of volatile components and sensors. The pattern recognition system works as human brain to distinguish samples with different fingerprints. A lot of applications have been found in existing researches of using E-nose to assess the volatile profile of meat (Tian et al. 2013; Kodogiannis 2017), fruits (Guohua et al. 2013; Wu et al. 2017), juices (Huang et al. 2015; Qiu et al. 2017; Qiu and Wang 2017), vegetables (Yao et al. 2015), nuts (Xu et al. 2017; Jiang et al. 2017), dairy products (Yang et al. 2015; Majcher et al. 2015) and honey (Dymerski et al. 2014).

In recent years, E-nose has also been applied in tea related researches, including the origin identification of green tea, the detection of optimum fermentation time for black tea and the evaluation of different green tea quality (Pasquini et al. 2016; Bhattacharyya et al. 2007). To obtain a satisfactory result of evaluating tea quality grades, some researches compared the performance of different experimental subjects, different data reductions and different classification algorithms, indicating the importance of those elements in applying E-nose to identify tea quality (Yu et al. 2009; Dai et al. 2015; Chen et al. 2011; Pławiak and Maziarz 2014). However, these researches only partially contributed to the selection of appropriate data processing and modelling methods. Hardly any research has been reported on systematically investigating the performance improving methods to detect tea quality by E-nose based on different data extractions and reductions and classification algorithms. Furthermore, the existed researches scarcely explored the efficiency of data reduction methods in building regression model and probed the potential correlation between E-nose signals and the contents of volatile components in tea samples.

The current research mainly apply E-nose to identify tea quality of different grades and mine the correlation relationship between E-nose signals and the content of volatile components. Data reductions were introduced to improve the efficacy of classification and correlation models. The main objectives are as follows: (1) by using PCA and MDS as unsupervised data reduction methods and LDA as a supervised data reduction method to process the original data and comparing the performance of their efficiency; (2) to set up classification models severally for identifying tea leaf and tea infusion samples based on different data extraction and data reduction methods and different classifiers, and to compare the classification results of these models; (3) to analyze the volatile components in tea leaf and tea infusion by using GC–MS, and to investigate the potential correlation between E-nose signals and the contents of volatile components in different tea grades by applying E-nose signals to build regression models for predicting the content main volatile components in tea samples.

Material and methods

Preparation of tea samples

Longjing tea were used as experimental samples in the research. The tea samples were picked from Longjing village (30°13′, 120°06′, Hangzhou city, Zhejiang province, China), which is an appropriate place for tea plant with good growing condition. All tea leaves were plucked during March to May in 2017 and processed by skilled workers with same method. The tea grade of Longjing tea is related to plucking time. Usually, tea leaves picked earlier means a higher price and higher tea grade. Six tea groups with different quality grades were applied in this study. There were 32 samples in each group and 5 g tea leaf for each sample. These samples were packed in aluminum foil and stored in 4 °C before testing. The six tea grades including their prices and plucking time were as follows: Grade 1 (1000 ¥/500 g, March 28), Grade 2 (800 ¥/500 g, April 2), Grade 3 (600 ¥/500 g, April 10), Grade 4 (400 ¥/500 g, April 18), Grade 5 (200 ¥/500 g, April 28), Grade 6 (100 ¥/500 g, May 3).

Electronic nose setup

The E-nose (PEN3, Airsense Corporation, Germany) used in this research are equipped with a metal oxide semiconductor (MOS) sensor array which composed of 10 different MOS sensors. The details of the sensor array were shown in Table 1.

Table 1 The name and main performance of each sensor in Electronic nose (PEN3)

Number	Name	Main performance	Reference
S1	W1C	Aromatic compounds	Toluene, 10 ppm
S2	W5S	Very sensitive, broad range sensitivity, react on nitrogen oxides, sensitive with negative signal	NO_2, 1 ppm
S3	W3C	Ammonia, used as sensor for aromatic compounds	Benzene, 10 ppm
S4	W6S	Mainly hydrogen, selectively, (breath gases)	H_2, 100 ppm
S5	W5C	Alkenes, aromatic compounds, less polar compounds	Propane, 1 ppm
S6	W1S	Sensitive to methane (environment) ca. 10 ppm. Broad range, similar to S8	CH_3, 100 ppm
S7	W1W	Reacts on sulfur compounds, H_2S 0.1 ppm. Otherwise sensitive to many Terrenes and sulfur organic compounds, which are important for smell, limonene, praline	H_2S, 1 ppm
S8	W2S	Detects alcohol's, partially aromatic compounds, broad range	CO, 100 ppm
S9	W2W	Aromatic compounds, sulfur organic compounds	H_2S, 1 ppm
S10	W3S	Reacts on high concentrations > 100 ppm, sometimes very selective (methane)	CH_3, 10 CH_3, 100 mg/kg

A tea leaf sample was prepared by placing 5 g tea leaves in a 500 mL beaker and was sealed with plastic wrap. In order to generate headspace gas, the sample was kept still for 60 min. Afterwards, the headspace gas was pumped into the gas sensor chamber at a constant rate of 200 mL min^{-1} and then reacted with the MOS sensors leading to the change of electric conductivity in sensors. In measuring stage, the ratio of conductance was measured as response signal of each sensor to tea sample. The measurement time was 70 s for each sample and a signal was recorded per second. After the measurement stage, the gas sensors chamber was cleaned for 70 s by cleaning air to ensure the response signal back to baseline. All the measuring signals were properly stored for later analyzing. 192 tea leaf samples (6 groups and 32 samples for each group) were detected by E-nose. The temperature during the detection was kept at 25 ± 1 °C.

A tea infusion sample was prepared by pouring 250 mL freshly boiled water into a 500 mL beaker with 5 g tea leaves and brewing for 5 min. After that, the tea leaves were filtered with a sieve and the tea infusion sample was cooled down to room temperature. All the tea leaf samples were brewed into tea infusion according to the brewing method above. Hence, there were 192 tea infusion samples for E-nose detection. The detection procedure and parameters for tea infusion samples were totally the same with those for tea leaf samples.

Gas chromatography-mass spectrometer

A HS-SPME (Head space-solid phase microextraction) method with a CVR/ DVB/PDMS fiber (65 μm film thickness, Supelco, Bellafonte, PA, USA) was applied to extract the volatile components of tea leaf and tea infusion samples. For the preparation of tea leaf samples, 2 g tea leaf were placed in a 50 mL glass vial sealed with a PTFE-coated septum (Beijing Bomex Co., China). For tea infusion samples, 15 mL tea infusion was placed in a 50 mL glass vial and ethyl caprate was added in tea infusion as internal standard. Before extraction, both of tea leaf and tea infusion samples need equilibration for 5 min in a bath at 60 °C. For each sample extraction, the fiber was inserted in the sample for 60 min. Then, the SPME fiber was desorbed in the injection port (240 °C) of the GC–MS for 5 min. After the desorption, the volatile components were entered into Agilent 7890A gas chromatograph and then analyzed with a mass selective detector (Agilent 5975C, Agilent, USA). Temperature programming was as follows: the temperature in the GC oven was initially kept at 40 °C for 2 min, then heated to 70 °C at 3 °C /min, heated to 110 °C at 2 °C /min, then increased to 200 °C at 5 °C /min, and retained for 5 min at 200 °C. Helium was applied as carrier gas (purity > 99.99%) and the flow velocity was constant at 1 mL min^{-1}. The mass spectrometer conditions were: ionization mode, EI; electron energy, 70 eV; interface temperature, 280 °C; ion source temperature, 230 °C; mass scan range, 47–401 u. The retention times for n-alkanes (C8-C20; Sigma. Aldrich [Shanghai] Trading Co., Ltd., Shanghai, China) were determined to calculate the retention indices (RI) of volatile components in tea samples. MS fragmented patterns with the National Institute of Standards and Technology (NIST) 11.0 database (NIST 11.0, USA) and published data were applied to identify volatile components. When the compounds were identified by NIST database, peaks with a similarity index more than 80% were the assigned name of corresponding compounds. The relative amounts of volatile components in tea leaf were presented as percentages of the total area of volatile component peaks. The concentration of volatile components in tea infusion were calculated in ng/g based on

internal standard. There were three replicates for each tea grade detection by GC–MS.

Data analysis

Principal component analysis

Principal component analysis (PCA) is a widely used unsupervised method for linear dimensionality reduction analysis, which aims at mapping high dimensional variables in a low dimensional space via a linear projection and maximization of variance in the low dimensions (Lai and Teoh 2016). The main procedure for linear dimensionality reduction analysis using PCA is the computation of covariance matrix of original variables. Afterwards, eigenvalues and eigenvectors are obtained by decomposing the covariance matrix and the eigenvectors are sorted according to descending order of the eigenvalues. The first eigenvector contains the highest variance of the original variables as well as the variances of succeeding eigenvectors are correspondingly decreasing in turn (Qiu et al. 2017).

Multi-dimensional scaling

Multi-dimensional scaling (MDS) mainly aims to place samples in N-dimensional space and ensure that the sample-between distance in K-dimensional space $(K > N)$ could be preserved as well as possible. $D = [d_{ij}]$ presents the distance matrix in K-dimensional space, where d_{ij} is the distance between the coordinates of ith sample and jth sample. In order to find the corresponding embedding matrix in N-dimensional space, another matrix $B = [b_{ij}]$ is computed with D. where

$$b_{ij} = -\frac{1}{2}\left[d_{ij}^2 - \frac{1}{N}\sum_{k=1}^{N}d_{kj}^2 - \frac{1}{N}\sum_{k=1}^{N}d_{ik}^2 + \frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}d_{kl}^2\right].$$

Afterwards, eigenvalues λ_i and corresponding eigenvectors $\gamma_i = [\gamma_{ij}]$ are computed by the decomposition of B matrix. the eigenvectors of the q-largest eigenvalues form the embedding matrix in N-dimensional space (Torgerson 1952; Kramer 2016).

Linear discriminant analysis

Linear discriminant analysis (LDA) is a supervised method for linear dimensionality reduction analysis and has been successfully applied in gas identification and classification (Akbar et al. 2016; Choi et al. 2017). For the purpose of extracting classified information and reducing the dimensions of variables, high dimensional variables are projected into a low dimensional vector space based on Fisher discriminant criterion by minimizing within-class variance and maximizing between-class variance in the low dimensional vector space, which ensures the variables in different classes are separated as far as possible.

Logistic regression

Logistic regression (LR) with multinomial is a supervised classification model with a discrete random variable set of {1, 2, 3…K}, in which K is the number of categories. Multinomial logistic regression is an Extension of binomial logistic regression in which the independent variable is either 0 or 1 (Prabhakar et al. 2015). Maximum likelihood method is applied to estimate parameters in LR model. The problem of model building turn to find optimal solution of likelihood function by usually using quasi-newton method or gradient descent method.

Support vector machine

Support vector machine (SVM) is a supervised learning method for both classification and regression. It was proposed by Cortesand Vapnik based on statistical learning theory (SLT) and worked as a linear machine in the high dimensional feature space formed by the non-linear mapping of the n-dimensional input vector into a K-dimensional feature space $(K > n)$ with a function (Cortes and Vapnik 1995). The construction of a hyper plane and the maximization of interval between adjacent categories in the K-dimensional feature space are significant targets for SVM model. The specialized learning procedure make SVM working as a trained network with good generalization (Brudzewski et al. 2004).

The python sklearn library (python 3.6.3) was introduced to complete PCA, MDS, LDA, LR and SVM process.

Results and discussion

E-nose sensor responses and data extractions

The conductance ratio (G/G0) was chosen as responding signal, where G was the sensor conductivity when sensor exposed to sample gas and G0 was the conductivity when sensor exposed to the zero gas. Figure 1 showed the typical E-nose responses for a typical tea leaf and a tea infusion sample. The x-axis represented detection time and the y-axis represented the changes of G/G0 values of E-nose sensors. Changes of G/G0 values for tea leaf sample with the increasing of detection time was shown in Fig. 1a, where only S2 had a significant change while the other
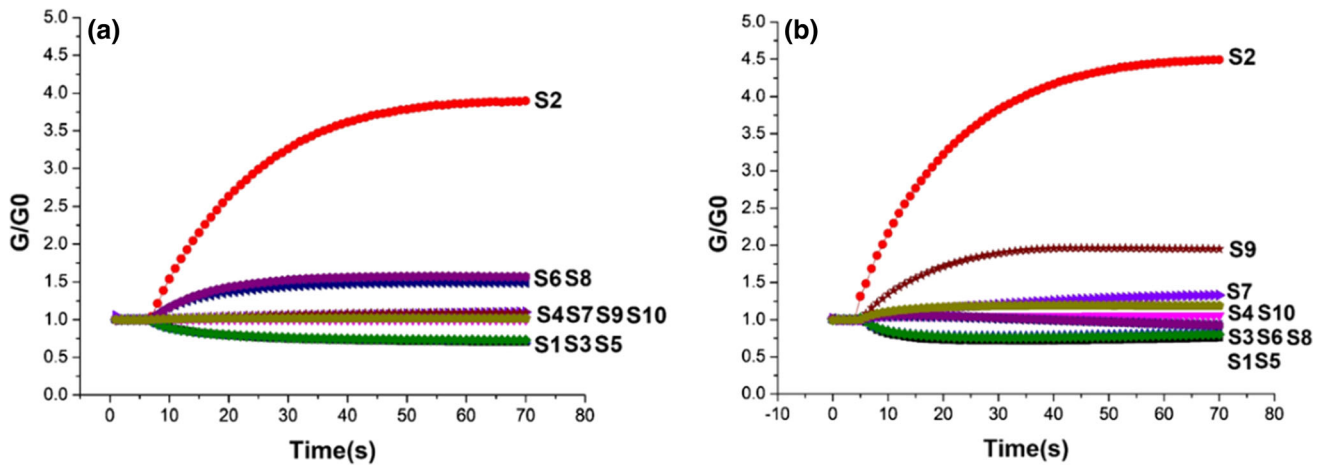
Fig. 1 Responding curves of ten E-nose sensors to **a** tea leaf and **b** tea infusion

sensors exhibited few variations. In Fig. 1b, the responding signals of S2 and S9 changed notably during the whole detection process for tea infusion sample.

Different data extractions including 70th s value, 35th s value and average differential value (ADV) were applied to extract characterizing information from original signals and to help in obtaining a model with good generalization and learning capacity. 35th s value represented the signal of each sensor at the half of the detection. 70th s value represented the signal of each sensor at the end of the detection. The formula for average differential value was

$$S_{ADV} = \frac{\int_1^{70} \frac{\Delta x}{\Delta t} dt}{70}.$$

The results of data reduction

PCA is a commonly used unsupervised method to project data into a low dimensional space by orthogonal transformation. Nine principle components (PCs) were selected for PCA model in the following research to retain the most of original information. As shown in Fig. 2, the first two PCs were chosen with PC1 as x-axis and PC2 as y-axis to visualize the distribution of samples. Figure 2a1–a3 presented the distribution of tea leaf samples based on the three data extractions, in which groups of Grade 1, Grade 2, Grade 3 and Grade 4 were overlapped, and the groups of Grade 4, Grade 5 and Grade 6 were overlapped. Figure 2a4–a6 presented the distribution of tea infusion samples based on the three data extractions. Figure 2a4, a6 exhibited that the tea infusion groups of Grade 1, Grade 2, Grade 3 and Grade 6 were overlapped, and the groups of Grade 3, Grade 4 and Grade 5 could not be classified clearly. By comparing the distribution of tea leaf samples with that of tea infusion samples, it could be found that the distances between groups for tea infusion samples were slightly greater than those for tea leaf samples, manifesting

that the PCA distribution for tea infusion samples were slightly better than those for tea leaf samples, although both of the results were not satisfactory.

MDS is also an unsupervised method to project data into a low dimension by keeping the distances between samples in new data space the same as they are in original space. In order to retain most of the data information, nine principle components were extracted for classification analysis. Figure 2b1–b6 presented the 2-dimensional distribution of all the samples by using the first two principle components of MDS projection. The MDS distribution for tea leaf samples were shown in Fig. 2b1–b3, in which all the samples overlapped with each other. As shown in Fig. 2b4–b6, for tea infusion samples, the similar results were obtained as for tea leaf samples that all of the tea infusion samples overlapped with each other and distributed disorderly, indicating the inefficiency of using the first two principle components of MDS projection in separating samples of different tea grades.

As unsupervised methods like PCA and MDS could not achieve satisfactory results in distributing tea samples, LDA was applied as a frequently-used supervised method for data reduction. LDA mainly concentrates on finding a proper project axes that samples from the same group can be distributed closely while samples from different groups can be distributed separately. The data dimension size after LDA processing should be less than the number of groups. In order to retain most of the original information, original data with ten-dimension were reduced to five-dimension. Two-dimensional plots (shown in Fig. 2c1–c6) were presented to visualize the distribution of samples in different groups. The distribution plots for tea leaf samples on the basis of 70th s value and 35th s value were shown in Fig. 2c1, c2. In the two plots, Grade 3 could be clearly discriminated from the other groups, and Grade 1 and Grade 2 were partially overlapped and were far away from
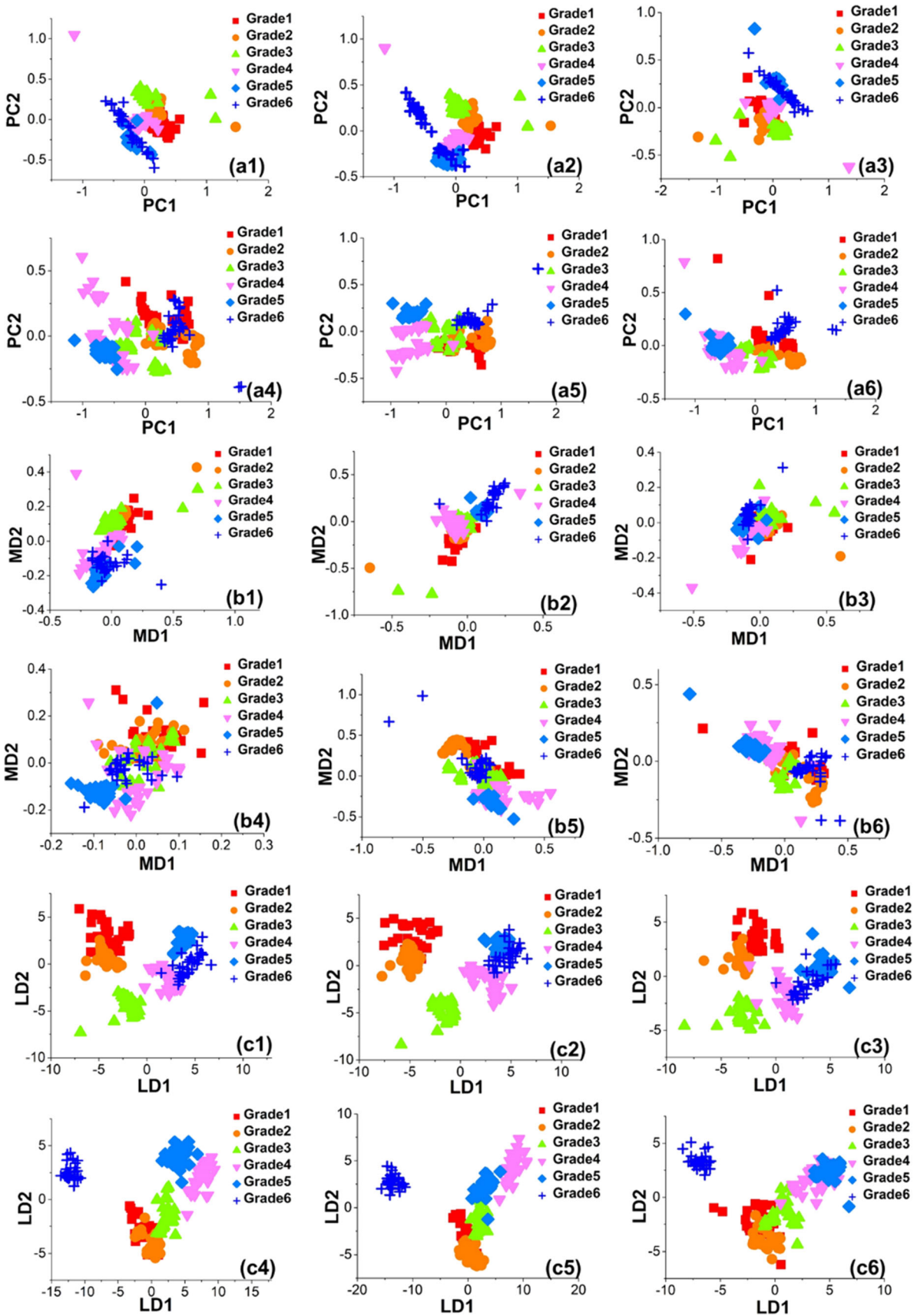
◄**Fig. 2** PCA results for tea leaf based on **a1** 70th s value, **a2** 35th s value, **a3** ADV, and for tea infusion based on **a4** 70th s value, **b5** 35th s value, **a6** ADV; MDS results for tea leaf based on **b1** 70th s value, **b2** 35th s value, **b3** ADV, and for tea infusion based on **b4** 70th s value, **b5** 35th s value, **b6** ADV; LDA results for tea leaf based on **c1** 70th s value, **c2** 35th s value, **c3** ADV, and for tea infusion based on **c4** 70th s value, **c5** 35th s value, **c6** ADV

Grade 4, Grade 5 and Grade 6 in which Grade 6 respectively overlapped with Grade 4 and Grade 5. When ADV used as data extraction, the distribution of samples in different tea leaf groups were similar to that based on the other two data extractions except that two samples in Grade 4 were mistakenly classified into Grade 2 and Grade 3. For tea infusion samples, when based on 70th s value as data extraction (Fig. 2c4), samples in Grade 4, Grade 5 and Grade 6 could be severally separated, while Grade 2 overlapped with Grade 1 as well as partially overlapped with Grade 3. The distribution of samples based on 35th s value (Fig. 2c5) was similar to that based on 70th s value (Fig. 2c4) except that Grade 5 was closer to Grade 3 in Fig. 2c5. In Fig. 2c6, only Grade 6 was clearly classified and almost all the other groups overlapped with one another. By comparing the LDA distribution results severally from tea leaf with those from tea infusion samples, tea infusion slightly outperformed tea leaves in distributing tea samples of different quality grades. It was also found that data extraction of 70th s value for tea infusion exhibited the best result in expressing differences of tea grades compared with the other data extractions. In comparison the two-dimension distribution plots based on PCA and MDS with those based on LDA, LDA provided superior performance in separating samples than PCA and MDS did. The reason might ascribe to the labels information only consisted in the LDA models.

The classification results based on Logistic Regression

Linear model of Logistic Regression (LR) was applied in this research to set up classification model. Quasi-newton method was decided to find optimal solution of likelihood function. The reciprocal value of regular parameter (C) in LR model was optimized in the range of 10^{-2}–10^{4}. 10-fold cross validation was applied for assessing the stability of the models. Dataset was randomly divided into two parts for training and testing model: 156 samples were used as training set and the rest of 36 samples were used as testing set. The classification results for tea samples based on different data extractions and data reductions were summarized in Table 2. For tea leaf samples, a consistent conclusion could be obviously achieved that PCA and

MDS reduction process contributed no effort to LR models and LDA improved the classification accuracy rate for most of the LR models. The LR model based on the 70th s value and LDA process achieved the best classification accuracy rate of 0.9444 when compared with the other LR models for identifying tea leaf samples. For tea infusion samples, among the three data extractions, the classification results based on LDA presented higher accuracy rate than those based on original data, PCA and MDS process. The accuracy rates for training set and testing set were 0.9808 and 0.9722 in LR model when based on LDA and the data extraction of 70th s value.

The classification results based on support vector machine

The SVM with radial basis function (RBF) was applied to identify tea leaf and tea infusion samples as a nonlinear supervised method. Penalty parameter C and kernel parameter gamma (γ) was optimized in SVM models by using a grid search method to search the best combination of C and γ. The grid search method was applied to search the best combination of C and γ in a range of 10^{-2}–10^{2}. 10-fold cross-validation was applied to evaluate the performance of each combination of parameters (C, γ). Table 3 summarized the best combinations of C and γ and classification results of SVM models for tea leaf and tea infusion samples. As for tea leaf samples, when using original data as input matrix, the accuracy rates of SVM models based on data extractions of 70th s value and ADV were higher than that based on data extraction of 35th s. The MDS process led to the decrease of accuracy rate for all the SVM models. The PCA and LDA process increased the classification accuracy rates of the SVM models by applying 70th s value and 35th s value as data extraction methods. Generally, the performance of SVM based on LDA were better than both PCA, MDS and original data in this case. For tea infusion samples, when using original data as input matrix, the accuracy rates of the SVM models based on data extraction of 70th s value and 35th s value were higher than those based on data extraction of ADV. Meanwhile, the LDA process provided higher accuracy rate in SVM models than PCA, MDS and original data did. Moreover, the LDA process coupled with SVM model could achieve the best classification accuracy rates of 100% when based on the data extractions of 70th s value. Through the comparison of results for tea leaf and tea infusion samples, the accuracy rates of SVM models in both training set and testing set for classifying tea infusion samples were generally higher than those for tea leaf samples, which may ascribe to the volatile components in tea infusion better characterize tea samples of different grades.

Table 2 Comparison of LR results based on different data extractions and data reductions for tea samples

Tea samples	Data extractions	Data reductions	Tenfold CV	Training set	Testing set
Tea leaf	70th s value	Original	0.9222	0.9872	0.9167
		PCA	0.8944	0.9936	0.8889
		MDS	0.8139	0.8718	0.8611
		LDA	0.9611	1	0.9444
	35th s value	Original	0.8389	0.8782	0.8333
		PCA	0.9000	0.9615	0.8611
		MDS	0.8249	0.9679	0.8889
		LDA	0.9361	0.9615	0.8889
	ADV	Original	0.8806	0.9808	0.8333
		PCA	0.8944	0.9744	0.8611
		MDS	0.8194	0.9231	0.6667
		LDA	0.7750	0.9551	0.8661
Tea infusion	70th s value	Original	0.9250	0.9744	0.75
		PCA	0.9333	0.9744	0.7778
		MDS	0.8972	0.9936	0.75
		LDA	0.9444	0.9808	0.9722
	35th s value	Original	0.9278	0.9615	0.7222
		PCA	0.9417	0.9808	0.7778
		MDS	0.9083	0.9744	0.8056
		LDA	0.9361	0.9551	0.9444
	ADV	Original	0.9250	0.9744	0.7222
		PCA	0.9250	0.9808	0.6944
		MDS	0.8722	0.9615	0.6667
		LDA	0.9417	0.9744	0.8889

When comparing the classification results of the SVM models with LR models, although SVM models generally outperformed LR models, some accordant conclusions could be obtained. Firstly, the LDA process provided better classification results than PCA, MDS and original data did. Secondly, after LDA process, the accuracy rates based on the models for tea infusion samples were generally superior than those for tea leaf samples. Thirdly, by applying LDA as data reduction, both of LR models and SVM models for tea infusion based on data extraction of 70th s value provided satisfactory results with accuracy rates of 97.22% and 100%.

The analysis of volatile components in tea leaf and tea infusion by GC–MS

The volatile components of both tea leaf and tea infusion were measured by GC–MS in this research. Table 4 presented the relative amounts of the identified volatile components in tea leaf as percentages of the total area of volatile components peaks and the contents of volatile components contained in tea infusion identified with reference to internal standard. For tea leaf samples, the content of nonanal was dominate among all the identified

volatile components. When comparing the quantities of volatile components in tea leaf samples of different grades, the contents of benzaldehyde, nonanal, α-Terpineol, decanal and β-cyclociral were decreased with the decreasing of tea grades, which indicated the positive contribution of these volatile components to tea grades. Meanwhile, it is notably found in this research that most of the volatile components in tea leaf including benzaldehyde, benzeneacetaldehyde, 1-octanol, nonanal, phenylethyl alcohol, naphthalene, α-terpineol, methyl salicylate, decanal, β-cyclociral, (z)-3-hexenyl hexanoate, tetradecane, α-cedrene, α-ionone, geranyl acetone, β-ionone, cadinene, pentadecane and heptadecane could also be found in tea infusion. Moreover, in this research, 1-octen-3-ol, β-myrcene, linalool, safranal, geraniol, citral, indole, cubebene, jasmone, nerolidol and methyl jasmonate were only identified in tea infusion might indicate that new volatile components were generated with the help of boiled water. In tea infusion, the main constituents of volatile components were linalool, geraniol and nonanal, which possessed rose-like or flora impression. Along with the tea grades decreasing, the contents of linalool and nonanal presented a descending tendency. Jasmone was only found in the tea infusion of Grade 1 and Grade 2. Meanwhile,

Table 3 Comparison of SVM results based on different data extractions and data reductions for tea samples

Tea samples	Data extractions	Data reductions	Parameters		Tenfold CV	Training set	Testing set
			C	γ			
Tea leaf	70th s value	Original	18.4207	22.2299	0.8917	1	0.8333
		PCA	39.0694	3.3932	0.8972	0.9872	0.9167
		MDS	2.3299	39.0694	0.8611	0.9872	0.75
		LDA	4.9417	0.0309	0.9389	0.9744	0.9444
	35th s value	Original	2.3299	39.0694	0.9361	0.9936	0.6944
		PCA	39.0694	15.2642	0.9250	1	0.8333
		MDS	1.9307	56.8987	0.9278	0.9936	0.7222
		LDA	3.3932	0.0176	0.9499	0.9679	0.9167
	ADV	Original	56.8987	1.5999	0.8778	0.9744	0.8611
		PCA	32.3746	5.9636	0.8806	0.9936	0.8611
		MDS	68.6649	4.0949	0.8806	0.9744	0.7778
		LDA	1.9307	0.0449	0.9167	0.9487	0.8611
Tea infusion	70th s value	Original	7.1968	5.9636	0.9528	1	0.9167
		PCA	8.6851	2.8118	0.9583	0.9936	0.9167
		MDS	7.1968	12.6486	0.9250	0.993	0.9444
		LDA	2.3299	0.0449	0.9667	1	1
	35th s value	Original	10.4811	1.9307	0.9333	0.9936	0.9444
		PCA	7.1969	2.8118	0.9417	0.9936	0.9167
		MDS	5.9636	4.0949	0.8972	0.9936	0.8889
		LDA	4.9417	0.1151	0.9333	0.9808	0.9722
	ADV	Original	4.0949	4.0949	0.9278	0.9743	0.8056
		PCA	5.9636	2.3299	0.9222	0.9743	0.8333
		MDS	10.4811	2.8118	0.9444	0.9872	0.8611
		LDA	8.6851	0.0449	0.9639	0.9872	0.9167

benzaldehyde, 1-octen-3-ol, β-cyclociral, citral, nerolidol and methyl jasmonate showed the same changing trend as linalool and nonanal, revealing the positive influence of these volatile components to tea aroma. The contents of benzeneacetaldehyde, phenylethyl alcohol, naphthalene, methyl salicylate and indole declined as the tea grade going up, manifesting that these volatile components may had negative contribution to tea perception. It could be found that the concentrations of the newly generated volatile components in tea infusion like indole, jasmone, linalool and nerolidol were changed with the variation of tea grades, which manifested that tea infusion might better characterize the differences among different tea grades than tea leaf did.

The correlations between E-nose signals and volatile components

In order to investigate the potential correlation between the content of volatile components and E-nose signals,

regression models were set up to quantitatively predict the content of volatile components. Tea infusion samples contained more volatile components and could be superiorly classified than tea leaf samples did. Therefore, the regression models were built based on E-nose signals obtained from tea infusion samples. As shown in Table 4, the contents of linalool, nonanal and geraniol were related to tea grades and were the highest among all the detected volatile components in tea infusion. Hence, the three volatile components were chosen for the following quantitative analysis. SVM was applied to set up regression models. To obtain the best performance, the radial basis function (RBF) was chosen as the core function and 10-fold cross validation was applied in SVM. Penalty parameter C and kernel parameter gamma (γ) were optimized by using a grid search method in SVM regression models to acquire the optimal model. The regression results were shown in Table 5. Squared correlation coefficient (R^2) as well as root mean square of error (RMSE) were used to assess the performance of the prediction models. As shown

Table 4 Volatile components in tea leaf and tea infusion of different grades by GC–MS

Tea samples	No	RI[a]	Voilate components	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6
Tea leaf	1	955	Benzaldehyde	0.9667 ± 0.0063	0.9422 ± 0.0435	0.7432 ± 0.0506	0.7745 ± 0.0764	0.7628 ± 0.0870	0.6093 ± 0.0390
	2	1023	D-Limonene	0.8631 ± 0.0571	0.6941 ± 0.0474	0.4388 ± 0.0251	0.4002 ± 0.0314	0.4318 ± 0.0535	0.3797 ± 0.0463
	3	1039	Benzeneacetaldehyde	0.2861 ± 0.0379	0.3550 ± 0.0647	0.3248 ± 0.0362	0.3604 ± 0.0287	0.2269 ± 0.0246	0.2779 ± 0.0258
	4	1069	1-Octanol	1.3972 ± 0.1133	1.5089 ± 0.1617	1.0023 ± 0.1327	1.0439 ± 0.1134	0.7800 ± 0.0262	1.0036 ± 0.1081
	5	1104	Nonanal	5.3231 ± 0.5593	5.2546 ± 0.6598	4.8897 ± 0.6450	4.5399 ± 0.1719	4.7337 ± 0.2369	4.2763 ± 0.4460
	6	1111	Phenylethyl alcohol	2.3471 ± 0.3261	2.1064 ± 0.2551	2.4099 ± 0.3070	2.7639 ± 0.1592	3.0273 ± 0.1347	2.9764 ± 0.2459
	7	1175	Naphthalene	0.4769 ± 0.0570	0.4813 ± 0.159	0.4137 ± 0.0414	0.3966 ± 0.0489	0.5343 ± 0.0249	0.6146 ± 0.0521
	8	1187	α-Terpineol	0.9015 ± 0.0598	0.8517 ± 0.0370	0.6598 ± 0.0733	0.8376 ± 0.0233	0.6483 ± 0.0234	0.4835 ± 0.0286
	9	1190	Methyl salicylate	0.4446 ± 0.0328	0.4902 ± 0.0223	0.6200 ± 0.0545	0.5804 ± 0.0307	0.5094 ± 0.0406	0.5235 ± 0.0715
	10	1205	Decanal	2.2042 ± 0.1530	2.2409 ± 0.4159	1.5872 ± 0.0667	1.3008 ± 0.1663	1.0675 ± 0.1310	1.2980 ± 0.1206
	11	1217	β-Cyclociral	0.4414 ± 0.0143	0.4294 ± 0.0529	0.4045 ± 0.0348	0.4169 ± 0.0372	0.3798 ± 0.0084	0.2911 ± 0.0316
	12	1300	Tridecane	1.1401 ± 0.1486	1.5785 ± 0.0577	1.5929 ± 0.1093	2.0249 ± 0.1817	1.3498 ± 0.1210	1.4526 ± 0.2381
	13	1384	(Z)-3-hexenyl hexanoate	2.3918 ± 0.4010	3.4518 ± 0.2218	3.4291 ± 0.5309	5.8061 ± 0.1646	4.3055 ± 0.3257	2.7531 ± 0.0554
	14	1400	Tetradecane	2.9982 ± 0.2620	2.9333 ± 0.1169	3.2221 ± 0.3833	2.8609 ± 0.2644	3.4062 ± 0.3941	4.3463 ± 0.1434
	15	1406	α-Cedrene	1.1554 ± 0.1004	1.0148 ± 0.0596	0.7396 ± 0.0333	0.9298 ± 0.0202	0.8352 ± 0.0623	1.3638 ± 0.0539
	16	1415	β-Cedrene	0.6088 ± 0.0425	0.5126 ± 0.0644	0.3844 ± 0.0399	0.2683 ± 0.0074	0.5315 ± 0.0455	0.6451 ± 0.0603
	17	1428	α-Ionone	0.2231 ± 0.0123	0.2963 ± 0.0335	0.3229 ± 0.0143	0.2166 ± 0.0083	0.3046 ± 0.0305	0.3534 ± 0.0298
	18	1459	β-Farnesene	0.5189 ± 0.0565	0.3657 ± 0.1507	0.2969 ± 0.0114	0.2714 ± 0.0442	ND	ND
	19	1455	Geranyl acetone	0.7641 ± 0.1542	1.0248 ± 0.1175	1.2350 ± 0.0730	0.9156 ± 0.0284	1.0210 ± 0.427	1.6496 ± 0.0615
	20	1485	β-Ionone	0.7836 ± 0.0815	0.8008 ± 0.0423	0.8426 ± 0.0763	0.7281 ± 0.0860	0.9571 ± 0.1152	1.0181 ± 0.1402
	21	1500	Pentadecane	1.4075 ± 0.2373	1.2362 ± 0.2863	1.2189 ± 0.0859	1.1506 ± 0.1531	1.5184 ± 0.0393	2.1824 ± 0.5294
	22	1522	Cadinene	1.3145 ± 0.1570	1.0216 ± 0.0887	0.7406 ± 0.0481	0.8059 ± 0.0458	1.2005 ± 0.0528	1.5816 ± 0.2364
	23	1600	Hexadecane	1.4588 ± 0.0815	1.2594 ± 0.2921	1.1253 ± 0.3454	0.9589 ± 0.3676	1.4785 ± 0.1289	2.2321 ± 0.3073
	24	1700	Heptadecane	0.2586 ± 0.0479	0.3062 ± 0.0375	0.2379 ± 0.0135	0.2422 ± 0.0073	0.2659 ± 0.0179	0.3342 ± 0.0251
Tea infusion	1	955	Benzaldehyde	287 ± 32	259 ± 18	259 ± 20	240 ± 20	262 ± 27	247 ± 19
	2	977	1-Octen-3-ol	111 ± 10	107 ± 8	99 ± 10	95 ± 12	93 ± 4	87 ± 9
	3	989	β-Myrcene	448 ± 42	442 ± 57	405 ± 32	387 ± 31	418 ± 45	446 ± 28
	4	1039	Benzeneacetaldehyde	232 ± 45	239 ± 31	318 ± 28	324 ± 11	336 ± 33	358 ± 28
	5	1069	1-Octanol	1065 ± 67	1001 ± 119	924 ± 67	884 ± 53	954 ± 30	1106 ± 79
	6	1099	Linalool	3136 ± 242	2937 ± 184	3012 ± 147	2974 ± 224	2717 ± 161	2649 ± 19
	7	1104	Nonanal	1748 ± 154	1627 ± 157	1547 ± 115	1570 ± 125	1449 ± 57	1312 ± 167
	8	1111	Phenylethyl Alcohol	223 ± 15	248 ± 25	317 ± 27	346 ± 32	350 ± 12	347 ± 8
	9	1175	Naphthalene	120 ± 14	103 ± 11	148 ± 12	134 ± 7	219 ± 29	180 ± 17
	10	1187	α-Terpineol	428 ± 57	420 ± 32	446 ± 25	455 ± 37	439 ± 16	391 ± 34
	11	1190	Methyl salicylate	223 ± 27	287 ± 22	299 ± 24	275 ± 12	331 ± 27	377 ± 16
	12	1196	Safranal	113 ± 13	95 ± 10	147 ± 19	160 ± 15	149 ± 11	103 ± 5

Table 4 continued

Tea samples	No	RI^a	Voilate components	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6
	13	1205	Decanal	825 ± 50	827 ± 91	657 ± 38	787 ± 46	839 ± 47	931 ± 72
	14	1217	β-Cyclociral	181 ± 8	172 ± 18	168 ± 7	156 ± 18	154 ± 12	151 ± 13
	15	1254	Geraniol	2021 ± 174	1941 ± 239	1777 ± 146	1524 ± 98	1894 ± 155	2124 ± 130
	16	1269	Citral	124 ± 13	113 ± 18	123 ± 10	89 ± 11	98 ± 16	109 ± 7
	17	1289	Indole	185 ± 25	195 ± 13	174 ± 17	206 ± 25	251 ± 31	232 ± 18
	18	1351	Cubebene	72 ± 8	65 ± 2	55 ± 6	53 ± 5	77 ± 5	74 ± 6
	19	1384	(Z)-3-hexenyl hexanoate	158 ± 16	178 ± 18	209 ± 10	212 ± 14	207 ± 20	137 ± 10
	20	1397	Jasmone	1116 ± 59	1060 ± 68	ND	ND	ND	ND
	21	1400	Tetradecane	75 ± 13	75 ± 2	71 ± 9	62 ± 5	77 ± 4	77 ± 6
	22	1406	α-Cedrene	15 ± 1.6	16 ± 0.6	15 ± 0.8	14 ± + 1.1	18 ± 0.7	21 ± 0.7
	23	1428	α-Ionone	53 ± 10	61 ± 12	58 ± 7	48 ± 2	68 ± 7	55 ± 8
	24	1455	Geranyl acetone	371 ± 68	417 ± 51	439 ± 38	466 ± 66	342 ± 6	362 ± 41
	25	1485	β-Ionone	472 ± 28	507 ± 58	515 ± 54	425 ± 19	494 ± 83	459 ± 29
	26	1500	Pentadecane	158 ± 15	148 ± 8	139 ± 12	147 ± 6	166 ± 23	165 ± 19
	27	1522	Cadinene	116 ± 8	105 ± 9	94 ± 6	97 ± 11	118 ± 13	135 ± 8
	28	1564	Nerolidol	112 ± 13	102 ± 9	87 ± 3	87 ± 4	77 ± 8	76 ± 4
	29	1649	Methyl jasmonate	138 ± 11	153 ± 17	134 ± 22	124 ± 14	104 ± 6	96 ± 7
	30	1700	Heptadecane	116 ± 14	118 ± 7	113 ± 14	106 ± 12	120 ± 6	122 ± 10

% Mean percentage and standard deviations of tea leaf constituents from triplicate experiments

ND not detected

^a Oder of elution are given on nonpolar column (HP-5 MS or HP-5)

Table 5 Comparison of SVM results based on different data reductions of predicting volatile components in tea infusion

Volatile components	Data reductions	Parameters		Training set		Testing set	
		C	γ	R^2	RMSE	R^2	RMSE
Linalool	Original	3.3932	4.0949	0.9611	0.0045	0.8529	0.0163
	PCA	68.6649	2.3299	0.9655	0.0040	0.8662	0.0159
	MDS	22.2299	3.3932	0.9574	0.0049	0.8654	0.0157
	LDA	0.9103	0.01	0.9523	0.0056	0.9343	0.0077
Nonanal	Original	12.6486	4.0949	0.9545	0.0045	0.8205	0.0181
	PCA	8.6851	4.0949	0. 9545	0.0050	0.8207	0.0198
	MDS	1.3257	12.6486	0.9580	0.0046	0.7696	0.0373
	LDA	4.9417	0.01	0.9617	0.0047	0.8980	0.0094
Geraniol	Original	3.3932	0.9103	0.9582	0.0043	0.8800	0.0122
	PCA	82.8643	0.7543	0.9609	0.0039	0.8289	0.0175
	MDS	10.4811	1.0985	0.9622	0.0039	0.7847	0.0219
	LDA	0.3556		0.9576	0.0043	0.9315	0.0069

in Table 5, the prediction results based on original data were not satisfactory for all the three volatile components. PCA and MDS process had no positive contribution to increase prediction accuracy. LDA process provided positive contribution for the prediction accuracy for all the three volatile components (Linalool: $R^2 = 0.9343$; Nonanal: $R^2 = 0.8980$; Geraniol: $R^2 = 0.9315$), indicating the effectiveness of LDA in improving the results of regression models. The results indicated that the LDA processed E-nose signals combined with SVM could be effectively applied for quantitatively predicting the content of volatile compounds contained in tea samples.

Conclusion

In this research, E-nose was employed to evaluate tea quality by detecting tea leaf and tea infusion samples. The volatile components of tea samples were analyzed by GC–MS as reference. LR and SVM were applied to set up classification models. The performance of data extractions including 35th s value, 70th s value and ADV, as well as data reductions including PCA, MDS and LDA were compared. The SVM model based on the data extraction of 70th s value combined with LDA reduction achieved satisfactory classification performance with the accuracy of 100%. Meanwhile, good correlations were found between

the contents of main volatile components in tea infusion including linalool ($R^2 = 0.9343$), nonanal ($R^2 = 0.8980$) and geraniol ($R^2 = 0.9315$) and the E-nose signals of 70th s value. The E-nose that composed of 10 metal-oxide based gas sensors could be successfully employed for qualitatively identifying tea quality grades and quantitative analyzing the content of main volatile compounds with the changing of tea quality grades. The research manifested the feasibility of applying E-nose to evaluate tea quality by characterizing tea aroma with electronic signals.

References

Akbar MA, Ali AAS, Amira A, Bensaali F, Benammar M, Hassan M, Bermak A (2016) An empirical study for PCA-and LDA-based feature reduction for gas identification. IEEE Sens J 16(14):5734–5746

Brudzewski K, Osowski S, Markiewicz T (2004) Classification of milk by means of an electronic nose and SVM neural network. Sens Actuat B Chem 98(2–3):291–298

Bhattacharyya N, Seth S, Tudu B et al (2007) Detection of optimum fermentation time for black tea manufacturing using electronic nose. Sens Actuator B Chem 122(2):627–634

Cortes C, Vapnik V (1995) Support-vector networks. Mach Learning 20(3):273–297

Chen YS, Liu BL, Chang YN (2010) Bioactivities and sensory evaluation of Pu-erh teas made from three tea leaves in an improved pile fermentation process. J Biosci Bioeng 109(6):557–563

Chen Q, Zhao J, Chen Z, Lin H, Zhao DA (2011) Discrimination of green tea quality using the electronic nose technique and the human panel test, comparison of linear and nonlinear classification tools. Sens Actuator B Chem 159(1):294–300

Choi SI, Jeon HM, Jeong GM (2017) Data Reconstruction Using Subspace Analysis for Gas Classification. IEEE Sens J 17(18):5954–5962

Dymerski T, Gębicki J, Wardencki W, Namieśnik J (2014) Application of an electronic nose instrument to fast classification of polish honey types. Sensors 14(6):10709–10724

Dai Y, Zhi R, Zhao L, Gao H, Shi B, Wang H (2015) Longjing tea quality classification by fusion of features collected from E-nose. Chemom Intell Lab Syst 144:63–70

Guohua H, Yuling W, Dandan Y, Wenwen D (2013) Fuji apple storage time predictive method using electronic nose. Food Anal Method 6(1):82–88

Huang XC, Guo CF, Yuan YH, Luo XX, Yue TL (2015) Detection of medicinal off-flavor in apple juice with artificial sensing system and comparison with test panel evaluation and GC–MS. Food Control 51:270–277

Jiang S, Wang J, Sun Y (2017) Qualitative and quantitative analysis of fatty acid profiles of Chinese pecans (Carya cathayensis) during storage using an electronic nose combined with chemometric methods. RSC Adv 7(73):46461–46471

Krstic M, Stojadinovic M, Smiljanic K, Stanic-Vucinic D, Velickovic TC (2015) The anti-cancer activity of green tea, coffee and cocoa extracts on human cervical adenocarcinoma HeLa cells depends on both pro-oxidant and anti-proliferative activities of polyphenols. RSC Adv 5(5):3260–3268

Kramer O (2016) Dimensionality reduction hybridizations with multidimensional scaling. In: Merenyi E, Mendenhall MJ, O'Driscoll P (eds) Advances in self-organizing maps and learning vector quantization. Springer, Cham, pp 155–163

Kodogiannis VS (2017) Application of an electronic nose coupled with fuzzy-wavelet network for the detection of meat spoilage. Food Bioprocess Technol 10(4):730–749

Lin J, Dai Y, Guo YN, Xu HR, Wang XC (2012) Volatile profile analysis and quality prediction of Longjing tea (Camellia sinensis) by HS-SPME/GC-MS. J Zhejiang Univ Sci B 13(12):972–980

Lai CQ, Teoh SS (2016) An Efficient Method of HOG Feature Extraction Using Selective Histogram Bin and PCA Feature Reduction. Adv Electr Comput Eng 16(4):101–108

Majcher MA, Kaczmarek A, Klensporf-Pawlik D, Pikul J, Jeleń HH (2015) SPME-MS-based electronic nose as a tool for determination of authenticity of PDO cheese. Oscypek Food Anal Methods 8(9):2211–2217

Prabhakar TN, Xavier G, Geetha P, Soman KP (2015) Spatial preprocessing based multinomial logistic regression for hyperspectral image classification. Procedia Comput Sci 46:1817–1826

Pławiak P, Maziarz W (2014) Classification of tea specimens using novel hybrid artificial intelligence methods. Sens Actuators B Chem 192:117–125

Pasquini B, Orlandini S, Goodarzi M, Caprini C, Gotti R, Furlanetto S (2016) Chiral cyclodextrin-modified micellar electrokinetic chromatography and chemometric techniques for green tea samples origin discrimination. Talanta 150:7–13

Qin Z, Pang X, Chen D, Cheng H, Hu X, Wu J (2013) Evaluation of Chinese tea by the electronic nose and gas chromatography–mass spectrometry: correlation with sensory properties and classification according to grade level. Food Res Int 53(2):864–874

Qiu S, Wang J, Du D (2017) Assessment of high pressure processed mandarin juice in the headspace by using electronic nose and chemometric analysis. Innov Food Sci Emerg 42:33–41

Qiu S, Wang J (2017) The prediction of food additives in the fruit juice based on electronic nose with chemometrics. Food Chem 230:208–214

Rawat R, Gulati A, Babu GK, Acharya R, Kaul VK, Singh B (2007) Characterization of volatile components of Kangra orthodox black tea by gas chromatography-mass spectrometry. Food Chem 105(1):229–235

Torgerson WS (1952) Multidimensional scaling: I. Theory and method. Psychometrika 17(4):401–419

Tian X, Wang J, Cui S (2013) Analysis of pork adulteration in minced mutton using electronic nose of metal oxide sensors. J Food Eng 119(4):744–749

Unachukwu UJ, Ahmed S, Kavalier A, Lyles JT, Kennelly EJ (2010) White and green teas (Camellia sinensis var. sinensis): variation in phenolic, 8-methylxanthine, and antioxidant profiles. J Food Sci 75(6):541–548

Wang L, Wei K, Cheng H, He W, Li X, Gong W (2014) Geographical tracing of Xihu Longjing tea using high performance liquid chromatography. Food Chem 146:98–103

Wu H, Wang J, Yue T, Yuan Y (2017) Variety-based discrimination of apple juices by an electronic nose and gas chromatography–mass spectrometry. Int J Food Sci Technol 52(11):2324–2333

Xu M, Ye L, Wang J, Wei Z, Cheng S (2017) Quality tracing of peanuts using an array of metal-oxide based gas sensors combined with chemometrics methods. Postharvest Biol Technol 128:98–104

Yu H, Wang J, Xiao H, Liu M (2009) Quality grade identification of green tea using the eigenvalues of PCA based on the E-nose signals. Sens Actuators B Chem 140(2):378–382

Yao Y, Pan S, Fan G, Dong L, Ren J, Zhu Y (2015) Evaluation of volatile profile of Sichuan dongcai, a traditional salted vegetable, by SPME–GC–MS and E-nose. LWT Food Sci Technol 64(2):528–535

Yang CJ, Ding W, Ma LJ, Jia R (2015) Discrimination and characterization of different intensities of goaty flavor in goat milk by means of an electronic nose. J Dairy Sci 98(1):55–67