Classification of Electronic Nose Data Using the Least Squares Support Vector Machine

Gaofeng Chen¹, Guifang Wu*¹
¹College of Mechanical and Electrical Engineering, Inner Mongolia Agricultural University, 306, Zhaowuda Road, Huhhot, 010018, P.R.China

* Corresponding author: Tel: +86-471-4309223; E-mail address: wgfsara@126.com (Guifang Wu)

Abstract—In this study, the response signals of three kinds of dry alfalfa volatile odors were collected by an electronic nose (E-nose), and the collected data were processed by principal component analysis (PCA) and linear discriminant analysis (LDA). A least squares support vector machine (LS-SVM) model was established to classify and evaluate the data. For the combined E-nose algorithm, the classification accuracies of the PCA-LS-SVM and LDA-LS-SVM models are 85% and 100%, respectively. LDA as the input model has better classification accuracy than the PCA-based model. The results show that the combination of the LDA and LS-SVM algorithms using an E-nose signal is effective in identifying different drying alfalfa. The performance of the LDA-based LS-SVM model is slightly higher than that of the PCA-based LS-SVM model. It can be concluded that the E-nose system combined with the LDA-based model has great potential to distinguish different dry alfalfa.

1. INTRODUCTION

Alfalfa, also referred to as perennial leguminous forage, or “forage crown”, not only produces a high yield but also exhibits good quality. This plant is not only good forage for livestock but also is a dish favored by consumers worldwide. Alfalfa is a high-quality food that can provide abundant protein. In addition to being used as feed, it can be directly eaten and has good nutritional, medical and health functions.

With the continuous improvement in people’s living standards, consumers have been paying more attention to the nutrition of healthy food. As a result, alfalfa is becoming increasingly popular among individuals. Every spring, many people eat the tender leaves of alfalfa, which are restricted for seasonal reasons. Only through dehydration and drying can alfalfa maintain its original quality, eating period and a fresh form to meet the requirements of consumers. The characteristics and quality of variously dried alfalfa products are different. The correct classification of dried alfalfa products in quality assurance is an important issue facing the commercial market.

Traditionally, hay classification usually uses sensory methods to determine observable characteristics, such as the color, odor, herbage species and inorganic hybrids, utilizes chemical methods to determine nutrient composition and uses digestibility tests to determine the solubility of the soluble content. The digestible nutrient content and aroma measurement technology gas chromatography–mass spectrometry (GC-MS), analysis) is usually used in production practice. Because these methods take a long time and are expensive, they also need special equipment and skilled operators with considerable experience. An electronic nose (E-nose) is a tool that imitates olfaction [8,9]. Using the chemical sensor array of the device, volatile substances can be detected and distinguished. Specifically, the device is designed to use...
a series of chemical sensors to detect and distinguish mixed substances [10]. The system can provide real-time information, which greatly facilitates the use of an E-nose to quickly monitor volatile components [11], and as a result, it can compensate for the shortcomings of traditional methods.

The E-nose system is designed to detect and distinguish complex volatile substances. The sensor group in the E-nose system consists of a series of nonspecific sensors that can generate fingerprint features of the sensor array. Aromatic patterns or fingerprints from known aromas are used to establish database systems and training models so that unknowns can be classified and identified[4]. E-nose technology has been successfully applied to agriculture [5], medicine [6], clinical diagnosis [10,11], environmental control [7] and other fields. During the past decade, E-nose technology has been mainly used to identify and analyze various food and agricultural products, such as milk [14], edible oil [15,16], beverages [12,13], fish [19], meat [17,18], vegetables [20,21], and fruits [22,23]. Current studies have mostly focused on the classification and recognition of tomatoes, peaches, citrus fruits, apples and other agricultural products. For animal husbandry, the classification of high-quality forage, grass products and livestock products has not been reported, and the real-time application and online recognition model based on E-nose technology has not yet been established. To solve this problem, the main work of this paper is as follows: first, the E-nose is used to collect the aroma data of different dry alfalfa, then principal component analysis (PCA) and linear discriminant analysis (LDA) are used to reduce the dimension, and finally, the least squares support vector machine (LS-SVM) model is established to evaluate the classification effect.

2. EXPERIMENTAL

2.1. Preparation of the Samples
First, three kinds of dry alfalfa, which were sun-dried, shady dried and mildewed, were selected from the laboratory of Inner Mongolia Agricultural University. Each dry alfalfa sample was ground into powder by a high-speed electric grinder, and then some impurities were removed by a 1 mm screen. Five grams of alfalfa powder were weighed and loaded into a 50 ml test tube. Thirty-five dry alfalfa samples were prepared for each kind, and a total of 105 samples were prepared. Finally, these samples were sealed with fresh-keeping film for the experiment.

2.2. Electronic Nose System
The experimental device is a metal oxide sensor (MOS) system based on an E-nose designed by PEN3 (Airsense Analytics, Germany) (Fig. 1). The system consists of five parts: a sample room, a sensor room, a data acquisition system, a control unit and a graphical user interface. Other parts in the system include pipes, electric valves and pumps. The sample chamber is a cylindrical 40 ml glass bottle, which is connected to the sensor room through an air diaphragm pump. Ten heating MOSs and temperature sensors are installed in the sensor room.

The measurement process is as follows: ambient air flows into and out of the sensor room for 240 seconds. This process ensures that the sensor chamber is free from any previous volatile interferences, while enabling the sensor to reach a stable baseline. The sample chamber is connected to the sensor chamber to absorb the sample odor into the sensor chamber for a period of 240 seconds. Then, the sensor room is swept for 240 seconds. During the measurement, the response signals of each sensor are recorded (Fig. 2). Ninety seconds is sufficient to stabilize the sensor. In this process, the volatile gas is pumped to the sensor room at a flow rate of 400 ml/min. The sampling interval is 1 second. At the end of the measurement, the average value of the last 5 seconds is selected as the characteristic variable. Therefore, each sample can extract 10 feature data points from the original sensor.
3. RESULTS

3.1. Principal Component Analysis (PCA)
PCA is suitable for data sets to identify different types of models. PCA applications convert the acquired nose data into two or three components to produce representative information data. The purpose of the transformation is to transform high-dimensional data into low-dimensional data to minimize the relevance of the original data set. The PCA diagram shows the similarities that are grouped together to extract functionality and reduce the dimensions. The results of the PCA are shown in Fig.3. From the PCA diagram, that different samples that belong to different categories exist independently. The first component explained 49.366% of the variance, the second component explained 20.688% of the variance, and the third component explained 17.117% of the variance. The cumulative contribution rate of the three components was 87.171%. There are overlapping phenomena on the edges of the three samples, but only individual samples are distributed separately. The results show that PCA cannot completely separate the three samples.
3.2. Linear Discriminant Analysis (LDA)

LDA is a commonly used method in the field of chemometrics and is widely used for classifying data. In our previous work, LDA was successfully used in various areas of classification. LDA aims to find a linear combination of the feature groups to represent the different types of interests. This functional combination is later used in combination with a linear classifier to perform the classification. However, for two types of separation cases, it was later defined as a multilevel classification case. Compared with complex classification, LDA has the advantages of simplicity and can ensure the accuracy of the model.

In the current work, a step-by-step LDA program is used to visualize the sample dispersion. This step-by-step discriminant method is used to select variables. In each step, the input variables, which can minimize the overall Wilks’ lambda distribution, include the maximum number of steps (20), the minimum partial F to be input (3.84), and the maximum partial F to be removed (2.71). As shown in Fig.4, two discriminant functions (LD) explain 100% of the total variance of the discriminant function (75.7% for the first and 24.3% for the second). It was found that three kinds of dry samples could be easily distinguished, and no overlap was found in all samples. Therefore, the LDA classification function is more effective than PCA.

3.3. Least Squares Support Vector Machine (LS-SVM)

As a nonlinear classification model, the (LS-SVM model is used) in this work. LS-SVM comes from the classification of two kinds of problems, in which LS-SVM can be considered as creating two sets of classification data of the hyperplane. If the linear boundary in a low-dimensional space is insufficient to divide into two categories, then a hyperplane can be created to allow linear partitioning into larger features. In LS-SVM, data are converted from low-dimensional space to high-dimensional space from the conversion function. The result is realized by the kernel function, which maps the indivisible data into the high-dimensional space and divides it into the high-dimensional space linearly. Generally, the LS-SVM model has three kernels: polynomial, Gaussian and sigmoid. Kernel functions have a great
impact on the LS-SVM performance. The functions of the three cores, namely, the Gaussian kernel function, are simple in structure and fast in operation. The Gaussian kernel function is usually the best choice. Therefore, only the Gaussian kernel function is tested, and its structure is a radial basis function (RBF), so it is called the RBF kernel function.

To optimize the performance of the model, two parameters ($\gamma$, $\sigma^2$) in the LS-SVM model must be optimized. In this experiment, for the LS-SVM classifier, cross-validation technology is used in combination with grid acquisition. The network searches for values in a specified area ($\gamma$, $\sigma^2$). Cross validation is used to avoid over fitting.

In the PCA-LS-SVM model, the optimal parameters specified by grid search are (20, 0.1). In the five-fold cross-validation procedure, the three-dimensional data sets are randomly divided into five equal parts. Four of them are selected for training and the remaining one is selected for testing. This process is repeated five times, and then the average value of the result is calculated as the classification accuracy. As shown in Fig. 5, the LS-SVM model can classify odor characteristics, and with PCA as the input, this model has a final classification accuracy of 95.238%. Further understanding shows that two dry samples were misclassified as sun-dried samples, six dry samples were misclassified as sun-dried samples, two moldy samples were classified as sun-dried samples, and a few of them were not identified. Therefore, it can be concluded that three kinds of dry alfalfa can be separated normally using odor characteristics.

In the LDA-LS-SVM model, the optimal parameters of the grid search are (10, 0.02). Similarly, two-dimensional data sets are divided into five equal parts, 20% of which are selected as tests, and the remaining 80% are selected as training. The process was repeated five times, with the average as the classification result. As shown in Fig. 6, three kinds of dried alfalfa, i.e., sun-dried, shade-dried and mildewed are distributed separately in a two-dimensional plane and can be completely separated. The results show that LDA can be used as input to separate the odor characteristics with a 100% accuracy.

![LS-SVM classification effect map based on PCA](image)
4. CONCLUSION
In this work, it is proven that an E-nose combined with the LS-SVM model can distinguish dry alfalfa of different qualities. A comparison of the classification results of the PCA and LDA shows that LDA-LS-SVM achieves the optimal discrimination results. The accuracy is 100%. This proves that LDA can extract the odor features of the classification. The LS-SVM classification model shows the feasibility and potential of the scheme, thus providing a fast and easy-to-operate new method for the market.

ACKNOWLEDGMENT
This study was supported by the National Natural Science Foundation of China (Project No: 32060414), Inner Mongolia Natural Science Foundation (Project No: 2018MS06012), Research Fund for the Doctoral Program of Higher Education (Project No: 20111515120004).

REFERENCES
[1] M. Peris and L. Escuder-Gilabert, “A 21st century technique for food control: Electronic noses,” Analytica Chimica Acta, vol. 638, pp. 1-15, April 2009.
[2] M. Ghasemi-Varnamkhisti, S. Mohtasebi, M. Rodriguez-Mendez, J. Lozano, S. Razavi, and H. Ahmadi, “Potential application of electronic nose technology in brewery,” Trends in Food Science & Technology, vol. 22, pp. 165-174, April 2011.
[3] J. Lozano, T. Arroyo, J. Santos, J. Cabellos, and M. Horrillo, “Electronic nose for wine ageing detection,” Sensors and Actuators B: Chemical, vol. 133, pp. 180-186, July 2008.
[4] M. Ghasemi-Varnamkhisti, S. S. Mohtasebi, and M. Siadat, “Biomimetic-based odor and taste sensing systems to food quality and safety characterization: An overview on basic principles and recent achievements,” Journal of food engineering, vol. 100, pp. 377-387, October 2010.
[5] M. Peris and L. Escuder-Gilabert, “A 21st century technique for food control: electronic noses,” Analytica Chimica Acta, vol. 638, 2009, pp. 1-15.
[6] H.M. Zhang and J. Wang, “Detection of age and insect damage incurred by wheat, with an electronic nose,” Journal of Stored Products Research, vol. 43, 2007, pp. 489-495.
[7] L.M Zhu, R.A. Seburg, E. Tsai, S. Puech, and J.C. Mifsud, “Flavor analysis in a pharmaceuti-cal oral solution formulation using an electronic-nose,” Journal of Pharmaceutical and Biomedical Analysis, vol. 34, 2004, pp. 453-461.
[8] M. Bernabei, G. Pennazza, M. Santortico, C. Corsi, C. Roscioni, R. Paolesse, C. Di Natale, and A. D’Amico, “A preliminary study on the possibility to diagnose urinary tract cancers by an electronic nose,” Sensors and Actuators B: Chemical, vol. 131, pp. 1-4, April 2008.
[9] C. Di Natale, A. Macagnano, E. Martinelli, R. Paolesse, G. D’Arcangelo, C. Roscioni, A. Finazzi-Agro, and A. D’Amico, “Lung cancer identification by the analysis of breath by means of an array of non-selective gas sensors,” Biosensors and Bioelectronics, vol. 18, 2003, pp. 1209-1218.

[10] Q. Ameer and S.B. Adeloju, “Polypyrrole-based electronic noses for environmental and industrial analysis,” Sensors and Actuators B: Chemical, vol. 106, 2005, pp. 541-552.

[11] B. Wang, S.Y. Xu, and D.W. Sun, “Application of the electronic nose to the identification of different milk flavors,” Food Research International, vol. 43, 2010, pp. 255-262.

[12] C. Apetrei, I.M. Apetrei, S. Villanueva, J.A. de Saja, F. Gutierrez-Rosales, and M.L. Rodriguez-Mendez, “Combination of an e-nose, an e-tongue and an e-eye for the characterisation of olive oils with different degree of bitterness,” Analytica Chimica Acta, vol. 663, pp. 91-97, July 2010.

[13] M.J. Lerma-Garcia, E.F. Simo-Alfonso, A. Bendini, and L. Cerretani, “Metal oxide semiconductor sensors for monitoring of oxidative status evolution and sensory analysis of virgin olive oils with different phenolic content,” Food Chemistry, vol. 117, 2009, pp. 608-614.

[14] J.A. Ragazzo-Sanchez, R. Chalier, D. Chevalier, M. Calderon-Santoyo, and C. Ghom-midh, “Identification of different alcoholic beverages by electronic nose coupled to GC,” Sensors and Actuators B: Chemical, vol. 134, 2008, pp. 43-48.

[15] H. Reinhard, F. Sager, and O. Zoller, “Citrus juice classification by SPME-GC-MS and electronic nose measurements,” Lwt-Food Science and Technology, vol. 41, 2008, pp. 1906-1912.

[16] N. El Barbri, J. Mirhisse, R. Ionescu, N. El Bari, X. Correig, B. Bouchikhi, and E. Llo-bet, “An electronic nose system based on a micro-machined gas sensor array to assess the freshness of sardines,” Sensors and Actuators B: Chemical, vol. 141, pp. 538-543, September 2009.

[17] S. Balasubramaninan, S. Panigrahi, C.M. Logue, C. Doetkott, M. Marchello, and J.S. Sherwood, “Independent component analysis-processed electronic nose data for predicting Salmonella typhimurium populations in contaminated beef,” Food Control, vol. 19, pp. 236-246, March 2008.

[18] J.S. Vestergaard, M. Martens, and P. Turkki, “Application of an electronic nose system for prediction of sensory quality changes of a meat product (pizza topping) during storage,” Lwt-Food Science and Technology, vol. 40, 2007, pp. 1095-1101.

[19] I. Concina, M. Falasconi, E. Gobbi, F. Bianchi, M. Musci, M. Mattarozzi, M. Pardo, A. Mangia, M. Careri, and G. Sberveglieri, “Early detection of microbial contamination in processed tomatoes by electronic nose,” Food Control, vol. 20, pp. 873-880, October 2009.

[20] A.H. Gomez, J. Wang, G.X. Hu, and A.G. Pereira, “Monitoring storage shelf life of tomato using electronic nose technique,” Journal of Food Engineering, vol. 85, pp. 625-631, April 2008.

[21] A.H. Gomez, J. Wang, G.X. Hu, and A.G. Pereira, “Discrimination of storage shelf-life for mandarin by electronic nose technique,” Lwt-Food Science and Technology, vol. 40, 2007, pp. 681-689.

[22] H.M. Zhang, M.X. Chang, J. Wang, and S. Ye, “Evaluation of peach quality indices using an electronic nose by MLR, QPST and BP network,” Sensors and Actuators B: Chemical, vol. 134, pp. 332-338, August 2008.

[23] M. Moghaddam, F. B. H. Ahmad, M. Basri, and M. B. A. Rahman, “Artificial neural network modeling studies to predict the yield of enzymatic synthesis of betulinic acid ester,” Electronic Journal of Biotechnology, vol. 13, pp. 3-4, January 2010.

[24] S. Panigrahi, S. Balasubramaninan, H. Gu, C. Logue, and M. Marchello, “Neural-network-integrated electronic nose system for identification of spoiled beef,” LWT, vol. 39, 2006, pp. 135-145.

[25] D. Dutta, E.L. Hines, J.W. Gardner, K.R. Ksahwan, and M. Bhuyan, “Tea quality prediction using a tin oxide-based electronic nose: an artificial intelligence approach,” Sensors and Actuators B, vol. 94, 2003, pp. 228-237.
[26] A.M. Pineiro, A. Fisher, and S.J. Hill, “The classification of tea according to region of origin using pattern recognition techniques and trace metal data,” J. Food Compos. Anal., vol. 16, 2003, pp. 195-211.

[27] S. Laberche, S. Bazzo, S. Cade, and E. Chanie, “Shelf life determination by electronic nose: application to milk,” Sens. Actuators B, vol. 106, pp. 199–206, April 2005.