Indonesian Journal of Electrical Engineering and Computer Science
Vol. 5, No. 3, March 2017, pp. 530 ~ 535
DOI: 10.11591/ijeecs.v5.i3.pp530-535

Aromatic Herbs Classification by Using Discriminant Analysis Techniques

N. F. M. Radzi*, A. Che Soh, A. J. Ishak, M. K. Hassan, U. K. Mohamad Yusof
Department of Electrical and Electronic Engineering, Universiti Putra Malaysia, 43400 Serdang, Malaysia, +60386567121
*Corresponding author, e-mail: nfmr86@gmail.com

Abstract

An electronic nose was used to distinguish between selected herb samples according to their family group species. This paper aims to evaluate the potential of using the electronic nose to characterize three groups of families of twelve herb species based on the discriminant analysis approach. The feature extraction involves the use of a signal processing technique that simplifies classification and yields optimal results. Two discriminant techniques: the principal component analysis (PCA) and the multiple discriminant analysis (MDA) were used to investigate the potential to distinguish herb species between several herbs within the same family group. The results showed that the twelve herb species can be better classified using the MDA method compared to the PCA method.

Keywords*: principal component analysis, multiple discriminant analysis, feature extraction*

1. Introduction

Phytochemicals in the form of volatile compounds give each herb different characteristic odors. Terpenes, steroids, phenolic compounds, amino acids, lipids, and alkaloids are common herbal constituents in herbs [1]. Herbs identification researches based on leaves characteristics have been proposed in various techniques and systems. This is because leaves have unique characteristics, such as shape, color, texture, and the nature of the odors [2-5]. However, dealing with herbs in the same group of family is more challenging than with different family groups since their physical appearances and aroma may be similar. Traditionally, botanists and forest rangers are the ones who are assigned to recognize, identify, and characterize the plant species. Unfortunately, many of the existing plant species on earth are still unknown due to the limited number of resources, and experts. Additionally, the human sensory panels are subjective and also inaccurate [2, 3].

Persaud, el al., [6] invented the mimicking nose system in 1982. The electronic nose (E-nose) is an analytical instrument, which consists of an array of chemical gas sensors with different gas selectivity, together with a pattern recognition system that can differentiate between odors and flavors [2, 3]. Currently, various techniques for automatic herbs recognition were proposed in pattern classification algorithms. E-nose for herbs identification relies on the selection of gas sensors, and the ability of the system to analyze the odor detection signal as well as to process the extracted information for discriminant analysis before the implementation of species classification. The selection of gas sensors depend on the volatile compounds that exist in the herbs. Gas chromatography mass spectrometer (GCMS) is used in many sectors to identify volatile compounds [7]. However, GCMS is a very complex experiment, and requires a huge budget for herbs odor analysis.

Some of the common challenges during the data collection process in signal processing are noise, duplication, and high dimensionality of data. The size of a dataset must be reduced during the pre-processing stage for further analyses, such as for classification, in order to obtain optimum results. In statistical pattern recognition, discriminant analysis is a learned discriminative feature transformation technique [8]. Discriminant analysis can reduce the dimension of the dataset, and it is one of the feature extraction processes [9, 10]. In this study, we will focus on two methods of data reduction: 1) principal component analysis (PCA), and 2) multiple discriminant analysis (MDA). Basically, PCA and MDA are two linear transformations

Received November 22, 2016; Revised February 5, 2017; Accepted February 27, 2017

that are involved in projecting their features from a higher to a lower dimensional space. The difference between these methods is that the PCA finds the directions that can maximize the variance, while the MDA finds the directions that can maximize the separation between classes. Furthermore, PCA is an unsupervised learning method, while MDA is a supervised learning method. This paper investigates the potential of using these types of discriminant analyses to discriminate three family groups of twelve aromatic herb species, which were collected from the UPM botanical garden. The electronic nose system was used for collecting the data of the herbs' odor responsesere evaluated to determine a linear transformation that are best discriminate among the classes.

2. Research Method
2.1. Overview of E-Nose Experiment

The development of the electronic nose was targeted to classify twelve herb species from three aromatic herb families, which were the Lauraceae, Myrtaceae, and Zingiberaceae family have been performed by Mohamad Yusof et al. (2015) [11]. The list of herbs was chosen with consultations from the botanists at the Bioscience Institute, Universiti Putra Malaysia based on the availability of the samples. Each sample was collected from the Agricultural Conservatory Park, Universiti Putra Malaysia. The scientific names of the samples are listed in Table 1. The selection of gas sensors for the electronic nose was conducted by Mohamad Yusof et al. (2014) [11], as listed in Table 2. FIGARO gas sensors were selected due to their fast response towards a broad range of chemical compounds, in addition to being affordable, with low power consumption, and a large number of target gas detection.

Table 1. Scientific names of the twelve herb species

Names of Herb Species		
Family *Lauraceae*	**Family *Myrtaceae***	**Family *Zingiberaceae***
Cinnamomum Iners	Syzygium Aromaticum	Scaphoclamys Kunstleri
Cinnamomum Verum	Syzygium Polyanthum	Etlingera Terengganuensis
Cinnamomum Porrectum	Melaleuca Alternifolia	Zingiber Zerumbet
Litsea Elliptica	Rhodomyrtus Tomentosa	Elettariopsis Curtisii

Table 2. The selected FIGARO MOS gas sensors for the electronic nose

Sensor Type	Abbreviation	Type of gas detection
TGS 2610	Sensor 1	Butane, propane, liquefied petroleum gas
TGS 2611	Sensor 2	Methane, natural gas
TGS 2620	Sensor 3	Alcohol, toluene, xylene, volatile organic compound
TGS 823	Sensor 4	Organic solvent vapors
TGS 832	Sensor 5	Halocarbon, Chlorofluorocarbon

2.2. Principal Component Analysis

Discriminant analysis is a powerful tool for analyzing the pattern of data discriminate in graphical representations. Various methods are available for discriminant analysis, such as the principal component analysis (PCA), linear discriminant analysis (LDA), multiple discriminant analysis (MDA), and so forth. Computationally, a discriminant analysis investigates the best discriminate data by transforming the high dimensionality of the data into a new dataset with low dimension, and without much loss of information. PCA was introduced by Pearson in 1901 [8]. It is an unsupervised multivariate data analysis technique that is used to visualize a pattern by projecting data to an eigenvector with largest eigenvalue that maximizes the variance of the projected data, as shown in Figure 1. The PCA is computed by determining the eigenvectors, and the eigenvalues of the covariance matrix. The covariance of two random variables is their tendency to vary together. Even when the PCA performs a coordinate rotation that aligns the transformed axes with the direction of the maximum variance, it does not guarantee a good discriminant between classes. Besides, it does not consider the class label of the feature vector because of its unsupervised dimension reduction [8]. The equation for PCA can be computed as:

$$y = \omega^T \cdot x \tag{1}$$

Where a set of *d*-dimensional data, $x = \{x^1, x^2, \ldots, x^N\}$, while $\omega^T = \omega_1 + \omega_2 + \cdots + \omega_1$ is an eigenvector, which gives the new projected data, y.

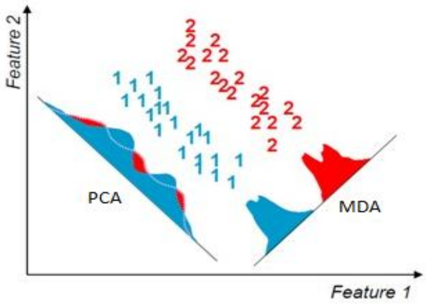


Figure 1. PCA versus MDA projection mapping

2.3. Multiple Discriminant Analysis

MDA is a supervised learning and discriminative technique for solving multiclass problems. This technique is inspired by Fisher's linear discriminant analysis for solving the discrimination between two-class problems. The MDA also project data in a similar way to the PCA but the projected data would be according to the best separated in the least-square sense. MDA will search for the optimal transformation whereby the projection should still maximize the ratio of between-class scatter to the within-class scatter, as shown in Figure 1. As a consequence, discrimination mapping between classes are well separated with less losses of information. The Fisher criterion function is defined in Equation (2);

$$J(W) = \frac{|W^T \cdot S_B \cdot W|}{|W^T \cdot S_W \cdot W|} \tag{2}$$
$$S_B = (\mu_1 - \mu_0) \cdot (\mu_1 - \mu_0)^T$$
$$S_W = (\mathrm{S}_1 + \mathrm{S}_2 + \cdots \mathrm{S}_{n\_})$$

Where S_B and S_W are the ratios of between-class scatter matrix and within-class scatter matrix, respectively. The scatter matrix of *n*-class are $S_1, S_2, \ldots, S_n$, while W is the projection matrix, and $J(W)$ is an eigenvalue that measures the difference between class means by measuring the within-class scatter matrix.

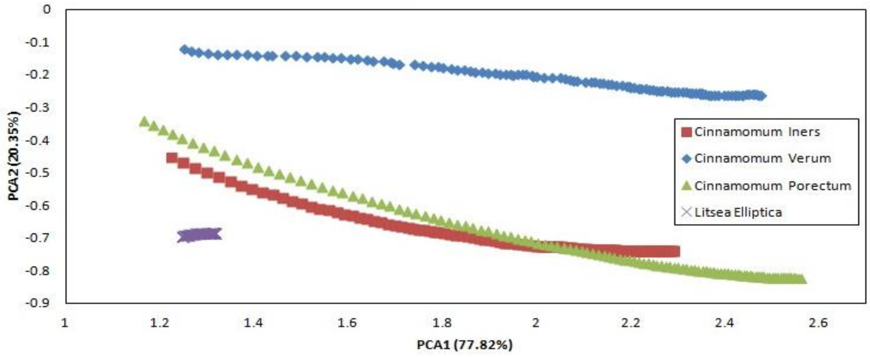
3. Results and Analysis

The discriminant analysis techniques in this research were implemented to discriminate multiclass problems by using the principal component analysis, and the multiple discriminant analysis. Twelve aromatic herb species from three group families were studied. The projected data results for the lower dimension space of four distinct clusters of herb species for each group family using PCA and MDA are shown in Figure 2 and Figure 3, respectively. Meanwhile, Table 3 lists the percentage results of the principal component (PCA_1 versus PCA_2) from the PCA method, and the percentage results of the multiple discriminant component from the MDA method (MDA_1 versus MDA_2).

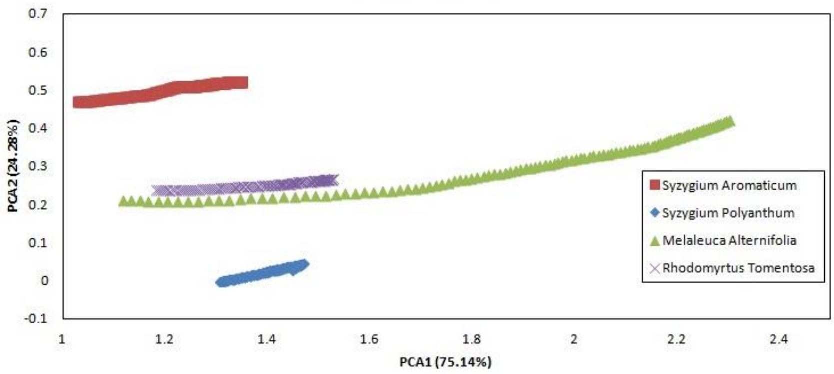
The percentage of component represents the percentage of the remaining data information after the transformation to a new coordinate. The first component (PCA_1 and MDA_1) contains the most information compared to the next number of component (PCA_2 and MDA_2). The first components for the MDA for all three family group species: *Lauraceae, Myrtaceae, and Zingiberacea* were 94.19%, 98.04%, and 70.99%, which were higher than the PCA percentages of 77.82%, 75.14%, and 62.40%, as shown in Table 3. These results showed that MDA has less of a total loss of information compared to PCA because the percentage of MDA_1 was greater than the percentage of PCA_1.

Table 3. The percentages of principal component analysis and multiple discriminant component

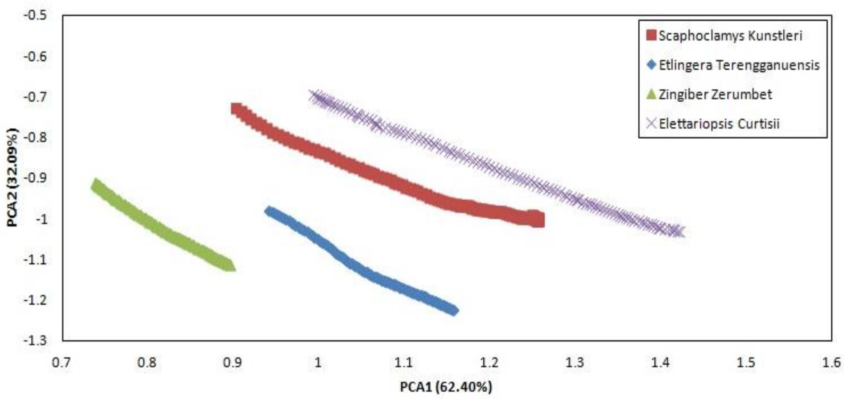
Family Name	PCA_1 (%)	PCA_2 (%)	MDA_1 (%)	MDA_2 (%)
Lauraceae	77.82	20.35	94.19	5.39
Myrtaceae	75.14	24.28	98.04	0.99
Zingiberaceae	62.40	32.09	70.99	27.56



(a) Family *Lauraceae*

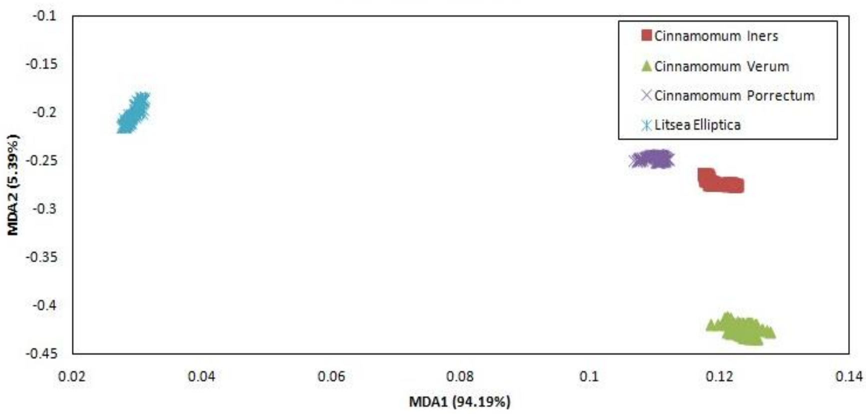


(b) Family *Myrtaceae*

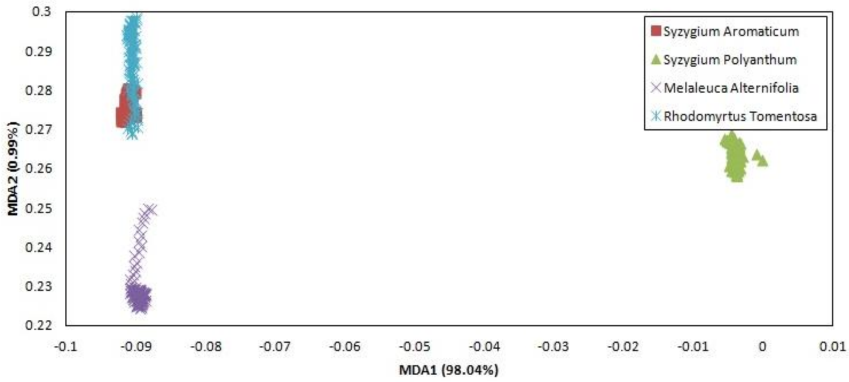


(c) Family *Zingiberaceae*

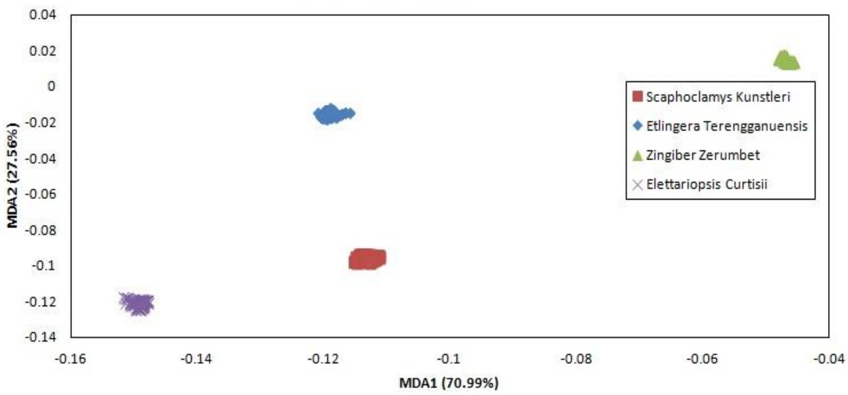
Figure 2. Discriminant analysis using Principal Component Analysis method, PCA_1 versus PCA_2



(a) Family *Lauraceae*



(b) Family *Myrtaceae*



(c) Family *Zingiberaceae*

Figure 3. Discriminant analysis using Multiple Discriminant Analysis method, MDA_1 versus MDA_2

MDA results in Figure 3 show better discriminations compared to results using PCA in Figure 2, whereby twelve herb species from three family groups were well separated, even though their physical appearances and aromatic characteristics are almost the same. MDA shows a greater percentage of information data on first projection vector compared to PCA. Therefore, the herb species within the same class were projected very close to each other, while simultaneously telling them farther apart from among as many different species or classes as possible. Better separation between classes will help in increasing the accuracy of the classification.

4. Conclusion

Both discriminant techniques represented a projection vector (dataset) onto a lesser dimensional space by extracting the most relevant information with less losses of information. This paper presents a comparison between two discriminant analysis methods when discriminating the classes of twelve herbs species among three group families. The results showed the advantages in using the MDA technique over the PCA. In general, discriminant analysis seeks to project data along a certain direction that can give better discrimination among several classes. Both discriminant techniques had performed the discrimination to visualize the classifications among several herb group species. MDA had shown a better separation between herbs species even for those in the same family when compared to the PCA technique using discriminant features.

Acknowledgements

This work is financed by the Universiti Putra Malaysia Grant Scheme (Geran Putra IPS) under the project title: Development of E-Tongue Device for Herb Recognition System, and a MyBrain PhD under the Ministry of Higher Education. Last but not least, an appreciation also goes to the Institute of Bioscience, Universiti Putra Malaysia for providing samples of the herbs.

References

[1] Ganora L. Herbal Constituents Foundations of Phytochemistry. Louisville, Colorado: Herbalchem Press. 2008: 1-15.
[2] Wilson AD. Diverse applications of electronic-nose technologies in agriculture and forestry. *Sensors (Basel)*. 2013; 13(2): 2295-2348.
[3] Wilson AD, Baietto M. Applications and advances in electronic-nose technologies. *Sensors (Basel)*. 2009; 9(7): 5099-5148.
[4] Ishak AJ, Hussain A, Mustafa MM. Weed image classification using Gabor wavelet and gradient field distribution. *Computers and Electronics in Agriculture*. 2009; 66(1): 53-61.
[5] Husin Z, Shakaff AYM, Aziz AHA, Farook RSM, Jaafar MN, Hashim U, Harun A. Embedded portable device for herb leaves recognition using image processing techniques and neural network algorithm. *Computers and Electronics in Agriculture*. 2012; 89: 18-29.
[6] Persaud KC, Dodd GH. Analysis of Discrimination Mechanisms of the Mammalian Olfactory System using A Model Nose. *Nature*. 1982; 299: 352-355.
[7] Jerome Jeyakumar J, Kamaraj M, Nandagopalan V, Anburaja V, Thiruvengadam. A study of Phytochemical Constituents in Caralluma Umbellata by Gc-Ms Analysis. *International Journal of Pharmaceutical Science Invention*. 2013: 37-41.
[8] Fukunaga K. Introduction to statistical pattern recognition. Academic Press Professional. 1990.
[9] Pearson K. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*. 1901; 6(2): 559-572.
[10] Geoffrey JM. Discriminant Analysis and Statistical Pattern Recognition. New York: John Wiley & Sons. 1992.
[11] Mohamad Yusof UK, Che Soh A, Radzi NFM, Ishak AJ, Hassan MK, Ahmad SA, Khamis S. Selection of Feature Analysis Electronic Nose Signals Based on the Correlation Between Gas Sensor and Herbal Phytochemical. *Australian Journal of Basic and Applied Sciences*. 2015; 9(5): 360-367.