EasyChair Preprint

№ 7272

Cardamom Quality Evaluation Employing Electronic Nose

Madhurima Ghosh, Devdulal Ghosh and Nabarun Bhattacharyya

December 28, 2021

Cardamom Quality Evaluation employing Electronic Nose

Madhurima Ghosh✉, Devdulal Ghosh, Nabarun Bhattacharyya

Abstract Cardamom is an export-oriented commodity of India, used for cooking and has medicinal values. This calls for the need of rapid and effective method for quality assessment of cardamom. The major parameters for quality determination of cardamom are flavour and aroma. The present practice of quality estimation involves assessment by human experts based on physical characteristics, such as freshness, shape, size, colour and aroma which is subject to biasness and error. The alternate evaluation technique comprises of various qualitative and quantitative chemical tests of the samples (using gas chromatography and mass spectrometer), which is an expensive, labourious, time-consuming and skilled man-power demanding process. This paper presents a novel approach to predict the oil yield and constituent chemicals 1,8-cineole and alpha-terpinyl acetate, which are responsible for the flavour of cardamom using a handheld electronic-nose (HEN) developed by Centre for Development of Advanced Computing (CDAC), Kolkata. We worked with thirteen different cardamom samples with varying oil yield percentages and different measures of the constituent chemicals of the extracted essential oil. We considered cardamom samples, with and without husk separately in our study. Initially, we have applied clustering algorithms like Principal Component Analysis (PCA) and Density-Based Spatial Clustering on Applications with Noise (DB-SCAN) on the data values collected by five different metal oxide semiconductor (MOS) sensors of HEN and the device was able to group the samples into distinct clusters with considerable accuracy. Then, Partial Least Square (PLS) Regression was applied on the dataset to train the system and eventually predict the quality of unknown cardamom samples. The model has given around 98% and 95% accuracy for oil yield and 1,8-cineole prediction respectively for the samples with husk.

Keywords Cardamom · Quality evaluation · Partial Least Square Regression · Principal Component Analysis · Density-Based Spatial Clustering of Applications with Noise · Electronic Nose · Metal Oxide Semiconductor Sensor

1 Introduction

Cardamom is ranked third in terms of price, followed by saffron and vanilla. Small cardamom (Elettaria cardamomum Maton) is grown mainly in India, Guatemala, Sri Lanka, Nepal, and also found in Tanzania, Indonesia, Vietnam, Thailand, Papua New Guinea, and El Salvador, and is exported worldwide specially to Middle Eastern countries. Cardamom extracts and cardamom essential oil are used as a flavouring agent and has numerous potential therapeutic uses.

Hence, rapid and effective method for quality evaluation of cardamom is the need of the hour. The quality varies with the percentage of essential oil extracted and its constituent chemicals. The major constituents of essential oil are 1,8-cineole and alpha-terpinyl acetate. The most popular practice of aroma estimation of cardamom, and thereby price determination, involves quality assessment by human experts based on physical characteristics, such as freshness, shape, size, colour and aroma. However, this method is subject to biasness of the expert, error prone and non-repetitive. In order to provide accurate estimation of quality, an alternative approach involves various chemical tests employing analytical instruments, such as gas chromatography (GC) and gas chromatography coupled with a mass spectrometer (GC/MS) [1]. The results of this method confirm that

M. Ghosh (✉)
B. Tech Final Year Student, St. Thomas' College of Engineering & Technology (under Maulana Abul Kalam Azad University of Technology), Kolkata
e-mail: mghosh.madhurima@gmail.com

D. Ghosh, N. Bhattacharyya
Centre for Development of Advanced Computing, Kolkata
e-mail: devdulal.ghosh@cdac.in

N. Bhattacharyya
e-mail: nabarun.bhattacharya@cdac.in

the proportion of the various compounds of the oil extracted from cardamom differ from each other among different varieties of cardamom [2], [3]. However, this process is time-consuming, requires expensive laboratory equipment and skilled personnel to operate these instruments.

Electronic nose (e-Nose) [4] is one of the prospective solutions for rapid quality estimation of various food products and beverages, such as meat [5], tea [6], coffee, dairy products, alcoholic beverages, oil [7], vinegar, cocoa beans and many more. This method provides considerable level of accuracy, removes human subjectivities and errors, and is repetitive. In an e-Nose, an array of sensors is embedded in the device to sense and generate an "aroma fingerprint" for each training sample. Since, most of these sensors are specific to a class of chemicals, and have cross sensitivities to other classes too, it is very difficult to quantify the amount of the constituent chemicals present in the data samples. Instead, the approach employed in e-Noses is to create an "aroma fingerprint" based on the sensor values for each type of food sample. Once the training is complete, the device tries to match the aroma fingerprints of unknown samples under test to that of the pre-trained samples resulting in evaluation of the quality of the unknown samples. In some cases, e-Nose is also used for several disease detection, such as asthma [8], cancer, tuberculosis, cystic fibrosis [9] or even monitoring the health of astronauts in space. Although, quality estimation of spices using e-Nose is a relatively new domain of research, some preliminary work on sensor selection and aroma fingerprint generation of cardamom are done by Leela et al. [10] and Ghosh et al [11].

We have adopted a novel approach by using a handheld electronic nose (HEN) [12] developed by Centre for Development of Advanced Computing (CDAC), Kolkata for the quality evaluation of cardamom by estimating the oil yield percentage and measure of 1,8-cineole and alpha-terpinyl acetate in the samples through Partial Least Square (PLS) regression algorithm. In this study, we have dealt with thirteen different samples separately for cardamom with husk and without husk.

2 Materials and Method

2.1 Cardamom Extraction Process

ICAR-Indian Institute of Spices Research (IISR), Kozhikode contributed in this study through extraction of essential oil [10] and measurement of the percentage of major constituent chemicals, namely, 1,8-cineole and alpha-terpinyl acetate of thirteen different samples of cardamom. 20 g of each sample including the husk were grinded and then hydrodistilled for three hours in Clevenger apparatus to extract the essential oil. The extracted essential oil was collected, dried over anhydrous sodium sulphate and stored in a refrigerator until the analysis was carried out.

2.2 Chemical Analysis of Samples

Gas chromatography (GC) [11] examination was performed on a Shimadzu (GC-2010) Gas chromatograph fitted to FID indicator and RTX – 5 column (30 m x 0.25 mm, film thickness 0.25 µm). Nitrogen was utilized as the transporter gas at a flow rate of 1.0 ml/min and the oven was programmed as at 60ºC up to 200ºC at the rate of 3.0ºC/min, again up to 220ºC at the rate of 5.0ºC/min, at which the column was kept up for 5 min. 0.1µl sample was infused. The injection port temperature was kept up at 250ºC and the detector temperature was 250ºC. 1:40 was the split ratio. The percentage composition component of the oil was dictated by area normalization.

Afterward, the analysis of essential oil was done using a Shimadzu GC-2010 Gas chromatograph furnished with QP 2010 mass spectrometer (GC/MS). RTX – 5 column (30 m x 0.25 mm, film thickness 0.25 µm) covered with polyethylene glycol was utilized. Helium was utilized as the transporter gas at a flow rate of 1.0 ml/minute. The stove was modified at 60ºC for 5 min and afterward expanded to 110ºC at 5.0ºC/min, then, at that point, up to 200ºC at the pace of 3.0ºC/min, again up to 220ºC at 5.0ºC/min, at which the column was kept up for 5 min. 0.1µl sample was infused. The injection port temperature was kept up at 260ºC and the detector temperature was 250ºC. 1:40 was the split ratio and ionization voltage was kept up at 70 eV. The retention indices were determined comparative with C8 – C20 n-alkanes. Table 1 shows the chemical analysis report for thirteen different samples carried out by IISR.

Table 1. Oil yield and major constituent chemical percentages of different cardamom samples obtained through chemical analysis

Sample	Name	Oil Yield (%)	1,8-Cineole (%)	Alpha-terpinyl acetate (%)
1	ELA	7	21	44
2	THIRU	5.5	22	44
3	GG1	7	18	48
4	GG2	6	22	42
5	WONDE	6.5	26	46
6	CON1	4	25	48
7	CON2	4.5	30	42
8	CON3	4	27	47
9	CON4	4.5	28	44
10	CON5	4	30	42
11	CON6	5	28	46
12	CON7	4	30	43
13	CON8	4.5	27	45

Chemical analysis is done by IISR, Kozhikode

2.3 Architecture of Handheld Electronic Nose (HEN)

This study has been conducted employing a Handheld e-Nose (HEN) [13] developed and patented by C-DAC, Kolkata. HEN is a true handheld e-Nose, built on embedded electronics and is suitable for field operation. The device is battery operated with integrated odour handling unit. The sample holder can easily be fitted onto the device during testing of samples. A touchscreen graphics display is provided as user interface and displaying of results. Combination of a micro-pump and micro valve serves the sample odour VOC excitation and thereafter delivery to the sensor array. The sensor array is built on five metal oxide semiconductor (MOS) sensors [14], namely, TGS 2602, TGS 2620, TGS 816, TGS 823 and TGS 832. Captured sensor data and results are stored in a SD Card in FAT32 file system.

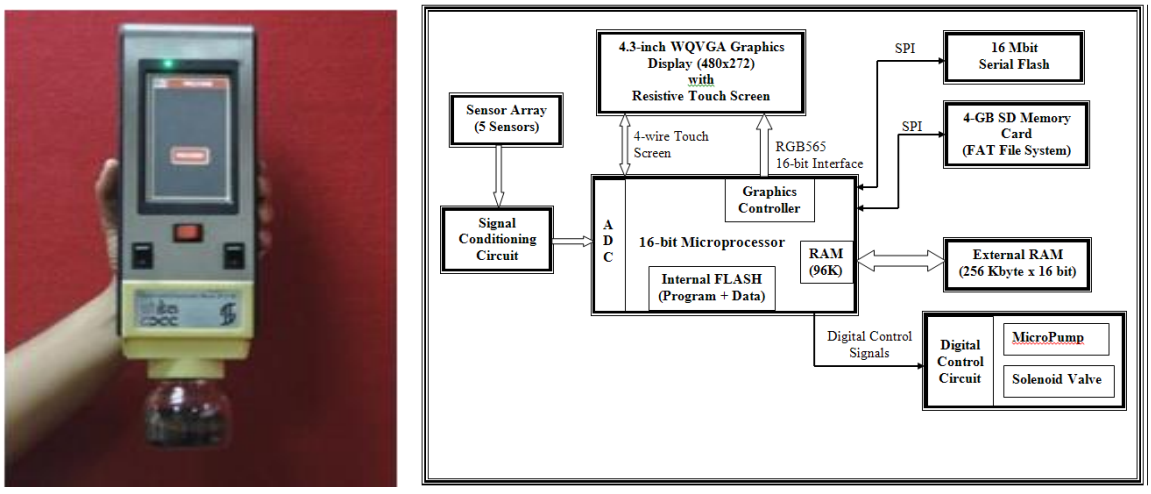

Fig. 1. Block diagram of Handheld Electronic Nose (HEN) architecture

Fig. 1 shows the hardware block diagram of HEN. A 16-bit PIC microprocessor is the heart and brain of HEN. After signal conditioning, the captured sensor data is fed to the in-built analog to digital converter of the processor for processing. The 4.3-inch touch screen and graphics display are interfaced to the processor through 4-wire and 16-bit RGB565 lines respectively. SD memory card and serial Flash memory are interfaced through SPI lines. Control of micro pump and micro valve is done through digital I/O lines of the processor. Software for HEN is developed on embedded C and graphics portion using Microchip graphics library.

2.4 Data Collection

For three consecutive days, the samples were tested on HEN and the corresponding values of the five different sensors were stored. 10 g of each sample with husk was first placed in the sample holder of HEN and the sensor data was recorded. Then the husk from these samples were removed, their corresponding weights were measured and the earlier stated process was repeated. The temperature and humidity recorded on 6th and 7th May were around 30^{o} C and 57% respectively. However, on 8th May, the temperature dropped down to 22^{o} C with heavy rainfall and the humidity increased to about 97%. The responses of MOS sensors are prone to variations in ambient temperature and humidity and thus, the data taken on 8th May had shown abrupt changes as compared to that of 6th and 7th May. Hence, we discarded the records taken on 8th May for training the device and analysis of results. Table 2 displays the details of the data taken for ana**lysis.**

Table 2. Details about the recorded dataset

Date	Sample	Name	Type	Weight (g)
6th May	1	ELA	With Husk	10
			Without husk	7.3
	2	THIRU	With husk	10
			Without husk	7.8
	3	GG1	With husk	10
			With husk	10
			Without husk	7.4
	4	GG2	With husk	10
			Without husk	7.4
			Without husk	7.4
	5	WONDE	With husk	10
			Without husk	7.5
			Without husk	7.5
			Without husk	7.5
	6	CON1	With husk	10
			Without husk	6.3
	7	CON2	With husk	10
			Without husk	7
	8	CON3	With husk	10
			Without husk	6.7
	9	CON4	With husk	10
			Without husk	6.3
7th May	10	CON5	With husk	10
			Without husk	6.5
	11	CON6	With husk	10
			Without husk	7
	12	CON7	With husk	10
			Without husk	6.9
	13	CON8	With husk	10
			Without husk	6.3

2.5 Analysis of Sensor Data

Fig. 3 shows the graph of 600 values recorded from the time of opening the valves and until the saturation point of all the five different sensors are reached, for the sample, CON7 (Sample 12) with husk recorded on 6th May. The graphs of all the other samples are similar to Fig. 2, but with variable slopes and different saturation points. However, we have considered the last 20 values of each sample for analysis where the readings are almost stable.

The sensor readings are then correlated to group the samples into distinct clusters and for prediction of the oil yield and major constituent chemical percentages responsible for the flavour of cardamom.

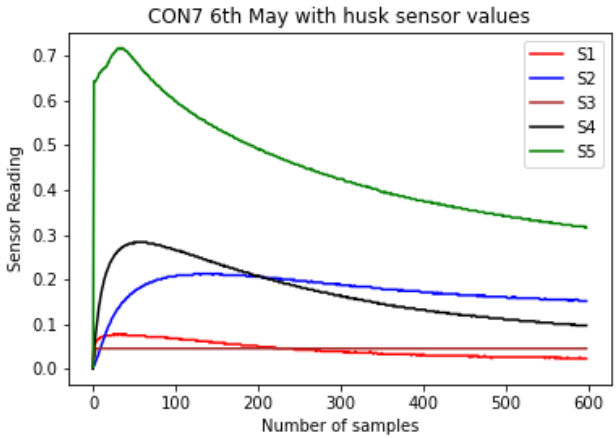

Fig. 2. Five sensor readings of HEN for CON7 (Sample 12)

2.6 Clustering Algorithms

Clustering refers to division of data points into various groups of similar data or values. Initially, the data for all the samples taken on both the days were fed to two popular clustering algorithms, namely Principal Component Analysis (PCA) and Density-Based Spatial Clustering of Applications with Noise (DB-SCAN). The system is able to segregate the samples into different clusters with considerable accuracy. This indicates that the oil yield percentage and composition of constituent chemicals of essential oil varies significantly and the data is fit for classification of samples into different categories.

Principal Component Analysis (PCA)

The dataset generated from HEN is difficult to interpret and visualize as it has dimensionality of five (five parameters). Hence, the data is visualized in 2D using Principal Component Analysis (PCA) [15] with minimum data loss to check whether the samples could be clustered into distinct groups. The algorithm of PCA is given in Fig. 3.

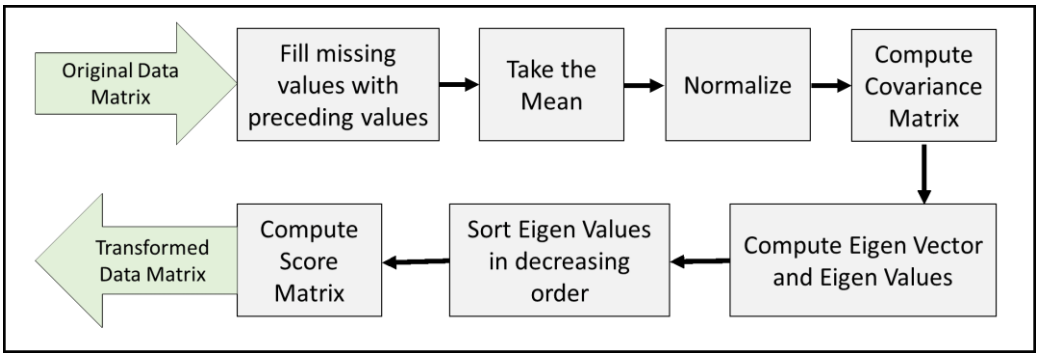

Fig. 3. Algorithm for Principal Component Analysis

Density-Based Spatial Clustering of Applications with Noise (DB-SCAN)

DB-SCAN (Density-Based Spatial Clustering of Applications with Noise) [16] is able to find arbitrary shaped clusters and also discards noise as outliers. Moreover, it does not need to provide the total number of clusters the dataset needs to be grouped to and hence, even in absence of domain knowledge, it would provide good clustering

results. However, the distance that specifies neighbourhoods and the minimum number of data points in each cluster much be specified for the algorithm to execute. The algorithm for DB-SCAN is given in Fig. 4.

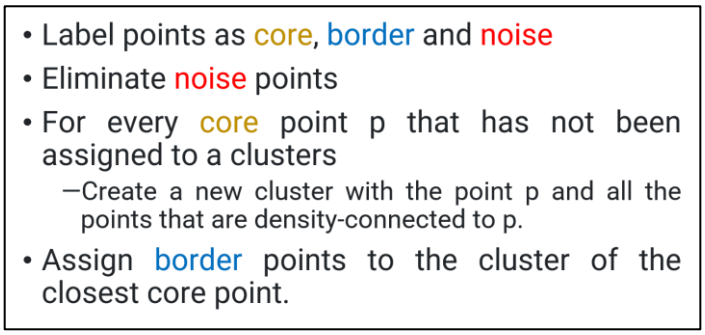

Fig. 4. Algorithm for DB-SCAN

2.7 Regression Algorithm

Regression models are used to establish relationship between the independent variables, i.e., the sensor values and the dependent variables, i.e., the oil yield, and measures of 1,8-cineole and alpha-terpinyl acetate in this use case of the training dataset. The established relationship is then utilized to predict the target(s), i.e., the oil yield, 1,8-cineole and alpha-terpinyl acetate percentages for unknown samples. We have used Partial Least Square (PLS) Regression algorithm for analyzing the data.

Partial Least Squares Regression (PLS)

Partial Least Square (PLS) [17] is an optimal regression algorithm based on covariance. PLS applies dimension reduction of the samples (similar to PCA) before applying a linear regression algorithm on the transformed data. In initial study, it was observed that the data is highly correlated and hence, PLS is chosen for prediction of the target(s). It also gives accurate results in spite of noisy data. Fig. 5 depicts the algorithm for PLS Regression.

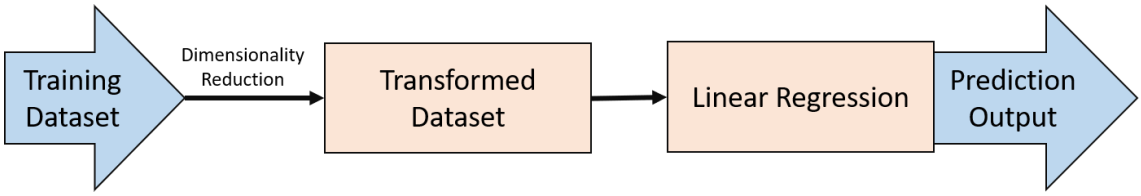

Fig. 5. Algorithm for Partial Least Square Regression

3 Result and Analysis

Table 3 depicts the amount of data explained by Principal components 1 and 2 for different datasets in order to reduce the dimension using PCA. It is observed that there is significant amount of data loss (approximately greater than 10%) when the number of samples taken under consideration is more than 5. However, this amount of data loss did not affect in proper clustering of the data.

Table 3. Percentage of variance explained by Principal Components

Date	Type	PC1 (%)	PC2 (%)	Data Loss (%)
6th May	With husk	60.5	29.1	10.4
	Without husk	61.2	20.9	17.9
7th May	With husk	79.4	18.9	1.7
	Without husk	66.6	32.5	0.9

Fig. 6 and 7 show the results of clustering using PCA on the samples on 6th and 7th May separately. The algorithm is able to properly cluster the data with considerable accuracy except certain imprecision in the samples CON3 (Sample 8) and CON4 (Sample 9) with husk on 6th May.

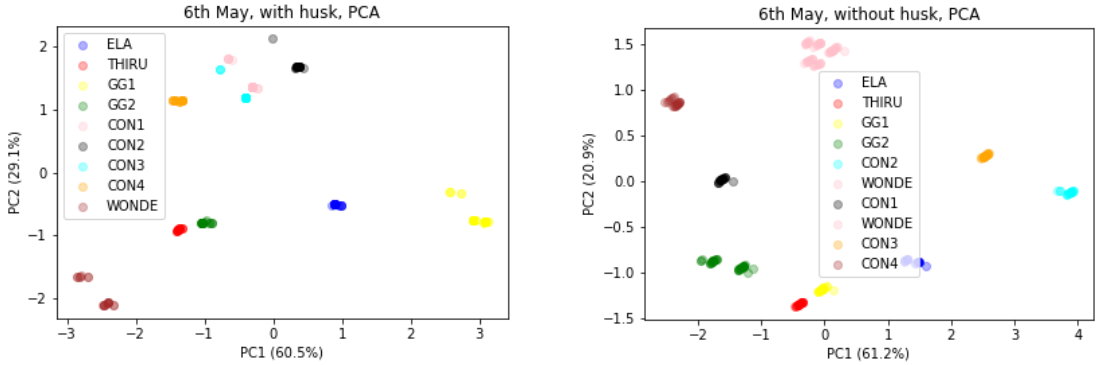

Fig. 6. Clustering using PCA for 6th May with and without husk

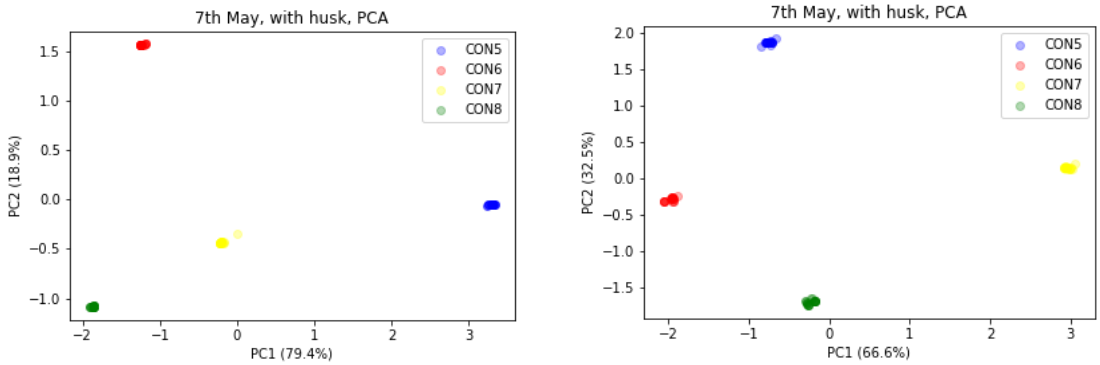

Fig. 7. Clustering using PCA for 7th May with and without husk

Fig. 11, 12, 13 and 14 compare the sample number and cluster number (result of DB-SCAN clustering) for the data recorded on 6th May and 7th May. Sample 3 (GG1), 4 (GG2) and 5 (WONDE) were recorded multiple times for experimental purpose. DB-SCAN algorithm could accurately cluster the datasets for with and without husk cardamom samples, except the dataset of with husk on 6th May i.e., the algorithm misplaced Sample 8 (CON3) into Sample 6 (CON1) cluster.

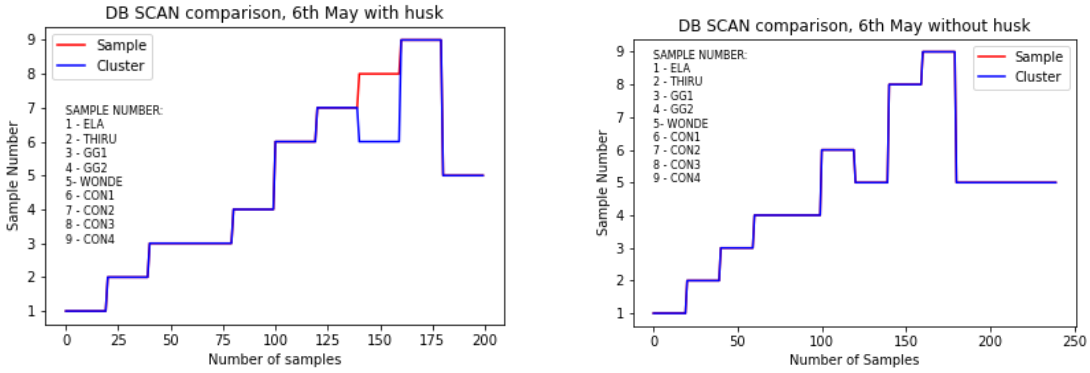

Fig. 8. Comparison of Cluster Number (result of DB-SCAN) with Sample Number on 6th May with and without husk[2]

[2] Sample 3 (GG1), 4 (GG2) and 5 (WONDE) were recorded multiple times for experimental purpose

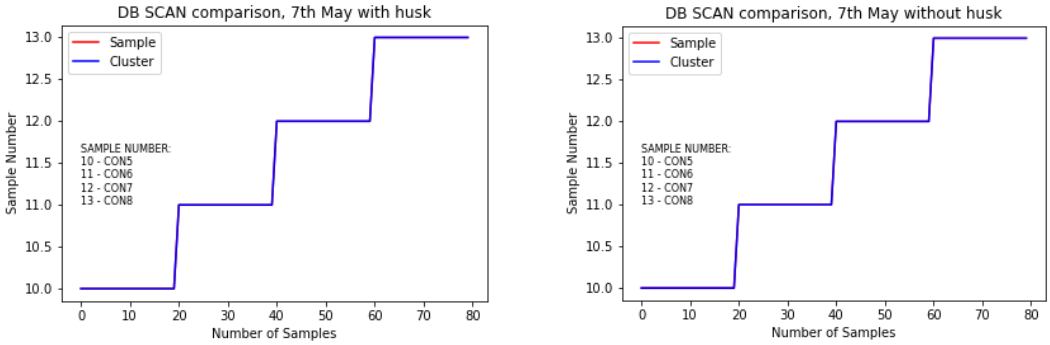

Fig. 9. Comparison of Cluster Number (result of DB-SCAN) with Sample Number on 7th May

The device was trained using PLS Regression model and Fig. 10, 11, 12 and 13 show the predictions of oil yield and 1,8-cineole percentages respectively by the model for unknown samples.

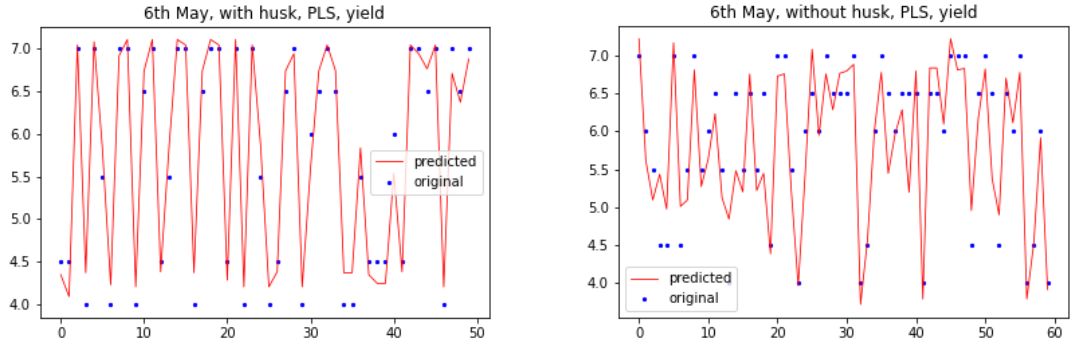

Fig. 10. Prediction of oil yield on 6th May with husk and without husk

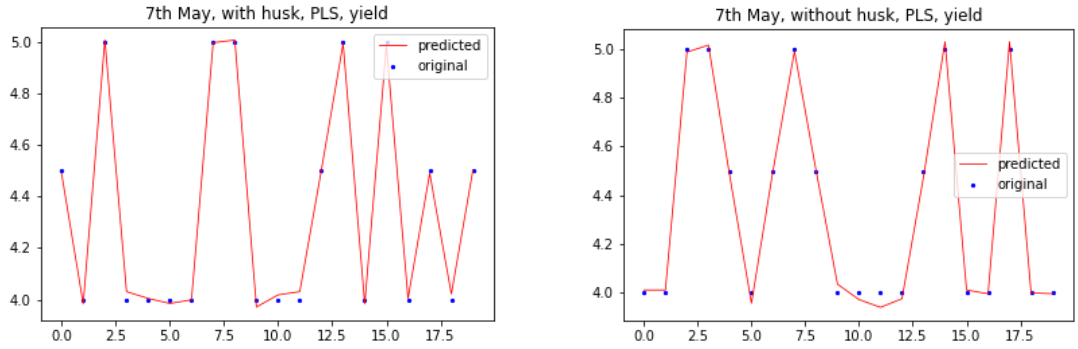

Fig. 11. Prediction of oil yield acetate on 7th May with husk and without husk

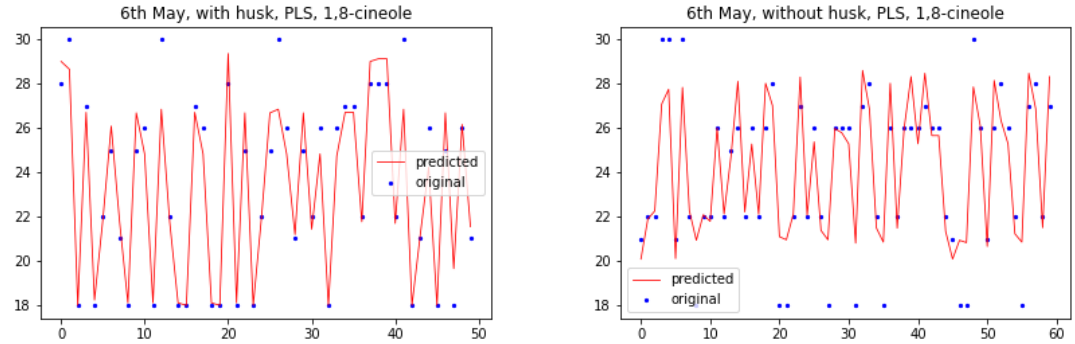

Fig. 12. Prediction of 1,8-cineole on 6th May with husk and without husk

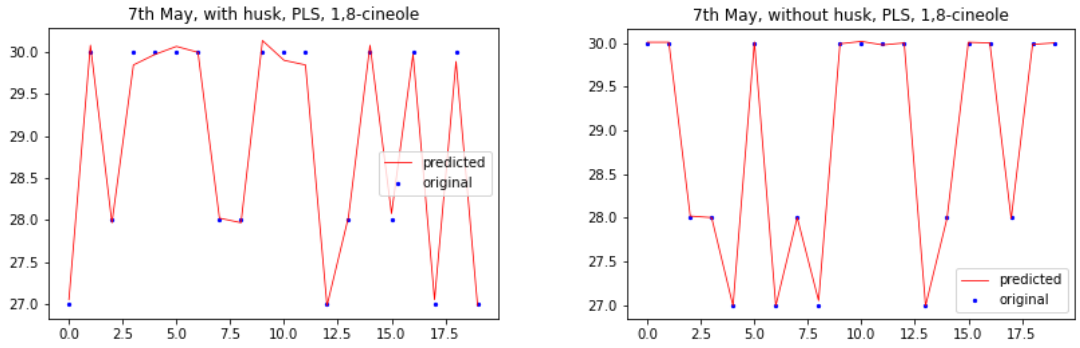

Fig. 13. Prediction of 1,8-cineole on 7th May with husk and without husk

Table 4 shows the results of PLS regression for estimation (average of 6th and 7th May datasets) of oil yield, 1,8-cineole and alpha-terpinyl acetate percentages in cardamom samples.

Table 4. Estimation of oil yield, 1,8-cineole and alpha-terpinyl acetate in cardamom samples by PLS Regression

Prediction	Type	Mean Squared Error (MSE)	Accuracy (%)
Oil Yield (%)	With husk	0.026	98.2
	Without husk	0.104	89.9
1,8-cineole (%)	With husk	0.688	95.7
	Without husk	1.228	90.2
Alpha-terpinyl acetate (%)	With husk	1.426	69.2
	Without husk	2.494	44.2

All results are average of datasets recorded on 6th and 7th May

Overall, the samples with husk gave better accuracy and precision as compared to the samples without husk. This confirms the fact that the husk of cardamom contains certain amounts of aroma constituents too.

The oil yield percentage prediction gave around 98% accuracy while 1,8-cineole predictions are almost 95% accurate for samples with husk. However, the result for estimation of alpha-terpinyl acetate is not at all satisfactory.

4 Conclusion

Our novel approach of cardamom quality estimation using HEN gave satisfactory results for oil yield and one of the major constituent chemicals of essential oil, 1,8-cineole and is fit for use in rapid evaluation of cardamom quality at affordable cost. However, the estimate for another constituent chemical of essential oil, alpha-terpinyl acetate is not reasonable.

The MOS sensors used for HEN in this study were specifically selected for quality analysis of tea and were not tuned for cardamom. Sensor array tuned for cardamom aroma constituent chemicals is expected to give more accurate results. A hybrid array of sensors including MOS, Quartz Crystal Microbalance (QCM), Conducting polymer, Electrochemical, etc., which may give optimum results for cardamom is a topic of our further research.

Moreover, the increased size of training dataset might also yield better precision and accuracy. Several other discriminant analysis algorithms and neural network-based regression models can be employed on the datasets to choose the optimum algorithm for our use case. Since, the MOS sensors are prone to temperature and humidity change, the parameters of ambient temperature, humidity and other environmental factors can be added in the feature list of the datasets to generalize the model and avoid training the system multiple times for various environmental conditions. The potential of e-Nose in evaluation of quality of cardamom extract and essential oil, as found in this study, can be explored further to make it suitable for commercial use.

Acknowledgments

The authors are thankful to the scientists of ICAR--Indian Institute of Spices Research (IISR), Kozhikode for providing labeled samples and chemical analysis of the samples.

References

1. Tambe, Esha & Gotmare, Sulekha & Scholar, (2019). Chemical Characterization of Three Cardamom Oils (Elettaria Cardamomum) By Gcms. World Journal of Pharmacy and Pharmaceutical Sciences. 08. 876 - 885. 10.20959/wjpps201911-14879.
2. Ashokkumar, K., Murugan, M., Dhanya, M., Raj, S., and Kamaraj, D. (2019). Phytochemical variations among four distinct varieties of Indian cardamom Elettaria cardamomum (L.) Maton. Nat. Prod. Res. 34, 1919–1922. doi: 10.1080/14786419.2018.1561687
3. B. Chempakam & Sindhu S (2008) Cardamom, pp 41-58, In: Chemistry of spices (Eds. V A Parthasarathy, B. Chempakam, T. John Zachariah), CABI.com
4. Karakaya, D., Ulucan, O. & Turkan, M. Electronic Nose and Its Applications: A Survey. Int. J. Autom. Comput. 17, 179–209 (2020). https://doi.org/10.1007/s11633-019-1212-9
5. J. S. Vestergaard, M. Martens, 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, no.6, pp.1095–1101, 2007. DOI: 10.1016/j.lwt.2006.06.008.
6. Amritasu Das, Alokesh Ghosh, Tarun Kanti Ghosh and Hena Ray “Portable Electronic Nose for Tea Aroma Detection”, International Journal of Compuying and Applications 6(1), 2011, pp. 39-43.
7. T. Majchrzak, W. Wojnowski, T. Dymerski, J. Gebicki, J. Namiesnik. Electronic noses in classification and quality control of edible oils: A review. Food Chemistry, vol.246, pp.192–201, 2018. DOI: 10.1016/j.foodchem.2017.11.013.
8. S. Dragonieri, R. Schot, B. J. Mertens, S. Le Cessie, S. A. Gauw, A. Spanevello, O. Resta, N. P. Willard, T. J. Vink, K. F. Rabe, E. H. Bel, P. J. Sterk. An electronic nose in the discrimination of patients with asthma and controls. Journal of Allergy and Clinical Immunology, vol.120, no.4, pp.856–862, 2007. DOI: 10.1016/j.jaci.2007.05.043.
9. C. M. H. H. T. Robroeks, J. J. B. N. van Berkel, J. W. Dallinga, Q. Jobsis, L. J. I. Zimmermann, H. J. E. Hendriks, M. F. M. Wouters, C. P. M. van der Grinten, K. D. G. van de Kant, F. J. van Schooten, E. Dompeling. Metabolomics of volatile organic compounds in cystic fibrosis patients and controls. Pediatric Research, vol.68, no.1, pp.75–80, 2010. DOI: 10.1203/00006450- 201011001-00143.
10. N K Leela, A M Muneeb, S Mukherjee, D Ghosh, N Bhattacharya (2017). Essential oil content of cardamom (Elettaria cardamomum Maton) by Hand-held Electronic Nose. Journal of Spices and Aromatic Crops. 131. 10.25081/josac.2017.v26.i2.882.
11. D. Ghosh *et al*., "Exploratory study on aroma profile of cardamom by GC- MS and Electronic nose," *2012 Sixth International Conference on Sensing Technology (ICST)*, 2012, pp. 399-403, doi: 10.1109/ICSensT.2012.6461708.
12. Nabarun Bhattacharyya, Rajib Bandyopadhyay, Manbendra Bhuyan, Bipan Tudu, Devdulal Ghosh, and Arun Jana, Electronic nose for Black Tea Classification and correlation of measurement with “Tea Taster” Marks, in IEEE Transactions on Instrumentation and Measurement, Vol. 57, No. 7, 2008.
13. Ghosh A., Ray H., Ghosh T., Das A., Bhattacharyya N., 2014, Generic handheld e-nose platform for quality assessment of agricultural produces and biomedical applications, Chemical Engineering Transactions, 40, 259-264 DOI: 10.3303/CET1440044
14. Bhattacharya Nabarun, Tudu Bipan, Jana Arun, Ghosh Devdulal, Bandhopadhyaya Rajib, Bhuyan Manabendra., 2008. Preemptive Identification of Optimum Fermentation Time for Black Tea Using Electronic Nose. Sensors and Actuators B, 131 (1): 110-116.
15. Sehgal, S., Singh, H., Agarwal, M., Bhasker, V., & Shantanu (2014). Data analysis using principal component analysis. 2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom), 45-48.
16. Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD'96). AAAI Press, 226–231.
17. M. Larbi, R. Trinchero, F. G. Canavero, P. Besnier and M. Swaminathan, "Analysis of Parameter Variability in Integrated Devices by Partial Least Squares Regression," 2020 IEEE 24th Workshop on Signal and Power Integrity (SPI), 2020, pp. 1-4, doi: 10.1109/SPI48784.2020.9218175.