In-situ cocoa beans quality grading by near-infrared-chemodyes systems

Article *in* Analytical Methods · September 2017

DOI: 10.1039/C7AY01751K

CITATIONS
19

READS
167

4 authors, including:

Felix Y.H. Kutsanedzie
Accra Technical University
40 PUBLICATIONS **1,057** CITATIONS

SEE PROFILE

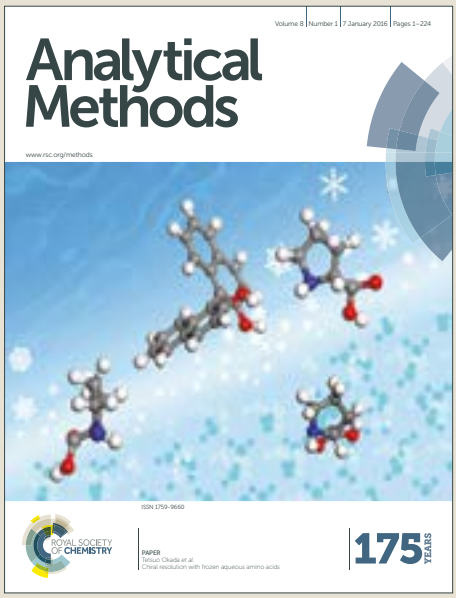
Quansheng Chen
Jiangsu University
421 PUBLICATIONS **11,150** CITATIONS

SEE PROFILE

Analytical Methods

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: F. Kutsanedzie, Q. Chen, H. Sun and W. Cheng, *Anal. Methods*, 2017, DOI: 10.1039/C7AY01751K.



ROYAL SOCIETY OF CHEMISTRY

rsc.li/methods

View Article Online
DOI: 10.1039/C7AY01751K

Analytical Methods

PAPER

In-situ cocoa beans quality grading by near-infrared-chemodyes systems

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Felix Y.H. Kutsanedzie[a], Quansheng Chen[a]*, Hao Sun[a], Wu Cheng[a]

Fermentation level is a key bean quality indicator in the cocoa industry. Colorimetric sensor e-nose (CS e-nose) and an innovatively designed near infrared chemo-intermediary-dyes spectra technique (NIR-CDS) combined with four chemometric algorithms – extreme machine learning (ELM), support vector machine (SVM), linear discriminant analysis (LDA) and k- nearest neighbor (k-NN), were applied to classify 90 sampled cocoa beans into three quality grades – fully fermented, partially fermented and non-fermented. CS e-nose (89% ≤ Rp ≤ 94%) and NIR-CDS (85% ≤ Rp ≤ 94%) achieved comparable classification rates; with the systems data cluster analysis yielding cophenetic correlation coefficients of (0.85 - 0.89). Both systems combined with SVM and ELM achieved high classification rate (Rp = 94%) and could be applied to cocoa bean quality classification on in-situ and nondestructive basis. This novel NIR-CDS technique proved a pragmatic approach for the selection of sensitive chemo-dyes used in the fabrication of e-nose colorimetric sensor array compared with the hitherto trial-and-error method, which is time-consuming and dye-wasteful. The technique could also be deployed in near-infrared systems for the detection of volatile (gaseous) compounds, which previously had been a limitation.

1. Introduction

Cocoa beans command a large market size globally and serve as raw material base for the chocolate, beverage and confectionery industries.[1]It is a commodity with enormous health benefits because of its rich polyphenol content, which has anti-oxidant, anti-carcinogenic, anti-inflammatory and anti-sclerotic properties; and thus capable of protecting the human body against cardiovascular diseases, cerebrovascular diseases, diabetes and rheumatoid arthritis.[2]

Despite there are several indicators of cocoa bean quality such as pH, total polyphenol content (TPC), methylxanthines content(MC), bean colour, moisture content, and flavour volatiles,[3, 4]the bean fermentation level has been the prime quality indicator used for cocoa bean quality classification, which is known to influence most of the afore-stated indicators.[3] As a result, well fermented cocoa beans are classified as high quality, with good chocolate flavour, and low astringent taste compared with unfermented beans.[4]

Different destructive techniques have been reportedly deployed for cocoa bean quality classification such as cut test, wet chemical analysis, electric tongue and GC-MS[5]; as well as non-destructive methods like near infrared spectroscopy (NIRS) and E-nose based on metal oxide sensors (MOS).[6, 7] Among the various quality classification techniques, the non-destructive ones are preferred because they have real time applicability, non-invasive, less labour intensive and environmentally friendly.[8] Excellent results have been reported on the use of commercial FT-NIRS systems for non-destructive total fungi count prediction in cocoa beans, cocoa bean authentication, classification and bean powder compositional analysis.[6, 8, 9] Albeit, the commercial FT-NIRS system is very accurate, it is expensive, has low accessibility and unable to detect food odour (gases). MOS based E-nose techniques have also been applied to cocoa bean quality classification non-destructively with their results affected by temperature and humidity shift.[10] Colorimetric sensor based e-

[a] School of Food & Biological Engineering, Jiangsu University, Xuefu Road 301, Zhenjiang 212013, P.R. China. * E-mail address: qschen@ujs.edu.cn (QS Chen). *Corresponding author: Tel.: +86-511-88790318; fax: +86-511-88780201

View Article Online
DOI: 10.1039/C7AY01751K

nose (CS e-nose), which provides antidote to the limitations of MOS based e-nose systems although has not yet been applied to cocoa beans quality classification, has its peculiar drawbacks of being labourious, time-consuming, and dyes-wasteful vis-à-vis its colorimetric sensor array (CSA) fabrication. NIR chemo-dyes spectra technique (NIR-CDS), an innovative technique designed in this study, employed near-infrared (NIR) to capture the spectra of volatile organic compounds present in cocoa beans exposed to chemo-dyes to attempt their quality classification.

Chemometric algorithms are usually combined with Non-destructive techniques for model establishment toward food quality monitoring, discrimination and chemical content prediction.[11] Different chemometric algorithms such as linear discrimination analysis (LDA), k nearest neighbour (k-NN), support vector machine (SVM), extreme learning machine (ELM) and artificial neural network (ANN) among others have been applied to the classification, identification and adulteration detection in various samples.[12] This study therefore combined two linear chemometric algorithms (LDA and k-NN) and two nonlinear ones (SVM and ELM) to CS e-nose and NIR-CDS (a novel technique), to attempt the non-destructive and real time quality classification of cocoa beans based on their fermentation levels and flavour volatiles present, as low-cost alternative techniques; as well as explore the NIR-CDS technique of combining chemo-dyes with NIRS system to provide antidote to chemo-dyes selection for CS e-nose sensor fabrication, and volatiles (gases) detection limitations in NIRS systems.

2. Experimental

2.1 Sample collection and preparation

Fresh cocoa beans pods obtained from cocoa growing regions in Ghana were cut open to remove the beans. Some of the beans were fermented for 7 days and others for 3 days by heaping and covering with banana leaves on the floor; and were solar-dried for a month to obtain fully fermented (FF) and partially fermented (PF) beans respectively. Similarly, portions of the fresh cocoa beans were solar-dried directly for a month to obtain the non-fermented (NF) beans. Thirty (30) beans from each of the three categories were ground into powder, cooled to a temperature of 20^{o}C and then sieved with 400mL mesh. The resulting bean powder from each category were subsequently weighed 3g per sample into nitrogen flushed bags and sealed to obtain a total of ninety (90) samples for the study.

2.2Systems description and data collection

CS e-nose system

The CS e-nose was built with a fabricated colorimetric sensor array composed of twelve (9) porphyrin dyes and three (3) pH indicators printed on reverse phase silica gel plates (Merck KGAA, Frankfurter, Germany) as sensors; a scanner (HP Scanjet G4050) connected to HP laptop for scanning of prior and post images of the fabricated sensor array exposed to the cocoa bean samples; and a laptop for data acquisition and processing as shown in Fig. 1.

[Here for Fig. 1]

The colorimetric sensor array for the system was fabricated using 2mg each of nine different metalloporphyrins (analytical grade) based dyes: 5, 10, 15, 20 – Tetraphenyl - 21H, 23H-porphine manganese (III) chloride; 5, 10, 15, 20 - Tetra phenyl-21H, 23H - porphine; 2, 3, 7, 8, 12, 13, 17, 18 - Octaethyl-21H, 23H - porphine manganese (III) chloride; 5, 10, 15, 20 – Tetrakis (4 – methoxyphenyl) - 21H, 23H - porphine iron (III) chloride; 5, 10, 15, 20 – Tetraphenyl - 21H, 23H - porphine; 5, 10, 15, 20 – Tetraphenyl - 21H, 23H - porphine copper (II); 5, 10, 15, 20 – Tetraphenyl - 21H, 23H - porphine zinc; 5, 10, 15, 20 - Tetra phenyl - 21H, 23H - porphine iron (III) chloride; 5, 10, 15, 20 - Tetrakis (4 – methoxyphenyl) - 21H, 23H - porphine cobalt(II); and three pH indicators - methyl red; bromocresol green; and bromothymol blue. The 9 dyes and 3 pH indicators were dissolved in 1mL dichloromethane and 1mL ethanol respectively to obtain 2 mg/mL of each dye separately in 2mL eppendorf tubes. These were ultrasonicated at room temperature for 30min to obtain 9 different porphyrin dyes and 3 pH indicator solutions. Each of these dyes and pH indicators were printed using 0.1µL micropipette in a 4 x 3 array arrangement on the reverse face of silica gel plates cut into the dimension of 3cm x 3cm. These printed silica gel plates were prepared, dried in a fume chamber, and then stored in nitrogen flushed bags for further used. For each of the 30 samples prepared per each cocoa bean category, a scanner (HP Scanjet G4050) connected to HP laptop was used to capture the prior images (I_1) and post exposure images (I_2) of 3g of each sample sealed in 20 mL glass beakers with clingfilm for 30min at room temperature. An illustration of the setup is captured in Fig. 1.

NIR-CDS system

NIR-CDS is an innovatively built system composed of a portable in-house developed NIRS system consisting of halogen light source with excitation wavelength range of 900nm – 2500nm, acquisition time of 3ms - 10s and a power source of 5V (SPL photonics Co., Hangzhou, china); spectrometer and optical fibre optic probe (ocean optics, Taiwan, China) connected to HP windows 7 laptop as illustrated in Fig. 2. Prior to the spectra data acquisition, the portable NIRS component of the system had its dark and white spectrum background corrected by subtraction, and its parameters set to: an exposure integration time, 4ms; number of scanning times, 3; and smoothing points, 5; to enable stable spectra data collection. The optical probe was used to scan each of the twelve (12) chemo-dye points on the reverse face of the silica gel plates of the colorimetric sensor array fabricated for the CS e-nose system, post their exposure to 3g each of the 90 cocoa beans samples in 20mL beakers sealed with clingfilm for 30mins at room temperature. The 12 different printed dye points per plate were scanned individually three times with near infrared

View Article Online
DOI: 10.1039/C7AY01751K

fibre optic probe and averaged to obtain the spectrum data of each sample's volatile organic compounds that bonded to the chemo-dyes.

[Here for Fig.2]

2.3 Data pre-processing

The spectral data obtained from cocoa beans samples using the NIR-CDS were pre-processed with standard normal variate (SNV) and multiplicative scatter correction (MSC) algorithms to remove irrelevant information contained in the acquired spectra capable of influencing the models prediction ability[13]. Also, the image data obtained via the CS e-nose were corrected with Pro 9 ACDsee image processing software.

2.4Data classification trend analysis

Principal components analysis (PCA) was performed on the data acquired with both systems to determine the classification trends of the data generated by compressing them into few variables - principal components (PCs) that retained the information contained in the original spectra[1]. The loadings plot of the top PCs obtained for the spectra data collected via NIRS – CDS system was analysed and used to explain the spectral discrimination of cocoa beans in the different categories. In addition, cluster analysis (CA) was used to generate dendrograms with cophenetic correlation coefficients to give an indication of how the classification patterns reflect the data acquired with both systems.

2.5 Chemometric calibrations

The pre-processed data obtained with the systems were processed using two nonlinear chemometric algorithms (ELM, SVM)[14, 15] and two linear ones (k-NN and LDA)[16, 17] to build models for the classification of the cocoa bean samples.

Extreme Learning Machine (ELM)

ELM, is a single-hidden layer feed forward neural networks (SLFN) based algorithm having its kernel matrix relating only to the input data and number of samples to be trained without considering the number of outputs nodes or the training target values.[18] The kernel function of the ELM was selected for this study; and the hidden nodes of its hidden layer established using trial-and-error experiment as described by Qiu *et al*.[18] It was selected as an algorithm coupled to both systems because it is fast, highly scalable and has less computational complexity;[19] hence appropriate for real time applicability.

Support Vector Machine (SVM)

SVM is founded on statistical learning theories and established as an effective classification algorithm superior to the traditional algorithms since it is supported on the principle of structural risk minimization (SRM), which minimizes greatly the expected risks rather than the error on the training data.[20] It transforms the original data into higher dimensional space using its kernel function known to influence its performance.[18] This study chose the Gaussian kernel function which gives a better SVM performance. The determination of its penalty parameter (δ^2) and kernel parameter (γ) were therefore done prior application and based on the grid search and root mean square error of cross-validation as described by Zhao *et al*.[21] The δ^2 and γ were optimized by selecting the best classification rate output with the lowest root mean square error of cross validation.

Linear Discrimination Analysis (LDA)

The LDA is an unsupervised linear chemometric algorithm that maximizes between groups variances and minimizes within groups variances in order to achieve sample classification. Models built with LDA are characterized by linear dependence on classification scores such that its linear discriminants (LDs) successively contain decreasing amount of variations, which explain decreasing information in the data set to be classified;[18] and its models achieve sample classification with the LDs contributing the highest variations in the dataset information.

K- nearest neigbhour (k-NN)

It is also a linear supervised chemometric algorithm based on the classification of an unknown sample from a test set using the k-nearest neighbour of the majority in the calibration set. Its k value influences the classification rate of the model. Thus, the optimum value of k that gives the lowest error rate was selected throughout the calibration process.

2.6 Data processing and modelling

The pre-processed spectra and the corrected image data of the ninety (90) cocoa bean samples acquired with the NIR-CDS and the CS e-nose systems were each divided randomly using the ratio 2:3 into two groups of datasets namely, prediction and calibration sets respectively. The calibration set made of 54 elements was used for building of the classification models whereas the prediction set consisting of 36 elements was used for testing the predictive strength of the built models. The four (4) chemometric algorithms were applied to the data acquired with both systems to develop models used to classify the samples. The models performances were assessed based on the largest recognition (classification) rates resulting from the built models expressed in terms of the correlation coefficient of determination of the prediction set (Rp) and that of the calibration set (Rc). All the classification models were cross-validated with 80% of calibration set (72) elements used as the validation set, with the coefficient of determination of cross-validation denoted by (Rv). Data pre-processing, processing, modelling and computations were all executed with Matlab R2013a application (MathWorks Inc., U.SA).

3. Results and discussion

3.1 Principal component analysis (PCA) and Cluster Analysis (CA)

PCA was performed on the data extracted from all the twelve (12) chemo-dyes points and point 9 printed with TPP-Co: 5, 10, 15, 20 - Tetrakis (4 – methoxyphenyl) - 21H, 23H - porphine cobalt (II) was selected as the only dye point that showed a clear classification trend based on the contributions of the

View Article Online
DOI: 10.1039/C7AY01751K

variations in it top three PCs (Table S1); and its classification trend shown in Fig. 3b.

PCA compressed the data acquired via both systems such that over 95% of total variations in them were contained in the first three principal components (PCs). The PCA for data acquired with the NIR-CDS based on dye point 9 had its total variations distributed per PCs as PC1 (98.36%), PC2 (1.02%) and PC3 (0.42%), with the variations in its first three PCs summing up to 99.38% of the total variations in the sample. Similarly, the sum of the first three PCs based on the CS e-nose data yielded 97.37% (PC1= 80.47%, PC = 15.16%, PC3 = 1.74%) of the total variations in the sample. Since the first three PCs of the data obtained from both systems contained over 95% of the cumulative variations, it therefore indicates that all variations in the cocoa beans samples can be catered for by PC1, PC2 and PC3.[22] This suggests the existence of classification pattern in the sample data as seen in Fig. 3.

[Here for Fig. 3]

Data acquired with the two systems were further subjected to cluster analysis (CA) with dendrograms. The cophenetic correlation coefficients obtained from the CA for the data acquired via CS e-nose and NIR-CDS yielded 0.88 and 0.85 respectively, indicating the classification patterns achieved are reflective of the sample data collected; and this classification pattern is clearly seen in the dendrograms captured in Fig. 4.

[Here for Fig. 4]

A careful examination of the dendrograms in Fig. 4 for both the CS e-nose and NIR-CDS clearly showed narrower Euclidean distances between the non-fermented and partially fermented beans, indicating these two bean categories are more closely related compared with the fully fermented beans. The Euclidean distance between the fully fermented bean class and the partially fermented bean class in the CS e-nose dendrogram yielded (5.5 - 4.5 = 1.0) and that in the NIR-CDS was (1.5 – 1.0 = 0.5); whereas the Euclidean distance between the partially fermented and the non-fermented beans class in the CS e-nose yielded (4.8 - 4.5 = 0.3) and that in NIR-CDS gave (0.7 – 0.6 = 0.1). Since fermentation is an oxidation reaction that produces many volatile and non-volatile compounds that give the cocoa bean it is characteristic chocolate aroma,[23] the fully fermented beans may have contained more volatiles compounds sensitive to the dyes employed in both systems compared with the partially fermented and non-fermented beans that contained fewer volatiles and their precursors.

3.2 Spectral analysis

The outcomes of the SNV and MSC pre-processed spectra obtained for the NIR-CDS based on dye point 9 produced similar effects (seen in Fig. S1). Also, both worked in the X space direction, as reported for SNV by Teye *et al.*[24] Comparison of the raw spectra acquired with the NIR-CDS and its SNV pre-processed spectra showed similar peaks and troughs as seen in Fig. 5.

[Here for Fig. 5]

Observably, the NIRS-CDS obtained spectra showed various characteristic peaks and troughs between 1000 - 2250nm; which could only be ascribed to the sensitivity of chemo-dye (TPP-Co) to the volatile organic compounds (VOCs) in the cocoa beans exposed to it in clingfilm sealed beakers.

However, because the highest loadings of a PCA contribute most to the discrimination of samples, the loadings of the top three PCs (PC1, PC2 and PC3) were plotted and analysed to explain the spectral discrimination of the cocoa beans into the three fermentation levels by the NIR-CDS system, which employed near-infrared to scan volatiles that interacted with TPP-Co from the cocoa bean samples. NIRS profile for chemicals and their functional groups was used to assign the identified spectra peaks of the largest PC loadings to their possible related chemicals. The PCs loadings plot shown in Fig. 6 revealed the largest loadings of PC1 showed peaks at 1345nm, assigned to first overtone O-H;[25] 1213nm, associated with first overtone of C-H combinations and linked to C-H and C-H_3 bonds;[1, 26] 1180nm, linked with C-H second overtone and contains C-H bonds;[26] and 1020nm, assigned to O-H second overtone, O-H and N-H.[27] These peaks of PC1 loadings are related to sugar, fats, esters and water respectively.[25-27] Peaks for the largest loadings for PC2 were observed at 1213nm, which was earlier established to be related to fats;[1, 26] and 1550nm, associated with O-H and is related to glucose.[28] PC3 largest loadings peak at 2220nm is associated with –CH=CH combination, CH stretching and C=O stretching, which are related to volatile acids (organic acid) such as lactic and acetic acid.[29] The volatile compounds related to the largest loadings of the three top PCs (PC1, PC2 and PC3) have been reportedly identified and established to vary in concentrations for the different fermentation levels of cocoa beans.[23] The variations in the concentrations of these compounds may have led to the quality classification of the beans into the different fermentation levels by the NIR-CDS and CS e-nose systems.

[Here for Fig. 6]

The presence of different peaks also suggest that the TPP-Co dye may be sensitive to different volatiles present in the cocoa beans to varied extents as revealed by Kutsanedzie *et al.*[30] and Askim *et al.*[31]

3.3 Classification models performance analysis

Data acquired from both systems were subsequently used in building classification models by employing ELM, K-NN, LDA and SVM. The classification rates achieved with the built models for CS e-nose and NIR-CD systems are respectively summarized in Table 1-2.

The single layer feed forward neural networks nodes (SLFNs) kernel function of the ELM was selected and the hidden nodes of its hidden layer was established based on trial-and-error experiments. It is possible accuracy rates were mapped with the number of the hidden nodes in the hidden layer from 10 to 19 nodes in 10 experimental trials as described by Qiu *et al.*,[18]

View Article Online
DOI: 10.1039/C7AY01751K

with a running time of 0.0312s obtained. Outcomes of its built model classification rates for cocoa bean samples in both systems are shown in Table 1 and 2. In the case of SVM, the best cross-validation accuracy was established via the search grid for the optimum penalty parameter (δ2) and kernel parameter (γ) of (3.0, 3.1) for the data acquired with both systems all within the range of (6, 7) as shown in Fig. 7. The running time achieved using the SVM for the cocoa beans classification was 0.001s, which was faster compared with that of the ELM (0.0312s).

[Here for Figure 7]

The optimized kernel parameter (k) values of k-NN algorithm were obtained for each system by calculating the prediction ability of different k-values that gave the highest classification rates shown in Fig. S2-3. The results of the classification and cross validation rates for k-NN algorithm combined with both systems are captured in Table 1 and 2.

[Here for Table 1]

For the LDA, its linear discriminant function (LDs) were used to separate the cocoa samples into the various fermentation levels via the maximization of the computed ratio of the between levels variances and the minimization of the within levels variances to generate the classification and cross-validation results summarized in Table 1 and 2.

Despite the classification rates recorded in the two systems combined with the four chemometric algorithms yielded satisfactory results, the nonlinear algorithms (ELM and SVM) gave comparatively higher classification rates (Rp = 94%) compared to the linear ones (LDA and k-NN), which ranged between (85% ≤ Rp ≤ 92%). The comparative lower performance of LDA and k-NN may be attributed to the fact that the cocoa beans fermentation process entails complex biochemical reactions that produces different volatile substances at various stages, for which reason they may not adequately be captured by the linear methods.[20] The NIR-CDS and CS e-nose systems involved the use of chemo-dyes, in which case the beans classification was solely based on the degree of sensitivity of the various chemo-dyes to the volatiles present in the beans. ELM and SVM yielded same classification rate (Rp = 94%), however, per the running time, the latter was faster. However, the k-NN recorded the lowest classification rate of (Rp = 85%), in that it considers the Euclidean distance to assign members to classes rather than the structure of the class.[32]

[Here for Table 2]

The bean classification rates achieved in the CS e-nose (89% ≤ Rp ≤ 94%) and NIR-CDS (85% ≤ Rp ≤ 94%) are comparable, suggesting both systems could be employed as alternatives for monitoring cocoa bean quality non-destructively and on real time basis.

According to Schwan and Fleet[33]during cocoa fermentation process, various volatile and non-volatile compounds are produced in varying quantities at different stages. The classification of the beans samples into their three fermentation levels (FF, PF and NF) with only a chemo-dye (TPP-Co) via the NIR-CDS system indicates that TPP-Co may have wide range and degrees of sensitivity to the various volatiles in the beans.[34] Based on this, the NIR-CDS system may therefore offer a pragmatic solution to the selection of sensitive chemo-dyes for fabrication of chemo-dyes used in e-nose systems as opposed to the trial-and-error methods currently being employed, which is time-consuming and dyes wasteful. This outcome also implies NIRS may be used to detect volatile (gaseous) compounds in samples via the use of chemo-dyes as capture probes, which hitherto had being its limitation - its inability to directly detect gaseous compounds.

4. Conclusions

CS e-nose and NIR-CDS systems combined with four different ELM, SVM, LDA and k-NN were applied successfully to classify cocoa beans into three quality categories such as fully fermented, partially fermented and non-fermented. Results revealed that the two systems combined with the nonlinear chemometric algorithms (SVM and ELM) achieved high classification rates (Rp = 94%). These results suggest that the two systems may possibly be deployed as alternative low-cost systems for cocoa beans quality monitoring on real time and non-destructive basis. In addition, the NIR-CDS system demonstrated the capability of using chemo-dyes to capture volatile (gaseous) compounds whose spectra could be scanned, extracted and modelled to solve quality monitoring problems. This technique therefore has significant application to FT-NIRS and NIRS systems for the detection of volatile (gaseous) compounds in materials, which had been a limitation for these systems.

Conflict of interest

We declare no conflict of interest.

Acknowledgements

We would like to acknowledge with profound appreciation the financial support from the National Natural Science Foundation of China (31471646) and Key R&D Program of Jiangsu Province, China (BE2015302).

Notes and references

1. A. Krähmer, A. Engel, D. Kadow, N. Ali, P. Umaharan, L. W. Kroh and H. Schulz, *Food chemistry*, 2015, **181**, 152-159.
2. A. Belščak, D. Komes, D. Horžić, K. K. Ganić and D. Karlović, *Food Research International*, 2009, **42**, 707-716.

View Article Online
DOI: 10.1039/C7AY01751K

3. P. C. Aculey, P. Snitkjaer, M. Owusu, M. Bassompiere, J. Takrama, L. Nørgaard, M. A. Petersen and D. S. Nielsen, *Journal of Food Science*, 2010, **75**, S300-S307.
4. R. Saltini, R. Akkerman and S. Frosch, *Food Control*, 2013, **29**, 167-187.
5. Y. Suazo, G. Davidov-Pardo and I. Arozarena, *Journal of Food Quality*, 2014, **37**, 50-56.
6. E. Teye, X. Huang, H. Dai and Q. Chen, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2013, **114**, 183-189.
7. F. Gu, L. Tan, H. Wu, Y. Fang, F. Xu, Z. Chu and Q. Wang, *Foods*, 2013, **2**, 183-197.
8. A. Vesela, A. S. Barros, A. Synytsya, I. Delgadillo, J. Copikova and M. A. Coimbra, *Analytica chimica acta*, 2007, **601**, 77-86.
9. F. Y. Kutsanedzie, Q. Chen, M. M. Hassan, M. Yang, H. Sun and M. H. Rahman, *Food chemistry*, 2017,DOI: http://dx.doi.org/10.1016/j.foodchem.2017.07.117.
10. M. Ghasemi-Varnamkhasti and M. Aghbashlo, *Trends in Food Science & Technology*, 2014, **38**, 158-166.
11. W. Pan, J. Zhao, Q. Chen and L. Yuan, *RSC Advances*, 2015, **5**, 60876-60883.
12. G. ElMasry, D.-W. Sun and P. Allen, *Journal of Food Engineering*, 2012, **110**, 127-140.
13. J. U. Porep, D. R. Kammerer and R. Carle, *Trends in Food Science & Technology*, 2015, **46**, 211-230.
14. W. Zheng, X. Fu and Y. Ying, *Chemometrics and Intelligent Laboratory Systems*, 2014, **139**, 42-47.
15. Q. Chen, J. Zhao, C. Fang and D. Wang, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2007, **66**, 568-574.
16. M. O'farrell, E. Lewis, C. Flanagan, W. Lyons and N. Jackman, *Sensors and Actuators B: Chemical*, 2005, **111**, 354-362.
17. E. González-Rufino, P. Carrión, E. Cernadas, M. Fernández-Delgado and R. Domínguez-Petit, *Pattern Recognition*, 2013, **46**, 2391-2407.
18. S. Qiu, L. Gao and J. Wang, *Journal of Food Engineering*, 2015, **144**, 77-85.
19. G.-B. Huang, Q.-Y. Zhu and C.-K. Siew, *Neurocomputing*, 2006, **70**, 489-501.
20. Q. Chen, J. Zhao and H. Lin, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2009, **72**, 845-850.
21. J. Zhao, H. Lin, Q. Chen, X. Huang, Z. Sun and F. Zhou, *Journal of Food Engineering*, 2010, **98**, 408-414.
22. H. Cheng, Z. H. Qin, X. F. Guo, X. S. Hu and J. H. Wu, *Food Research International*, 2013, **51**, 813–822.
23. V. T. T. Ho, J. Zhao and G. Fleet, *International Journal of Food Microbiology*, 2014, **174**, 72-87.
24. E. Teye, X. Y. Huang, W. Lei and H. Dai, *Food Research International*, 2014, **55**, 288-293.
25. F. Westad, A. Schmidt and M. Kermit, *Journal of Near Infrared Spectroscopy*, 2008, **16**, 265-273.
26. J. Moros, F. A. Iñón, S. Garrigues and M. de la Guardia, *Analytica Chimica Acta*, 2007, **584**, 215-222.
27. D. Cozzolino, H. E. Smyth and M. Gishen, *Journal of Agricultural and Food Chemistry*, 2003, **51**, 7703-7708.
28. H. Cui, L. An, W. Chen and K. Xu, *Opt. Express*, 2005, **13**, 6887-6891.
29. T. R. Viegas, A. L. Mata, M. M. Duarte and K. M. Lima, *Food Chemistry*, 2016, **190**, 1-4.
30. F. Y. Kutsanedzie, L. Hao, S. Yan, Q. Ouyang and Q. Chen, *Sensors and Actuators B: Chemical*, 2017, **254**, 597–602.
31. J. R. Askim, M. Mahmoudi and K. S. Suslick, *Chemical Society Reviews*, 2013, **42**, 8649-8682.
32. L. A. Berrueta, R. M. Alonsosalces and K. Héberger, *Journal of Chromatography A*, 2007, **1158**, 196-214.
33. R. F. Schwan and G. H. Fleet, *Cocoa and coffee fermentations*, Crc Press, Boca Raton, 2014.
34. L. Jing, X. Jianhua and X. Shuang, *Energy Procedia*, 2011, **12**, 625-631.

View Article Online
DOI: 10.1039/C7AY01751K

View Article Online
DOI: 10.1039/C7AY01751K

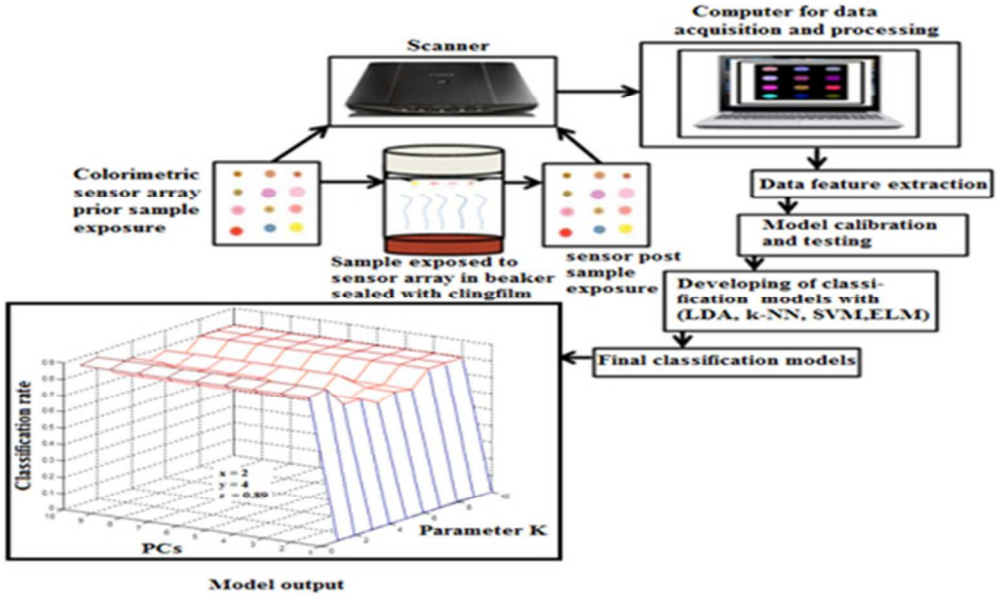


Fig. 1: CS e-nose system setup for cocoa bean fermentation levels-based classification

49x30mm (300 x 300 DPI)

View Article Online
DOI: 10.1039/C7AY01751K

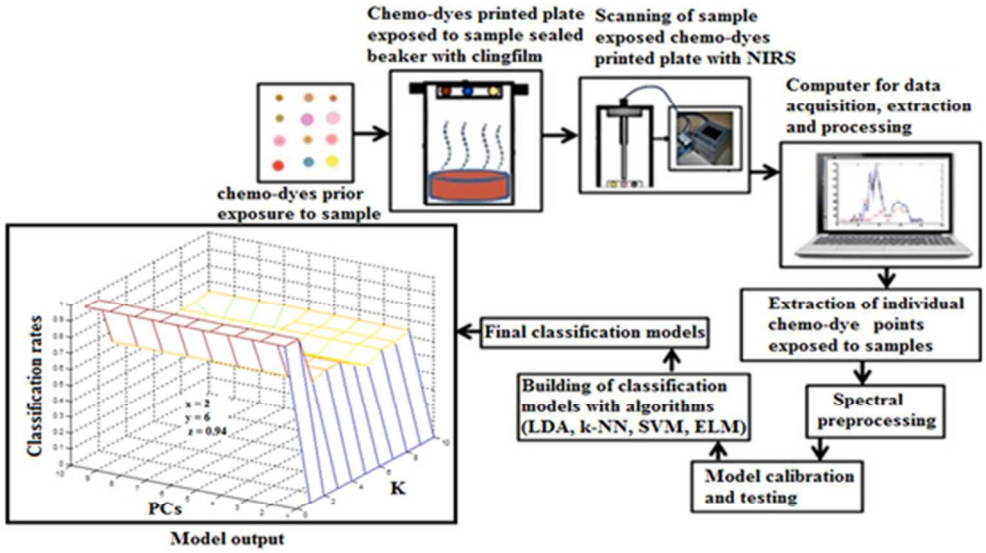


Fig. 2: NIR-CDS system setup for cocoa bean fermentation levels-based classification

49x30mm (300 x 300 DPI)

View Article Online
DOI: 10.1039/C7AY01751K

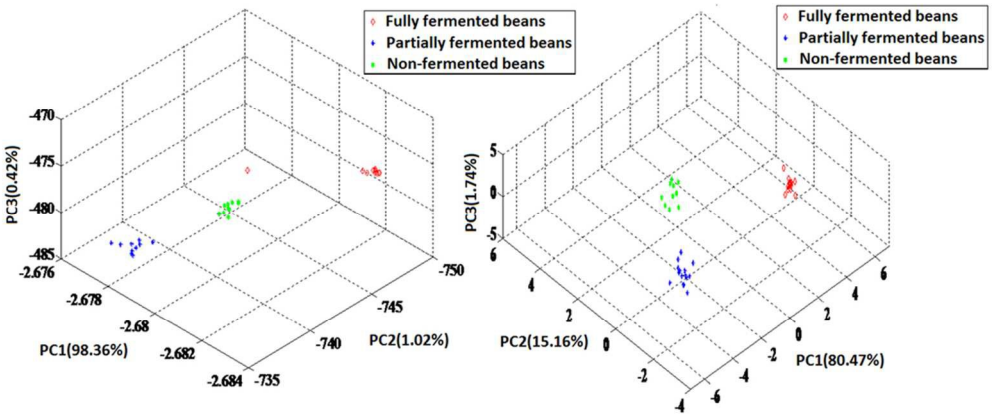


Fig. 3: PCA scatter plot for data obtained from (a) CS e-nose system; (b) NIRS-CDS system based on TPP-Co

91x38mm (300 x 300 DPI)

View Article Online
DOI: 10.1039/C7AY01751K

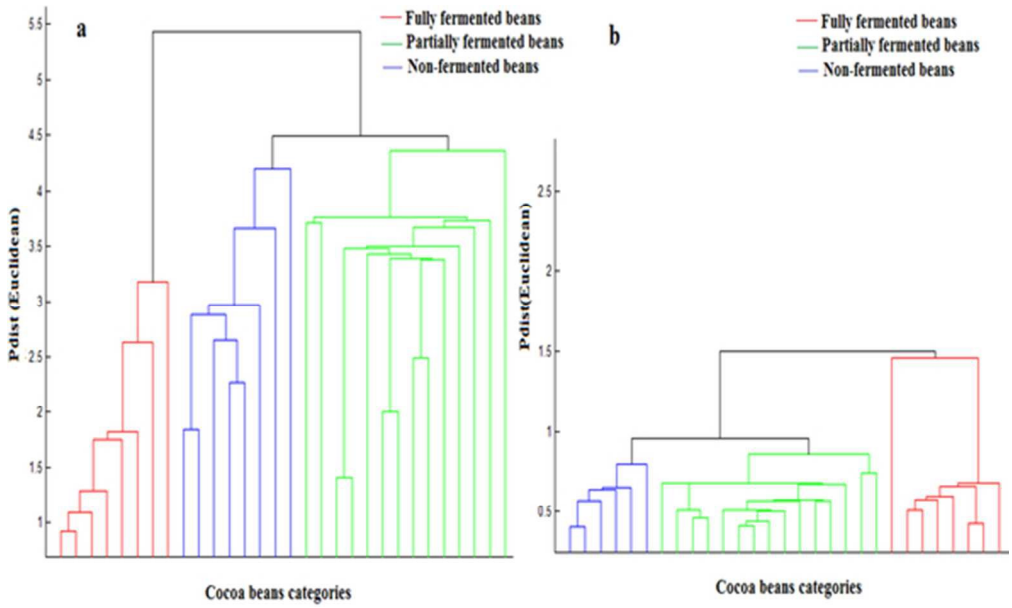


Fig. 4: Dendrograms for the data obtained from (a) CS e-nose and (b) NIR-CDS systems

49x30mm (300 x 300 DPI)

View Article Online
DOI: 10.1039/C7AY01751K

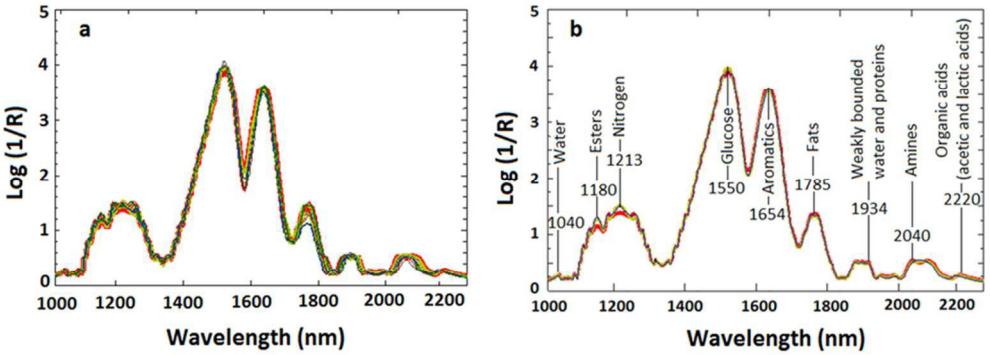


Fig. 5: Raw spectra of cocoa beans acquired with (a) NIR-CDS based on TPP-Co; and (b) its SNV preprocessed spectra

78x28mm (300 x 300 DPI)

View Article Online
DOI: 10.1039/C7AY01751K

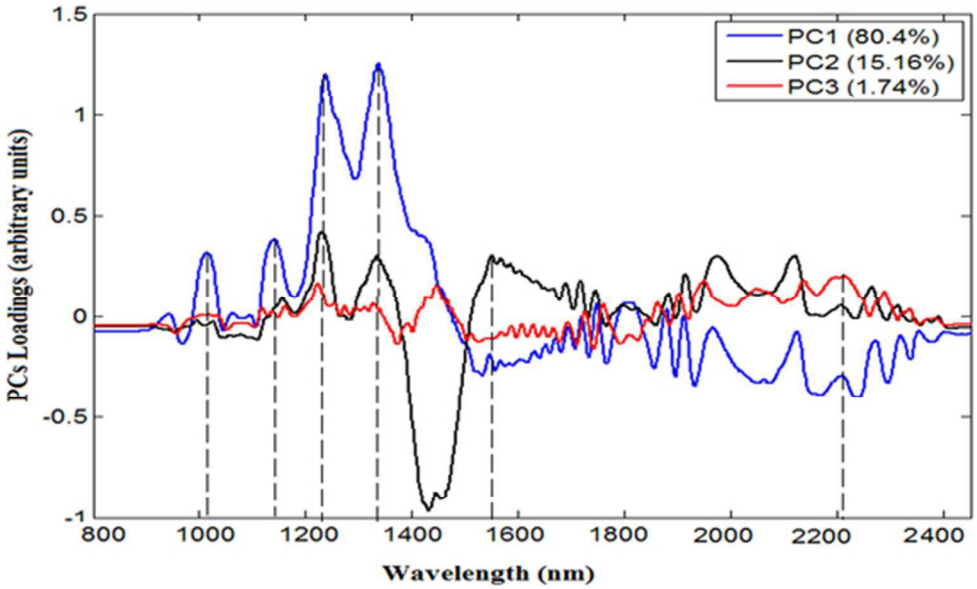


Fig. 6: Plot of top three PCs loadings against wavelength of cocoa beans spectra with NIRS-CDS system

49x30mm (300 x 300 DPI)

View Article Online
DOI: 10.1039/C7AY01751K

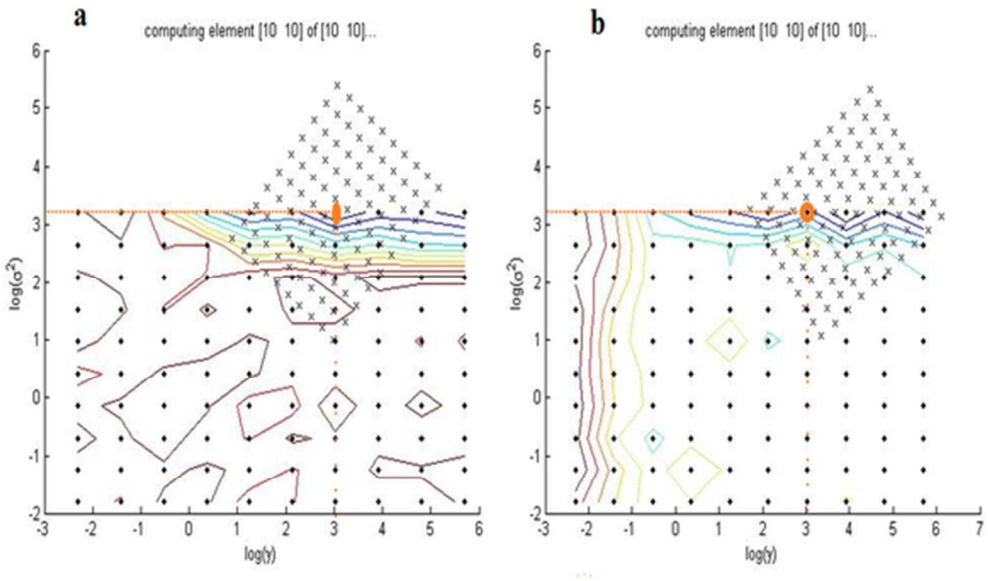


Fig. 7: Generated search grid diagram for the SVM optimum penalty parameter (δ2) and kernel parameter (γ) in (a) CS e-nose and (b) NIR-CDS systems

49x30mm (300 x 300 DPI)

View Article Online
DOI: 10.1039/C7AY01751K

Table 1: Results of the Cocoa beans fermentation levels classification models achieved in CS e-nose system combined with different chemometric algorithms.

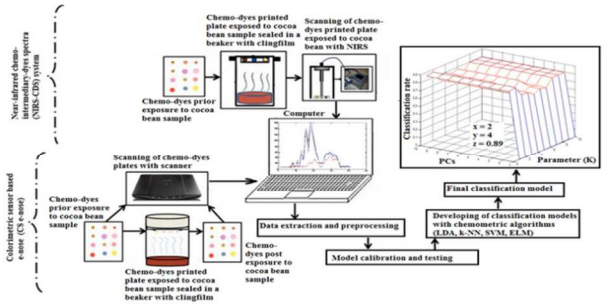
System	Chemometric algorithms	Principal component (PCs)	Classification rate (%)		
			Calibration set (Rc)	Validation Set (Rv)	Prediction set (Rp)
CS e-nose	ELM	3	98	97	94
	K-NN	7	97	97	90
	LDA	4	93	91	89
	SVM	3	100	100	94

View Article Online
DOI: 10.1039/C7AY01751K

Table 2: Results of the Cocoa beans fermentation levels classification models achieved in NIR-CDS system combined with different chemometric algorithms.

System	Chemo-dye point	Pre-processing methods	chemometric algorithms	PCs	Classification rate (%)		
					Calibration set (Rc)	Validation Set (Rv)	Prediction set (Rp)
NIR-CDS	9 (TPP-Co)	SNV	ELM	3	100	100	94
			K-NN	4	89	89	89
			LDA	4	100	92	89
			SVM	4	100	100	94
		MSC	ELM	3	98	94	94
			K-NN	6	94	90	85
			LDA	4	100	92	92
			SVM	4	100	100	94

View Article Online
DOI: 10.1039/C7AY01751K



Cocoa beans were quality graded innovatively using a near-infrared chemo-dyes system as aroma sensor to capture and detect their volatiles.