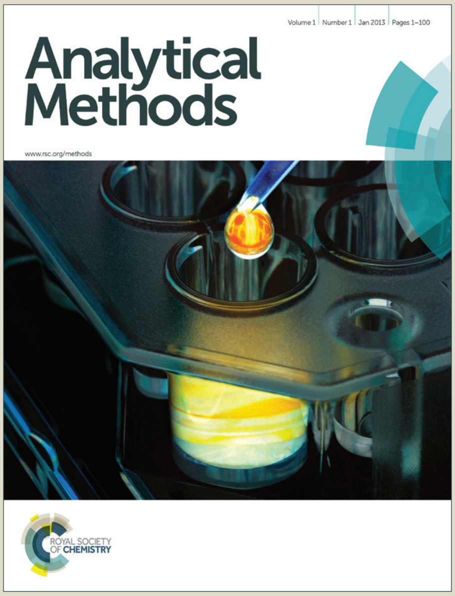
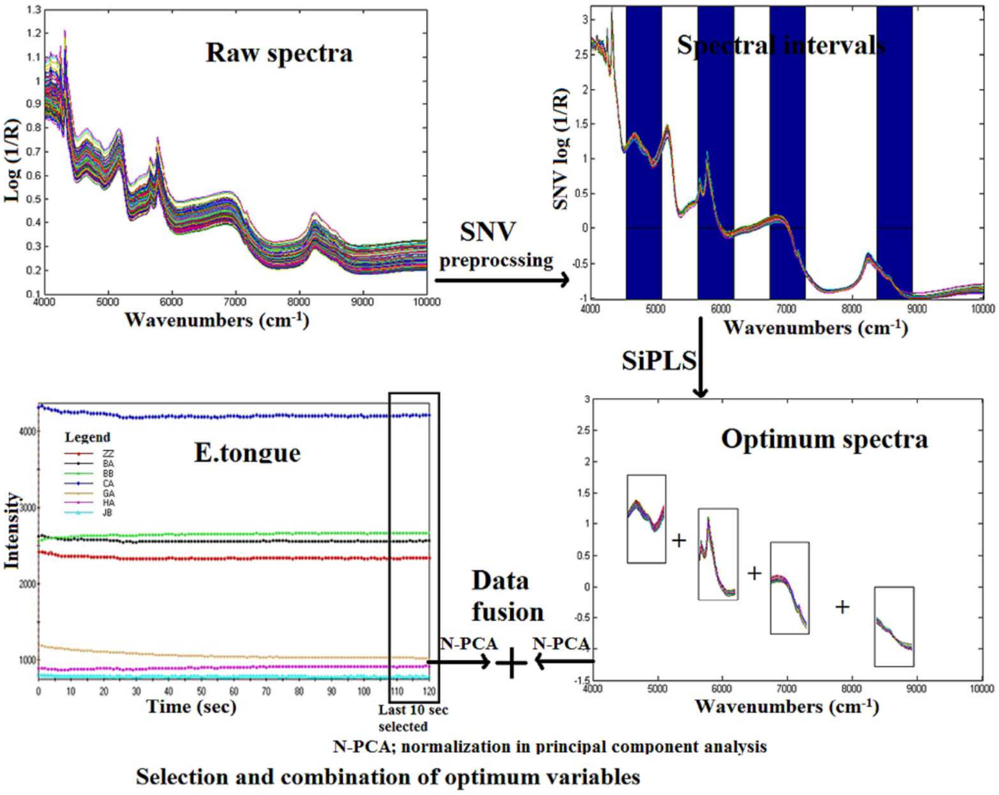
Analytical Methods

Accepted Manuscript


Volume 1 | Number 1 | Jan 2013 | Pages 1–100
Analytical Methods
www.rsc.org/methods
ROYAL SOCIETY OF CHEMISTRY

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/methods



1428x1147mm (96 x 96 DPI)

Rapid measurement of total polyphenols content in cocoa beans by data fusion of NIR Spectroscopy and Electronic tongue

*Xingyi Huang[1], Ernest Teye[1, 2], Livingstone K. Sam-Amoah[2], Fangkai Han[1], Liya Yao[1] and William Tchabo[1]

[1]School of Food and Biological Engineering, Jiangsu University,
Xuefu Road 301, Zhenjiang 212013, Jiangsu, P. R. China

[2]School of Agriculture, University of Cape Coast, Cape Coast, Ghana

*Corresponding author: Email: h_xingyi@163.com/teyernest@gmail.com

ABSTRACT

Total polyphenols content (TPC) is an important phytochemicals in cocoa beans due to its numerous health benefits. This work attempted to measure the total polyphenols content in cocoa beans by using a novel approach of integrating near infrared spectroscopy (NIRS) and Electronic tongue (ET). 110 samples of cocoa beans with different polyphenols content were used for data acquisition by NIRS and ET respectively. The optimum individual characteristic variables were extracted from technique and scaled by normalization in principal component analysis (PCA). Support vector machine regression (SVMR) was used to construct the model. The performance of the final model was evaluated according to: correlation coefficient (R_{pre}), root mean square error of prediction (RMSEP) and bias in the prediction set. Compared with a single technique (NIRS or ET), the data fusion was superior for the determination of TPC in cocoa beans. The optimal data fusion model was achieved with: R_{pre}=0.982, RMSEP=0.900 g/g and bias=0.013 in the prediction set. The overall results demonstrate that integrating NIRS and ET is possible and could improve the prediction of TPC in cocoa beans.

Keywords**:** Near infrared spectroscopy, Electronic tongue, Data fusion, Support vector machine regression, Total polyphenols content

1. INTRODUCTION

Cocoa bean products are increasingly becoming a popular beverage worldwide due to its numerous health benefits or medicinal properties. Recent studies have shown that the consumption of cocoa bean products can enhances the general well being of humans due to the present of polyphenols. Cocoa bean contains more polyphenols per serving than tea and coffee [1]. These phytochemicals have important role in preventing coronary artery disease, cancers and it is a myocardial stimulant, diuretic, coronary dilator and muscle relaxant[2-4]. Also, polyphenols compounds in cocoa beans are mainly responsible for the characteristic; taste, flavour and astringency of the fermented cocoa beans. It is therefore very important to determine the total polyphenols in cocoa beans and over the years, the methods employed for the determination of total polyphenols content include: colorimetric [5],thin-layer chromatography [6] and high-performance liquid chromatography [7]. However, these reputable analytical methods are expensive, time consuming, destructive, involves chemical usage, and very tedious.

To overcome these drawbacks, near infrared spectroscopy (NIRS) and electronic tongue (ET) has emerged as a novel tool for qualitative and quantitative measurements. These methods are; fast, accurate, reliable and non-destructive with no chemical usage .Together with recent advancement in computers and chemometrics, NIRS and ET have been used in various fields such as agricultural, nutritional, medicinal and petrochemical [8] and process monitoring, freshness evaluation, authentic assessment, foodstuffs recognition and quality analysis [9] respectively. Specifically in previous studies, NIRS has been used to determine various phytochemicals in cocoa beans such as protein, fat, carbohydrate, nitrogen and moisture content[10-12]. Furthermore, Alvarez and co-workers [13] determined fats, caffeine, theobromine and epicatechin in unfermented and sun dried criollo cocoa and Whitacre and others [14] predicted the content of cocoa procyanidins. While, ET has been used by other research such as: Teye, et al[15] for discrimination of cocoa beans according to geographical origin, Chen, et al. [16] for identification of green tea grade level. Also, Chen, et al. [17] used taste sensor technique to determine caffeine and main catechin content in green tea. Other studies are analysis of goat milk adulterated with bovine milk [18], detection of sugars and acids in tomatoes[19].

Although the combination of NIRS and ET is most likely to increase the performance of measurements, articles in this area are lacking. Also, upon a thorough literature search, little information is available on the use either NIRS or ET for rapid analysis of total polyphenols contents in cocoa bean. More so, the use of NIRS for the prediction of cocoa procyanidins, theobromine, and epicatechin was done with partial least squares (PLS) regression and modified PLS model. The modified PLS was used to manually select different spectral band and this might weaken the performance of the claibration model without prior experienced in the knowedge about NIRS. Nørgaard et al[20] developed synergy interval PLS (SiPLS) to select several intervals spectra data which could split the whole wavelength range into a number of intervals and calulate all possible PLS model combination of 2, 3, or 4 subintervals for optimum prediction. Furthermoe, the analysis of total polyphenol content is a complex and complicated process. Total polyphenol contents in cocoa bean is made up of mainly catechin (37%), procyanidins (58%), and anthocyanins (4%) [6]. These chemical compounds affect both external attributes and internal chemical properties of polyphenols. For instance, polyphenol in cocoa beans gives some uique taste and astrigent characteristics known as polyphenol bitterness and astringency [21] and higher polyphenol concentration leads to an increase in astringent-tasting chocolate [22] and cocoa liquor

[23]. Bonvehi and Ventura [24] also, found a correspondence between sensory data and polyphenolic compounds. Moreso, polyphenol imparts a red to purple to brown colour through oxidation of anthocyanins to quinonic compounds [24, 25]. For instance, anthocyanins are the most important group of plant pigment that are responsisble for colour [6]. A single prediction technique can atmost describe one aspect and multiply sensor fusion could prove very useful. Therefore, data fusion of NIRS and ET is most likely to increase the quantitative prediction performance of total polyphenols content in cocoa bean.

Data fusion of sensor is an effective way for the optimum utilization of two or more sensors, which seeks to combine information from multiple sensors to achieve inferences that are more feasible than a single sensor [26]. Literature information on data fusion is very few. Huang et al [27] predicted total volatile basic nitrogen in port by data fusion of three sensors techniques, Winquist et al [28] combined ET and electronic nose for solving classification problem and Ulla and coworkers [29] also, determined the botanical origin of honey by sensor fusion of ET and optical spectroscopy. The objectives of this studies were (1) to analyse the total polyphenols contents in cocoa beans by NIRS spectroscopy and ET, (2) to extract the optimum individual characteristic variables from each sensor data, and (3) data fusion of NIRS and ET for accurate and reliable prediction of total polyphenols contents in cocoa beans.

2. MATERIALS AND METHODS

2.1. Sample preparation

In this experiment, 110 cocoa bean samples were collected from different cocoa growing regions of Ghana under the supervision of the quality control division of the Ghana cocoa board. The beans samples were accurately labelled and transported to Jiangsu University, School of Food and Biological Engineering laboratory for further analysis. Considering the heterogeneities of the beans each sample was ground separately for 15 seconds by a small multi-purpose grinder (QE-100, Zhejiang YiLi Tool Co., Ltd. China). The powders of each sample were sieved with a 500 μm mesh before further analysis.

2.2. NIR Spectra collection

The spectra of each sample were collected in the reflectance mode by Antaris II Near Infrared Spectrophotometer (Thermo Electron Company, USA) with an integrating sphere and the reflectance (R) data were stored as absorbance (A) = Log (1/R). 10 g of the sample was collected into a standard sample cup and the spectra were scanned three times (after rotating the cup120^0) with a spectral resolution of 8.0 cm^{-1}. The experiment was conducted at a temperature of 25 ^{0}C and at humidity of 60%. Each spectrum was an average of 32 scans in the range of 4000-10000 cm^{-1} and the raw data were measured in 3.856 cm^{-1} interval resulting in 1557 variables. The mean of the three spectra collected from the same cocoa bean sample was used for subsequent analysis.

2.3 Electronic tongue data acquisition

The electronic tongue device used was α-Astree brand (Alpha MOS Company, Toulouse, France). The sensor array used comprises seven potentiometric chemical sensors such as ZZ, BB, CA, GA, HA, and JB and a reference electrode. The sensitivity of the sensors, differs from the five tastes; sourness, saltines, sweetness, bitterness and savoury [30]. The sensors are made with silicon transistors and organic coated to ensure that they are sensitive and selective to liquid samples. 1.0 g of each sample was accurately weighed into a beaker and 100 ml boiled distilled water added (0.01 gm/l). It was allowed to cool and then filtered through a filter paper. 80 ml of the filtrate was poured into a beaker and sent to the electronic tongue. Five samples were detected at once and the intensity values of each sensor values recorded. The data were collected at room temperature of 25 ^{0}C and humidity of 60%.

2.4. Determination of total polyphenols content

The determination of total polyphenolic content was done by a colorimetric assay using Folin-Ciocateu phenol reagent [31] with few modifications according to Romero-Cortes and co-workers [25]. The values for the total phenolics were expressed in percentage, in terms of gallic acid equivalents (g GAE /g of dry matter) with a standard curve of Pearson's correlation coefficient (R^2)=0.9970. Gallic acid was used because it is more stable and pharmacologically active antioxidant, quantitatively equivalent to most phenolics and gives consistent and reproducible results [32-34].The difference between two parallel measurements was less than 0.10%.

2.5 Software

All calculations and algorithms were carried out in Matlab Version 7.14 (Mathworks Inc., USA) with Windows 7 ultimate for data processing. Antaris II System (Thermo Electron Company, USA) was used for spectra acquisition.

2.5 Initial data processing

Standard normal variate (SNV)[35] was applied on the NIR spectra to remove slope variation and correct scatter effects due to particle size, so that the performance of PCA and the model will be based mainly on chemical spectral information [36].

Si-PLS proposed by Norgaard and co-workers [20] was used to select the optimum NIR spectra wave band range (several subintervals) with the highest predictive performance and the lowest prediction errors for the analyses of total polyphenols content in cocoa beans. Si-PLS works by splitting the data set into a number of intervals and then calculates accurately all possible PLS regression models for all possible combinations of 2, 3, or 4 intervals and the combination of intervals with the lowest RMSECV for optimum performance are obtained [37]. Furthermore, for the E.tongue data, the last 10 s of the entire 120 s were selected after several attempt and this time was found to be effective, because they were more stable and this was similar to other researchers [15, 30].

2.7. Calibration and Prediction set

The data set used in this experiment was made up of 110 samples. These were divided into two subsets called: calibration set (80 samples) and prediction set (30 samples). The calibration set was used to develop the model, while the prediction set was used for evaluating the actual predictive ability of the developed models. The individual sample in each set was selected randomly in order to come to approximately 3/1 division of calibration set/prediction set. To avoid bias in subset division, the subset was done as follows: for every 4 samples, about 3 were randomly selected as the calibration set while the remaining was used as the prediction set.

2.8. Theory of data fusion techniques

Data fusion techniques are normally classified according to abstraction levels at which data from different instruments are merged [38] and these include: high level abstraction (HLA), mid-level abstraction (MLA) and low-level abstraction (LLA). HLA consists of merging information at a higher level of abstraction that is; combining the results from multiple algorithms to yield a final fused decision (decision making fusion) [26]. HLA assumes that data from each sensor system are analyzed as a stand-alone set and afterwards the important features are extracted from each data set before merging them. Thus, the most prominent feature is selected before data integration and this leads to huge data processing with tremendous information losses [27]. This approach is biologically inspired, because in the human multisensory systems it is possible to retain the perception of each single sense and, at the same time merge them together to form a complex judgment [39]. MLA is also known as feature level fusion, it involves the integration of feature variables of two or more sensor signals. MLA is strong in keeping enough of the original variables. This approach was previously used to study the correlation between different instrumental techniques applied to the same samples [40] and has recently been used by Huang and co-workers [27]. LLA refers to original data fusion, it requires that, all the data from different sensors are simply concatenated before constructing the model [41]. Thus; after data fusion, the data matrix has the number of rows equal to the number of samples and the number of columns equal to the total number of information from all sources [39]. According to Haddi and co-workers [41], the merging of measurements from two sources in LLA could potentially provide more redundant information and this can grievously affect the results. To overcome this bottle-neck it is more suitable to couple low-level abstraction to a feature selection technique [42]and principal component analysis. Thus, in this study, LLA together with feature selection technique was employed and PCA was used as sensor fusion technique after normalization. Among the three data fusion techniques, LLA is mostly used and generally gives a good results [43]. Fig.1 shows the process of optimum selection of sensors characteristic variables and fusion.

Theory of SVMR model

Support vector machine is a strong non-linear multivariate algorithm originally invented by Vapnik in 1995 and the current standard incarnated by Vapnik and Cortes [44]. SVM has recently found its application in food analysis for solving classification and regression problems. SVM algorithm constructs a hyperplane or set of hyperplanes in a high dimensional space for classification, and this principle is also applied to regression tasks [45]. Generally, the higher dimensional space is implemented by a kernel function [45]. There are three classical kernel functions namely: polynomial kernel function, radial basis function and sigmoid kernel function and the type of kernel function used influences the performance of SVM model. Among these three kernel functions, radial basis function is mostly selected, because it can handle the linear and non-linear relationships between the class labels and the spectra data, also it is capable of reducing the computational complexity of the training set thereby providing a good performance under general smoothness assumptions[46-48]. Therefore, in this study, radial basis function was computed by using equation 1.

$$K(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{\gamma^2}\right) \qquad (1)$$

Where the parameter γ is the bandwidth parameter of the radial basis function

To generate a good performance for SVMR model, penalty parameter C and kernel parameter γ were optimised in this work. Penalty parameter C determines the trade-off between minimizing the training error and minimizing model complexity, while kernel parameter γ implicitly defines the bandwidth of the radial kernel function [49]. The appropriate selection of parameter c and y guarantees a satisfactory SVMR results. In this work, the pairs of (C and γ) were tried and the model with the best performance was chosen.

2.9. Development of total polyphenols content prediction model

The total polyphenols content in cocoa beans is very complex and complicated, hence the relationship between the characteristic variable from a single sensor tends to show low correlation and appears to be non-linear. In this study, Synergy interval partial least squares (Si-PLS) was used to select the optimum variables from the SNV pre-treated NIR spectra. Si-PLS is a very powerful multivariate technique that involves the selection of variables, where the data set is split into a number of intervals (variable wise) and calculates accurately all possible PLS model combinations of 2, 3, or 4 intervals. In this study, the full spectral range of 4000-10000 cm^{-1} of the samples were divided into 8, 9, 11,···,16 intervals combined with 2, 3, or 4 subintervals were used. The optimal combination of intervals and the number of PLS factors were optimized by cross-validation. The best combinations of intervals (optimum spectra selection) were chosen according: the lowest root mean square error of cross-validation (RMSECV), root mean square error of prediction (RMSEP) and correlation coefficient (R) by the equations used by Chen and co-workers[37] respectively. Si-PLS model has been used in recent times in food analysis, and found to be superior to others for selecting optimum spectra interval for accurate prediction [37, 50].

After the optimum selection of NIR spectra variables, the last 10 seconds of the E.tongue sensor data were extracted. Principal component analysis (PCA) was implemented on the selected variables: NIRS, ET and data fusion, because the characteristic variables of each sensor contain useful correlation and some redundant information. PCA is a popular dimensionality reduction technique that is used to eliminate redundant variables and decreases the computational burden [17]. PCA is also among the most popular and effective fusion algorithm [26]. The top principal components (PCs) were also extracted from each as the input data for Support vector machine regression (SVMR) in developing the TPC prediction model respectively. SVMR is a very powerful non-linear regression based on the classical support vector machine. It has been widely used and shown its superiority over others; because it has a good generalization property, self-learning and self-adjustment characteristics and embodies structural minimization principle [51]. The number of PCs and some parameters were optimized by cross-validation in calibrating the model in the calibration set. In this study, the leave one out cross-validation (LOO-CV) was performed [52] as done by other researchers for RMSECV[17, 37]. LOO-CV was done as follows:

firstly one sample in the calibration set is removed, and the predictive model is built with the remaining samples in the calibration set. The sample removed is then predicted with the model, and the procedure is repeated with each sample left out in the calibration set. LOO-CV is the simplest, and the most common used procedure [53]. The model's performance was evaluated by these parameters: RMSECV, RMSEP, R and bias as done by other authors[37, 54].

3. RESULTS AND DISCUSSION

3.1. Reference measurement of TPC

The 110 samples of cocoa bean used in this study showed a wide range of total polyphenols contents as seen from Table1. These total phenolics were found to be between 23.02-33.67% (g GAE/g) and were consistent with other researchers [55, 56]. Furthermore, the range in calibration set cover the range in the prediction set and also the standard deviations in the calibration set and prediction set are not significantly different, therefore the distribution of the samples are appropriate in the two sets [37]. This means that, bias in the distribution of samples in the two sets were negligible and the distribution of the reference data in the calibration and prediction sets are almost equal.

3.2. Selection of optimum NIRS data (Spectral variables)

In this study, after SNV pre-processing of NIR spectra, Si-PLS algorithm was used to select optimum spectral variable. The spectrum of 4000-10000 cm^{-1} was divided into 6, 7,… and 20 intervals and the number of intervals were optimized by cross-validation. The lowest RMSECV and the best R were achieved when the full NIR spectrum (4000-10000 cm^{-1}) was split into 11 intervals and the optimum combinations of intervals were [2, 4, 6, and 9]. These intervals corresponds to the spectra range of 4547-5091 cm^{-1}, 5643-6187 cm^{-1}, 6738-7282 cm^{-1}

and 8373-8913 cm^{-1} totalling to 83 variables, as shown in Fig.1. The total efficient variables (83) were then analyzed by PCA.

3.3. Extraction of optimum Electronic tongue (ET) data

The taste sensor array measured the dissolved chemical compounds in the solution and gives the voltage difference between the sensors and the reference electrode (called Ag/AgCl, which as a fixed voltage) i.e. the voltage difference obtained refers to the voltage of the sensor (V_s) minus the voltage of the reference electrode (V_e), because the dissolved compounds in the solution and the sensor affects the voltage of the given by sensor. After the measurement, each individual sensor gave a different intensity value based on their selectivity and sensitivity characteristics to the chemical properties in the cocoa bean samples. However, the response values at the last 10 (110-120) seconds was selected as the optimal range, because it was found to be more stable and this was similar to other researches [15, 30]. Thus, 7 characteristic variables were obtained as the best selection and were analyzed by PCA.

3.4 Data fusion

After the selection of efficient variables from the two sensors (ET had 7 variables and NIR spectra had 83 variables), each sensor data was scaled by normalization [57] before PCA. The data were then merged into one, totalling 90 variables. PCA is a unique technique that is popularly used for dimensionality reduction with the aim of eliminating redundant variables and diminishing the computational burden[49]. The total variables from the fused data were extracted as an input data for SVMR modelling. The computed models were compared.

3.5. SVMR models of TPC in cocoa beans

The total polyphenols content in cocoa beans is very complicated; made up of several phenolic compounds such as theobromine, xanthine, catechin, caffeine, epicatechin, quinones and anthocyanidins etc. Some of these chemicals influence taste, aroma and colour. For instance, high polyphenols content is related to bitter-astringency properties of the cocoa beans. Therefore, cocoa beans are normally fermented, because fermentation lessens the bitter-astringent properties of the beans, an effect that is attributed to loss of polyphenols (flavan-3-ols) during fermentation[58, 59]. Support vector machine regression (SVMR) as a powerful non-linear multivariate algorithm was attempted to develop the TPC prediction model in this study, because it has been found to be superior than other in cocoa beans study[36,60]. From Table 2 and Fig. 2, it could be seen that, the best TPC model by SVMR for ET and NIRS were achieved at PCs=5 and 9 respectively. The correlation coefficient (R_{cal}) for ET and NIRS were 0.813 and 0.920 in the calibration set respectively. From this table it is observed that, when the model was tested in the prediction set, there was a reduction for both techniques especially for ET techniques. The performance for ET in the prediction set was R_{pre}=0.70, RMSEP=1.796 and bias=0.564, while NIRS was R_{pre}=0.91, RMSEP=1.674 and bias=0.276. However, the single sensor in this experiment could not give the optimum predictive performance therefore, data fusion was attempted. In fact, the data fusion of different sensor could prove useful. It can acquire more information than a single sensor and could be used to predict the TPC in cocoa bean samples. Hence, the model based on data fusion was performed and compared with the single sensors. It revealed that, the model based on data fusion was found to be superior to the others as seen in Table 2 and Fig. 3. From Fig. 3 it could be seen that data fusion model was significantly stable in the prediction set when the mode was tested as compared to the other single techniques. i.e the differences between RMSECV and RMSEP was not significant. The data fusion results showed that R_{cal}, RMSECV and bias were 0.987, 0.890 and 0.006 in the calibration set.

These results could further be explained that, NIRS wavelength range selected by SiPLS had optimum combinations of five intervals [2, 4, 6, and 9] which correspond to 4547-5091 cm^{-1}, 5643-6187 cm^{-1}, 6738-7282 cm^{-1} and 8373-8913 cm^{-1} that are related to external and some internal attributes of TPC in cocoa beans. These selected spectra are related to: (4547-5091 cm^{-1}) = CON-H amide combination bands, CH_3 +Alcoholic O-H, (5643-6187 cm^{-1}) = CH_3, CON-H amide H-bond first overtone and alcoholic O-H first overtone, (6738-7282 cm^{-1}) = CON-H amide free first overtone and (8373-8913 cm^{-1}) = CH aromatic + CH_3 second overtone. All these functional groups are associated with in catechin and theobromine [61] which are major components in cocoa bean polyphenols. Also, the first overtone of O-H and -CH stretching vibration of methyl, methylene, and ethylene are characteristics of functional groups in catechins and epicatechins [8, 62]. In addition, polyphenols correlates with fermented cocoa bean colour [6], i.e. during fermentation polyphenol oxidases converts polyphenols into quinones and these complexes with other polyphenols to give rise to brown colouration [22, 63]. Also, anthocyanins are known to be plant pigments that are responsisble for colour [6].

On the other hand, the ET data provided information on the bitter astringent property of the cocoa beans that could be related to some part of the phenolic compounds especially, flavan-3-ols as it is related to the bitter-astringent properties of cocoa beans [21]. Furthermore, higher concentration of plyphenols was found to contribute to very astringent-tasting chocolate [22]. Also, Bonvehi and Ventura [24] found a correspondence between sensory data and polyphenolic compounds. Therefore, the model based data fusion provided both internal and external attributes that are directly related to total polyphenols contents in cocoa beans.

4. CONCLUSIONS

This work has demonstrated the feasibility of integrating NIR spectroscopy and Electronic tongue technique for an improved prediction of total polyphenols content in cocoa beans. Data fusion of NIRS and ET together with SVMR algorithm could be attempted for other related quality parameters. The overall results have proved that, data fusion (NIRS and ET) technique could improve the efficiency of measuring total polyphenols content in cocoa beans.

Acknowledgements

The authors wish to acknowledge the financial assistance provided by National Natural Science Foundation of China (No. 31071549), Priority Academic Program Development of Jiangsu Higher Education Institutes and University of Cape Coast (AS/86A/V6/1735). We are also grateful to Quality Control Company of Ghana COCOBOD (Q.C.C-Ltd) for their assistance.

REFERENCES

1. K. W. Lee, Y. J. Kim, H. J. Lee and C. Y. Lee, *Journal of Agricultural and Food Chemistry*, 2003, **51**, 7292-7295.
2. A. Di Castelnuovo, R. di Giuseppe, L. Iacoviello and G. de Gaetano, *European Journal of Internal Medicine*, 2012, **23**, 15-25.
3. H. Kim and P. Keeney, *Journal of Food Science*, 2006, **49**, 1090-1092.
4. P. M. Kris-Etherton and C. L. Keen, *Current Opinion in Lipidology*, 2002, **13**, 41-49.
5. J. A. Vinson, Y. Hao, X. Su and L. Zubik, *Journal of Agricultural and Food Chemistry*, 1998, **46**, 3630-3634.
6. J. Wollgast and E. Anklam, *Food Research International*, 2000, **33**, 423-447.
7. R. Nazaruddin, L. K. Seng, O. Hassan and M. Said, *Industrial Crops and Products*, 2006, **24**, 87-94.
8. V. Sinija and H. Mishra, *LWT-Food Science and Technology*, 2009, **42**, 998-1002.
9. L. Escuder-Gilabert and M. Peris, *Analytica Chimica Acta*, 2010, **665**, 15-25.
10. K. Kaffka, K. Norris, F. Kulcsar and I. Draskovits, *Acta Alimentaria*, 1982, **11**, 271-288.
11. J. J. Permanyer and M. L. Perez, *Journal of Food Science*, 1989, **54**, 768-769.
12. A. Vesela, A. S. Barros, A. Synytsya, I. Delgadillo, J. Čopíková and M. A. Coimbra, *Analytica Chimica Acta*, 2007, **601**, 77-86.
13. C. Álvarez, E. Pérez, E. Cros, M. Lares, S. Assemat, R. Boulanger and F. Davrieux, *Journal of Near Infrared Spectroscopy*, 2012, **20**, 307.
14. E. Whitacre, J. O. Liver, R. Van Den Broek, P. Van Engelen, B. K. Remers, B. Van Der Horst, M. S. Tewart and A. Jansen-Beuvink, *Journal of Food Science*, 2003, **68**, 2618-2622.
15. E. Teye, X. Huang, F. Han and F. Botchway, *Food Analytical Methods*, 2014, **7**, 360-365.
16. Q. Chen, J. Zhao and S. Vittayapadung, *Food Research International*, 2008, **41**, 500-504.

17. Q. Chen, J. Zhao, Z. Guo and X. Wang, *Journal of Food Composition and Analysis*, 2010, **23**, 353-358.
18. L. A. Dias, A. M. Peres, A. C. A. Veloso, F. S. Reis, M. Vilas-Boas and A. A. S. C. Machado, *Sensors and Actuators B: Chemical*, 2009, **136**, 209-217.
19. K. Beullens, D. Kirsanov, J. Irudayaraj, A. Rudnitskaya, A. Legin, B. M. Nicolaï and J. Lammertyn, *Sensors and Actuators B: Chemical*, 2006, **116**, 107-115.
20. L. Nørgaard, A. Saudland, J. Wagner, J. P. Nielsen, L. Munck and S. Engelsen, *Applied spectroscopy*, 2000, **54**, 413-419.
21. S. Jinap, P. Dimick and R. Hollender, *Food Control*, 1995, **6**, 105-110.
22. E. O. Afoakwa, in *Chocolate Science and Technology*, John Wiley & Sons, Ltd, 2010, pp. 12-34.
23. Misnawi, S. Jinap, B. Jamilah and S. Nazamid, *Food Quality and Preference*, 2004, **15**, 403-409.
24. J. Serra Bonvehi and F. Ventura Coll, *Food Chemistry*, 1997, **60**, 365-370.
25. T. Romero-Cortes, M. A. Salgado-Cervantes, P. García-Alamilla, M. A. García-Alvarado, G. del C Rodríguez-Jimenes, M. Hidalgo-Morales and V. Robles-Olvera, *Journal of the Science of Food and Agriculture*, 2013, **93**, 2596-2604.
26. J. Dong, D. Zhuang, Y. Huang and J. Fu, *Sensors*, 2009, **9**, 7771-7784.
27. L. Huang, J. Zhao, Q. Chen and Y. Zhang, *Food Chemistry*, 2014, **145**, 228-236.
28. F. Winquist, I. Lundström and P. Wide, *Sensors and Actuators B: Chemical*, 1999, **58**, 512-517.
29. P. A. Ulloa, R. Guerra, A. M. Cavaco, A. M. Rosa da Costa, A. C. Figueira and A. F. Brigas, *Computers and Electronics in Agriculture*, 2013, **94**, 1-11.
30. Z. Wei, J. Wang and W. Liao, *Journal of Food Engineering*, 2009, **94**, 260-266.
31. V. L. Singleton, R. Orthofer and R. M. Lamuela-Raventos, *Methods in enzymology*, 1999, **299**, 152-178.
32. R. Singh, K. Chidambara Murthy and G. Jayaprakasha, *Journal of agricultural and food chemistry*, 2002, **50**, 81-86.
33. V. Singleton and J. A. Rossi, *American journal of Enology and Viticulture*, 1965, **16**, 144-158.
34. D. Sreeramulu and M. Raghunath, *Food Research International*, 2010, **43**, 1017-1020.
35. R. Barnes, M. Dhanoa and S. J. Lister, *Applied Spectroscopy*, 1989, **43**, 772-777.
36. F. Aouidi, N. Dupuy, J. Artaud, S. Roussos, M. Msallem, I. Perraud-Gaime and M. Hamdi, *Food Chemistry*, 2012, **131**, 360-366.
37. Q. Chen, J. Zhao, M. Liu, J. Cai and J. Liu, *Journal of pharmaceutical and biomedical analysis*, 2008, **46**, 568-573.
38. A. Rudnitskaya, D. Kirsanov, A. Legin, K. Beullens, J. Lammertyn, B. M. Nicolaï and J. Irudayaraj, *Sensors and Actuators B: Chemical*, 2006, **116**, 23-28.
39. C. Di Natale, R. Paolesse, A. Macagnano, A. Mantini, A. D'Amico, A. Legin, L. Lvova, A. Rudnitskaya and Y. Vlasov, *Sensors and Actuators B: Chemical*, 2000, **64**, 15-21.
40. A. Barros, M. Safar, M. Devaux, P. Robert, D. Bertrand and D. Rutledge, *Applied spectroscopy*, 1997, **51**, 1384-1393.
41. Z. Haddi, S. Mabrouk, M. Bougrini, K. Tahri, K. Sghaier, H. Barhoumi, N. El Bari, A. Maaref, N. Jaffrezic-Renault and B. Bouchikhi, *Food Chemistry*, 2014, **150**, 246-253.
42. P. Boilot, E. Hines, M. Gongora and R. Folland, *Sensors and Actuators B: Chemical*, 2003, **88**, 80-88.
43. Z. Haddi, H. Alami, N. El Bari, M. Tounsi, H. Barhoumi, A. Maaref, N. Jaffrezic-Renault and B. Bouchikhi, *Food Research International*, 2013.
44. C. Cortes and V. Vapnik, *Machine learning*, 1995, **20**, 273-297.
45. U. Thissen, M. Pepers, B. Üstün, W. Melssen and L. Buydens, *Chemometrics and Intelligent Laboratory Systems*, 2004, **73**, 169-179.

46. L. A. Berrueta, R. M. Alonso-Salces and K. Héberger, *Journal of Chromatography A*, 2007, **1158**, 196-214.

47. H.-T. Lin and C.-J. Lin, *submitted to Neural Computation*, 2003, 1-32.

48. S. S. Keerthi and C.-J. Lin, *Neural computation*, 2003, **15**, 1667-1689.

49. D. Cozzolino, W. Cynkar, N. Shah and P. Smith, *Food Research International*, 2011, **44**, 1888-1896.

50. J. Cai, Q. Chen, X. Wan and J. Zhao, *Food Chemistry*, 2011, **126**, 1354-1360.

51. J. Zhao, Q. Chen, X. Huang and C. H. Fang, *Journal of Pharmaceutical and Biomedical analysis*, 2006, **41**, 1198-1204.

52. M. Dong and N. Wang, *Applied Mathematical Modelling*, 2011, **35**, 1024-1035.

53. J. Wu, J. Mei, S. Wen, S. Liao, J. Chen and Y. Shen, *Journal of Computational Chemistry*, 2010, **31**, 1956-1968.

54. M. Zhang, J. Luypaert, J. Fernández Pierna, Q. Xu and D. Massart, *Talanta*, 2004, **62**, 25-35.

55. N. d. Ortega, M.-P. Romero, A. Macià, J. Reguant, N. Anglès, J.-R. n. Morelló and M.-J. Motilva, *Journal of agricultural and food chemistry*, 2008, **56**, 9621-9627.

56. C. da Silva Oliveira, L. F. Maciel, M. S. Miranda and E. da Silva Bispo, *British Food Journal*, 2011, **113**, 1094-1102.

57. F. Han, X. Huang, E. Teye, F. Gu and H. Gu, *Analytical Methods*, 2014, **6**, 529-536.

58. J. Clapperton, S. Yow, J. Chan, D. Lim, R. Lockwood, L. Romanczyk and J. Hammerstone, *Tropical agriculture*, 1994, **71**, 303-308.

59. K. B. Miller, D. A. Stuart, N. L. Smith, C. Y. Lee, N. L. McHale, J. A. Flanagan, B. Ou and W. J. Hurst, *Journal of Agricultural and Food Chemistry*, 2006, **54**, 4062-4068.

60. E. Teye, X. Huang, H. Dai and Q. Chen, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2013, **114**, 183-189.

61. A. Bedini, V. Zanolli, S. Zanardi, U. Bersellini, E. Dalcanale and M. Suman, *Food Analytical Methods*, 2013, **6**, 17-27.

62. A. M. M. Jalil and A. Ismail, *Molecules*, 2008, **13**, 2190-2219.

63. N. Camu, T. De Winter, S. K. Addo, J. S. Takrama, H. Bernaert and L. De Vuyst, *Journal of the Science of Food and Agriculture*, 2008, **88**, 2288-2297.

Table Caption

Table 1 Chemical measurements of polyphenols in the calibration and prediction set

Table 2 Comparison of SVMR models of TPC of cocoa beans by different sensors

Figure Caption

Fig. 1 Selection and combination of optimum variables (data fusion of NIRS-ET)

Fig. 2 Scatter plots between predicted values and the reference measured values in the calibration set: (A) model based on ET and (B) model based on NIRS

Fig. 3 Reference measured versus ET-NIRS prediction of polyphenol in cocoa bean (A) calibration set and (B) prediction set by SVMR

Table 1.0 Chemical measurements of polyphenols in the calibration and prediction set

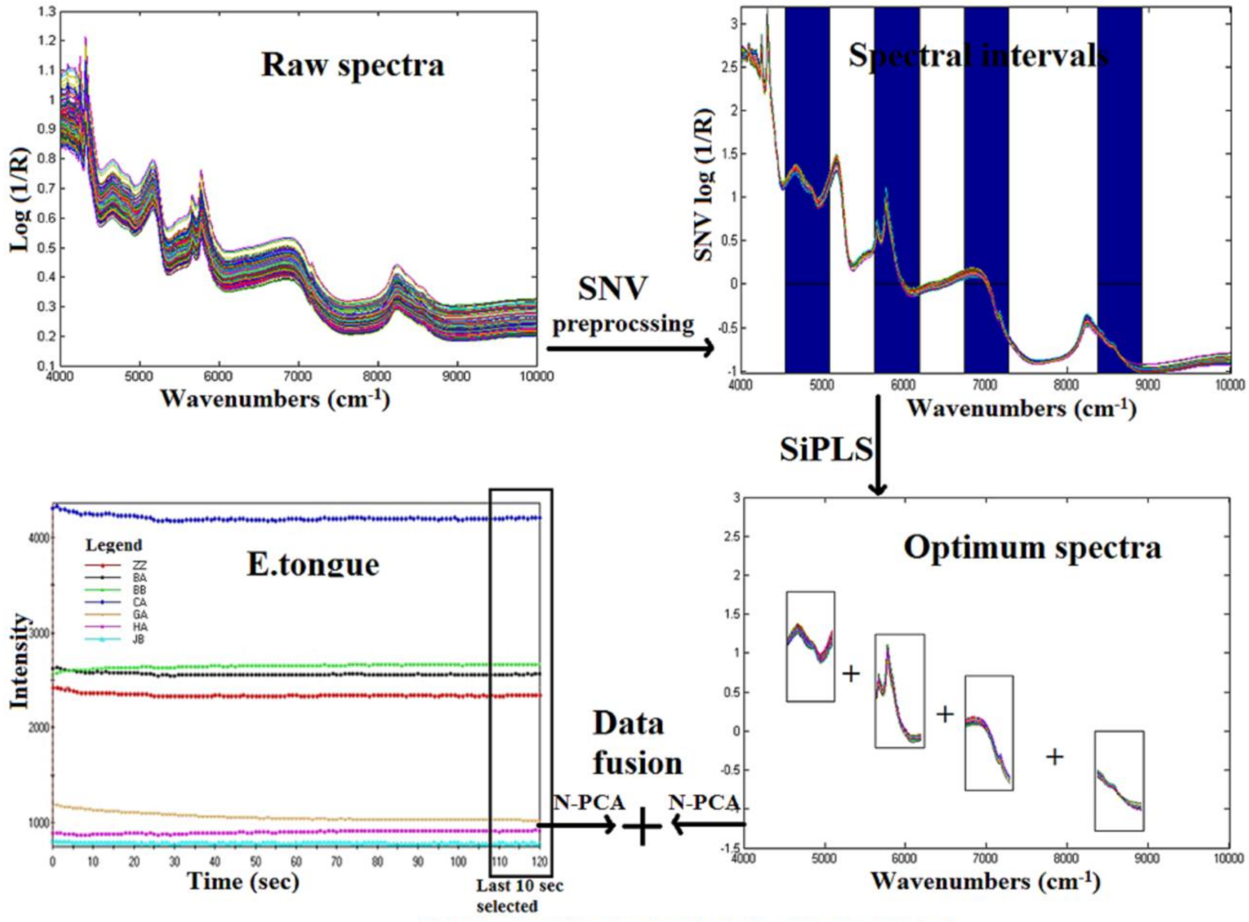
Sets	Units (%)	Subsets	SN	Range	Mean	Stdv
Calibration	g/g	Calibration	80	23.02–33.67	28.43	2.28
Prediction	g/g	Prediction	30	23.53–33.67	28.85	2.20

SN: Number of samples, Stdv; standard deviation

Table 2.0 Comparison of SVMR models of TPC of cocoa beans by different sensors

Models	*Vs	*PCs	Calibration set			Prediction set		
			R_{cal}	RMSECV (g/g)	Bias	R_{pre}	RMSEP (g/g)	Bias
ET data	7	5	0.813	1.346	0.312	0.706	1.796	0.564
NIRS data	83	9	0.920	1.148	0.152	0.917	1.164	0.276
Data fusion	90	7	0.987	0.890	0.006	0.982	0.900	0.013

*Vs; variables, *PCs; principal components



N-PCA; normalization in principal component analysis

Fig. 1 Selection and combination of optimum variables

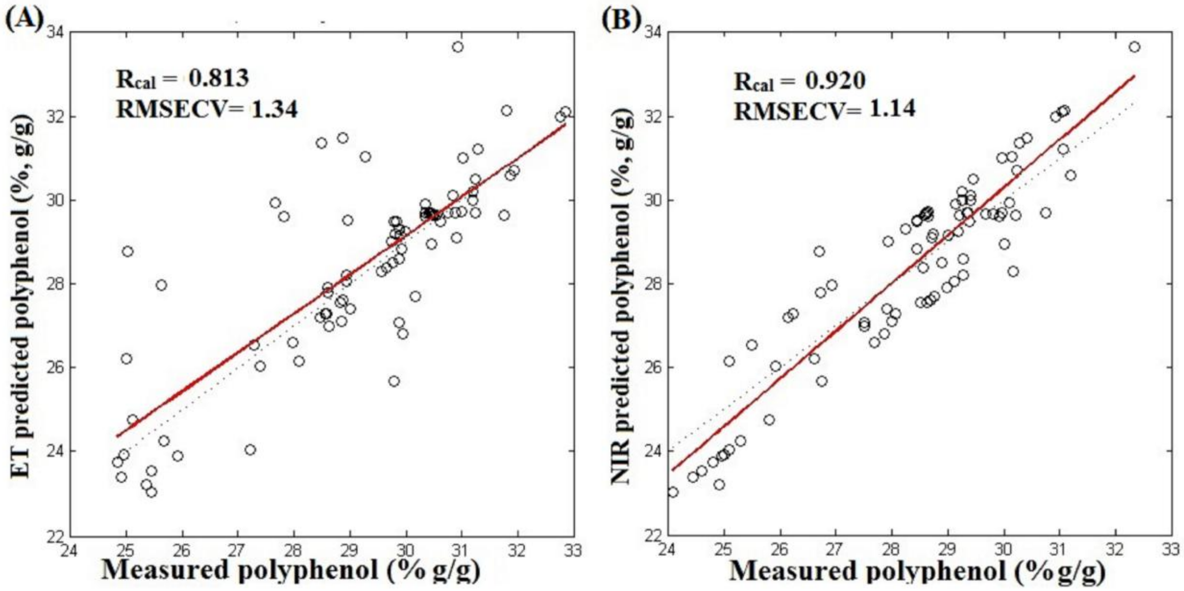


Fig. 2 Scatter plots between predicted values and the reference measurements values in the calibration set; (A) model based on ET and (B) model based on NIRS

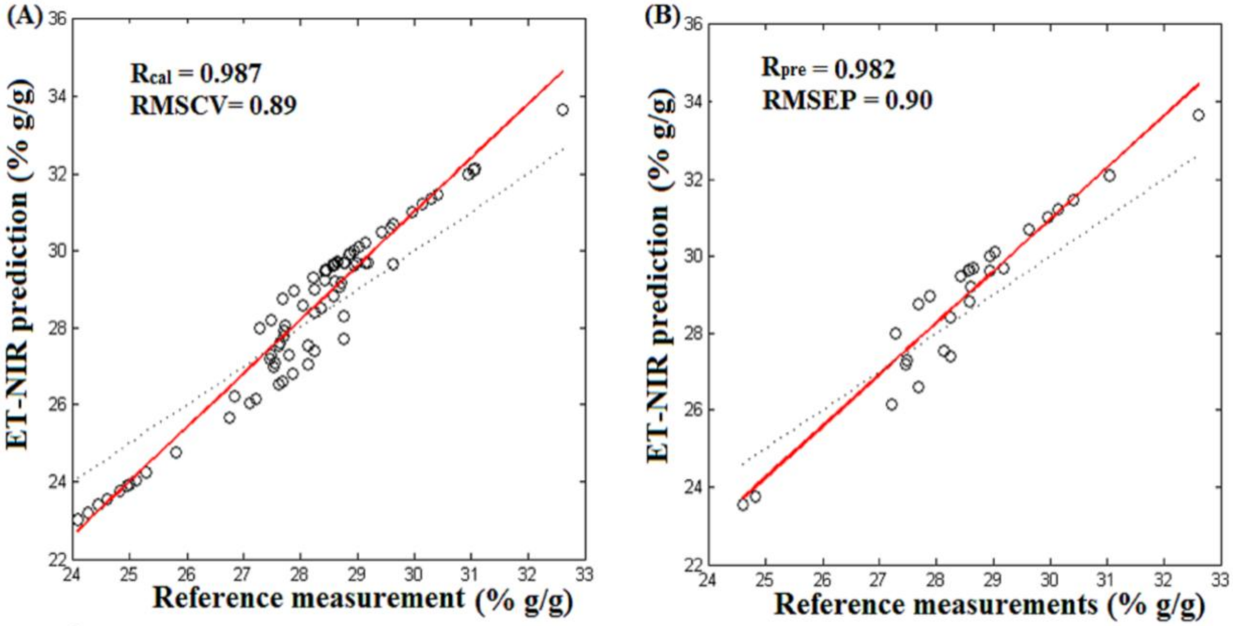


Fig. 3 Reference measured versus ET-NIR prediction of polyphenol in cocoa bean (A) calibration set and (B) prediction set by SVMR