

Classifying green teas with near infrared hyperspectral imaging

Puneet Mishra¹, Alison Nordon^{1*}

¹*WestCHEM, Department of Pure and Applied Chemistry and Centre for Process Analytics and Control Technology, University of Strathclyde, 295 Cathedral Street, Glasgow, G1 1XL, United Kingdom*

Keywords: quality, non-destructive, origin, multivariate

Email : puneet.mishra@wur.nl , *alison.nordon@strath.ac.uk

Abstract

Tea products analysis is currently limited to high-end analytical techniques such as high-performance liquid chromatography (HPLC), gas chromatography (GC) and isotope analysis. However, these techniques are time-consuming, expensive, destructive and require trained experts to perform the experiments. In the present work, an application of NIRS for classification of similar appearing green tea products is demonstrated. The tea products were classified based on their origin and utilising support vector machine classifier. Results showed good accuracy ($96.36 \pm 0.17\%$) to perform the classification of green tea products with seven different countries of origin.

Introduction

Tea is the most widely consumed manufactured drink in the world. Most of the tea products available in the market are obtained through processing of hand plucked fresh leaves of the plant *Camellia sinensis*. Based on the extent of fermentation, tea products are mainly categorised into three main classes: green (unfermented), oolong (semi-fermented), and black (fermented). In recent years, there has been an increase in the awareness of the potential health benefits of teas. Of the different tea products, the minimally processed green tea has attracted the most attention as it contains natural polyphenolic antioxidants. A wide range of health benefits of bioactive polyphenols have been reported including potential anti-cancer effects, influence on the human brain function, protection from environmental toxins, anti-angiogenesis effects and managing oxidative stress [1].

In the globalised tea market, there is an increasing demand for high-quality green tea products of specific geographical origin. High-quality green tea products are mainly characterised by the flavour that they impart, which involves two primary sensory perceptions, i.e., taste and aroma. The distinct taste and aroma of any tea product are linked to its geographical origin [2]. At present, tea manufacturers utilise sensory analysis by human experts for discrimination of green tea products. Sensory analysis involves assessment of tea products by appearance, colour, aroma and taste, along with the overall quality of the samples. However, distinguishing tea products based on sensory analysis is a time-consuming, expensive task and cannot be implemented online as it requires an expert human panel. Furthermore, sensory analysis is subjective, and it can be inconsistent and unpredictable owing to physiological and psychological differences between tasters [3][4]. Different high-end analytical techniques such as high-performance liquid chromatography (HPLC), gas chromatography (GC) and isotope analysis have also been explored for assessment of tea products. However, these techniques are time-consuming, expensive, destructive and require trained experts to perform the experiments. Two other popular techniques for the assessment of green tea products, which mimic the taste and odour assessment by the human test panel, are the e-nose and the e-tongue.

However, the e-nose and the e-tongue utilise non-specific sensor arrays and so fail to discriminate between similar complex mixtures [3]. Further, both sensors are affected by humidity and temperature [5].

In recent years, there has been increasing interest in the use of optical spectroscopic techniques for rapid, non-destructive assessment of food products. Near infrared (NIR) spectroscopy is particularly attractive for this purpose, as NIR spectral profiles can provide indirect access to different quality parameters for tea via qualitative and quantitative analysis. NIR hyperspectral imaging (HSI) combines NIR spectroscopy and imaging. The aim of this work is to demonstrate the use of NIR-HSI for classification of green tea samples, which visually appear similar, based on geographic origin.

Samples and measurements

Sixteen green tea samples differing in geographical origin were obtained from Unilever, Colworth Park, UK. The 16 samples originated from 7 different countries: Argentina (1), India (5), Sri Lanka (2), China (2), Japan (2), Kenya (3) and Sumatra (1). Imaging was performed with a push-broom line scan HSI camera (*Model name*: RedEye 1.7) from INNO-SPEC (Nurnberg, Germany). The camera has an InGaAs sensor and generates a spatial map of 320 x 256 pixels in the spectral range of 950 - 1760 nm. The pixel size was 30 x 30 μm^2 and the spectral resolution was 3.2 nm. The lighting was provided by two halogen light sources 50 W each and the integration time used was 300 ms. Imaging was performed by placing the samples over the translation stage which was controlled by an independent stage motor connected to the computer system (Zolix TSA 200 BF).

Data analysis

Radiometric calibration was performed using dark and white reference images acquired along with the samples. The effects of light scattering were removed utilising the standard normal variate (SNV) transform. In SNV, the mean and standard deviation of each spectrum for each pixel are calculated, the mean is subtracted, and the standard deviation is used to normalise the difference. This transformation normalises each spectrum to zero mean and unit standard deviation. Before applying the SNV transform, the spectral range was reduced from 950 - 1760 nm to 967 nm - 1700 nm, to remove the noisy regions at the edges of the spectral range, and converted to absorbance. Further, the spectral absorbance profiles were smoothed with a Savitzky-Golay filter. PLS Toolbox (version 8.11, Eigenvector Research Inc., USA) and Matlab (R2016b, Mathworks, USA) was used for data analysis. Classification was performed utilising support vector machines (SVMs). SVMs are supervised non-probabilistic learning models that utilise hyperplanes to define the decision boundaries for performing classification. A total of 400 spectra from each class were used for the modelling as explained in [6]. SVM algorithms are usually developed to perform a binary classification, however, SVMs can be used for multi-class classification problems by utilising several independent binary classifiers. This can be performed by combining it with ensemble methods such as error correcting output codes (ECOC). The ECOC deals with the multi-class classification problem by converting it into several independent binary classification problems. In the present case a 7 class classifier was developed.

Results

Figure 1 (a) presents the SNV pre-processed mean absorbance spectral profiles of the green tea samples corresponding to 7 countries of origin. The spectra were pre-processed to reduce the effects of light scattering originating from the inhomogeneous surface of the tea products as recommended in [7]. In Figure 1(a), it can be noted that variability is present at various regions over the complete range for the NIR spectra recorded. Such variability can be noted higher at specific locations compared to others. To understand this variability and the specific NIR regions, the standard deviation was calculated for each wavelength from the mean spectra of samples belonging to different countries of origin as plotted in Figure 1(b). The standard deviation plot highlights the wavelength regions with higher variability in the mean spectra of tea products.

In Figure 1(b), various peaks (depicted by the green vertical lines) can be identified at representative wavelengths. In previous works, the peaks at 1131, 1654 and 1666 nm were found to be representative of the total tea polyphenols, 1361 nm is representative of moisture content, 1093 – 1121 nm for thearubigin components of TRS, 1492 nm corresponds to free amino acids, 1176 nm is a second overtone C-H and 1390 nm for the CH₂ overtone [8][9][10].

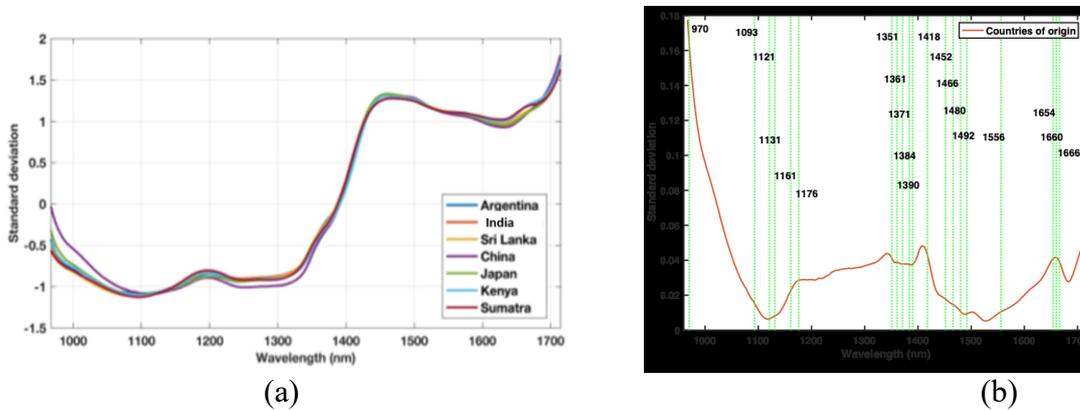


Figure 1: (a). Mean SNV pre-processed absorbance spectral profiles of tea samples corresponding to 7 countries of origin, and (b). Standard deviation calculated for NIR spectra of tea products of different countries of origin (red). The vertical green lines denote the assignment of wavelengths to different components present in tea products.

Figure 2(a) presents the classification maps obtained from the 10-fold cross-validated multiclass SVM-ECOC models for the seven class problems explaining classification based on country of origin. The classification accuracy obtained was $96.36 \pm 0.17\%$. The confusion matrix in Figure 2(b) shows the classification accuracies for individual classes obtained from the SVM model. The presence of misclassification can be observed in Figure 2(a) for all the tea samples, especially near the interface of the samples and containers. Further to that, there is also misclassification in samples such those from China and Japan, which were misclassified with each other. This could arise if there is high variability in the samples.

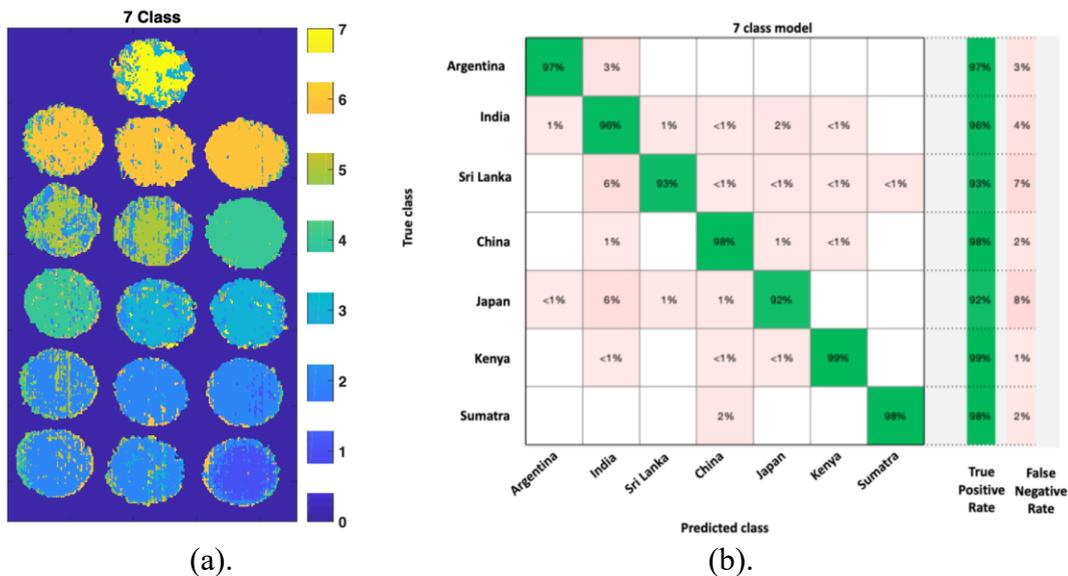


Figure 2: (a). Classification maps for the 16 green tea samples obtained using classification models corresponding to country of origin. Class 0 is the background Class 1-7 represents the sample from Argentina, India, Sri-Lanka, China, Japan, Kenya and Sumatra respectively. (b). Confusion matrix corresponding to the classification model.

Conclusions

NIR HSI can be used as a non-destructive tool for classification of different green tea products, which visually appear to be similar. The SVM-ECOC model provided a classification accuracy of $96.36 \pm 0.17\%$ for identifying the country of origin of the green teas. However, there was also misclassification between the samples, which could arise from variability in the samples and the similarity of the NIR spectra for different products. Therefore, experiments should be performed incorporating various factors responsible for the variability in the samples, and robust models should be developed capable of dealing with these variabilities. In conclusion, NIR HSI in conjunction with machine learning could be a potential tool for on-line assessment of green tea products and the monitoring of tea fermentation processes.

Acknowledgments

This work has received funding from the European Union's Horizon 2020 research and innovation programme named 'MODLIFE' (Advancing Modelling for Process-Product Innovation, Optimization, Monitoring and Control in Life Science Industries) under the Marie Skłodowska-Curie grant agreement number 675251.

References

- [1]. S. Bansal, S. Choudhary, M. Sharma, S.S. Kumar, S. Lohan, V. Bhardwaj, N. Syan, S. Jyoti, Tea: A native source of antimicrobial agents, *Food Res. Int.* 53 (2013) 568–584.
- [2]. Q. Chen, J. Zhao, H. Lin, Study on discrimination of Roast green tea (*Camellia sinensis* L.) according to geographical origin by FT-NIR spectroscopy and supervised pattern recognition, *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* 72 (2009) 845–850.

- [3]. D. Huo, Y. Wu, M. Yang, H. Fa, X. Luo, C. Hou, Discrimination of Chinese green tea according to varieties and grade levels using artificial nose and tongue based on colorimetric sensor arrays, *Food Chem.* 145 (2014) 639–645.
- [4]. W. He, X. Hu, L. Zhao, X. Liao, Y. Zhang, M. Zhang, J. Wu, Evaluation of Chinese tea by the electronic tongue: Correlation with sensory properties and classification according to geographical origin and grade level, *Food Res. Int.* 42 (2009) 1462–1467.
- [5]. E.A. Baldwin, J. Bai, A. Plotto, S. Dea, Electronic Noses and Tongues: Applications for the Food and Pharmaceutical Industries, *Sensors (Basel)*. 11 (2011) 4744–4766.
- [6]. P. Mishra, A. Nordon, M. S. Mohd Asaari, G. Lian, S. Redfern, Fusing spectral and textural information in near-infrared hyperspectral imaging to improve green tea classification modelling, *J. Food Eng.* 249 (2019) 40-47.
- [6]. P. Mishra, A. Nordon, J. Tschannerl, G. Lian, S. Redfern, S. Marshall, Near-infrared hyperspectral imaging for non-destructive classification of commercial tea products, *J. Food Eng.* 238 (2018) 70-77.
- [7]. Q. Chen, J. Zhao, X. Huang, H. Zhang, M. Liu, Simultaneous determination of total polyphenols and caffeine contents of green tea by near-infrared reflectance spectroscopy, *Microchem. J.* 83 (2006) 42–47.
- [8]. M. Bian, A.K. Skidmore, M. Schlerf, T. Wang, Y. Liu, R. Zeng, T. Fei, Predicting foliar biochemistry of tea (*Camellia sinensis*) using reflectance spectra measured at powder, leaf and canopy levels, *ISPRS J. Photogramm. Remote Sens.* 78 (2013) 148–156.
- [9]. S. M. Tan, R. M. Luo, Y. P. Zhou, H. Gong, Z. Tan, Rapid and non-destructive discrimination of tea varieties by near infrared diffuse reflection spectroscopy coupled with classification and regression trees, *African J. Biotechnol.* 11 (2012), 2303–2312.