Near-infrared hyperspectral imaging for non-destructive classification of commercial tea products

Puneet Mishra[1], Alison Nordon[1], Julius Tschannerl[2], Guoping Lian[3,4], Sally Redfern[3], Stephen Marshall[2]

[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*

[2] *Hyperspectral Imaging Centre, Department of Electronic and Electrical Engineering, University of Strathclyde, 204 George Street, Glasgow, G1 1XW, United Kingdom*

[3] *Unilever R&D Colworth, Colworth House, Sharnbrook, Bedford MK44 1LQ, United Kingdom*

[4] *Department of Chemical and Process Engineering, University of Surrey, Guildford GU2 7XH, United Kingdom*

Corresponding authors email: puneet.mishra@strath.ac.uk , alison.nordon@strath.ac.uk

Abstract

Tea is the most consumed manufactured drink in the world. In recent years, various high end analytical techniques such as high-performance liquid chromatography have been used to analyse tea products. However, these techniques require complex sample preparation, are time consuming, expensive and require a skilled analyst to carry out the experiments. Therefore, to support rapid and non-destructive assessment of tea products, the use of near infrared (NIR) (950-1760 nm) hyperspectral imaging (HSI) for classification of six different commercial tea products (oolong, green, yellow, white, black and Pu-erh) is presented. To visualise the HSI data, linear (principal component analysis (PCA) and multidimensional scaling (MDS)) and non-linear (t-distributed stochastic neighbour embedding (t-SNE) and isometric mapping (ISOMAP)) data visualisation methods were compared. t-SNE provided separation of the six commercial tea products into three groups based on the extent of processing: minimally processed, oxidised and fermented. To perform the classification of different tea products, a multi-class error-correcting output code (ECOC) model containing support vector machine (SVM) binary learners was developed. The classification model was further used to predict classes for pixels in the HSI hypercube to obtain the classification maps. The SVM-ECOC model provided a classification accuracy of 97.41±0.16 % for the six commercial tea products. The methodology developed provides a means for rapid, non-destructive, *in situ* testing of tea products, which would be of considerable benefit for process monitoring, quality control, authenticity and adulteration detection.

Keywords: Imaging spectroscopy, hypercube, multivariate, data visualisation, neighbourhood methods

1. Introduction

Being the oldest beverage, tea is the most consumed drink in the world (Sang, 2016). Different tea products exist due to different processes for freshly harvested tea leaves (Lv et al., 2013). There are six main types of tea products, i.e. oolong, green, yellow, white, black and Pu-erh (Chang, 2015), which differ in terms of processing (see Figure 1). Green, yellow and white tea products undergo minimal processing, oolong and black tea products have been oxidised while Pu-erh tea has been fermented. The chemical composition of fresh tea (*Camellia sinensis*) leaves is a complex mixture of caffeine, polyphenols, polysaccharides and nutrients such as protein, amino acids, lipids, and vitamins (Ruan et al., 2010). Typically, during the processing of fresh tea leaves, such as oxidation and fermentation, they undergo chemical compositional changes. Free amino acids, total tea polyphenols, soluble sugars, and caffeine are the four major chemical components that determine the nature and quality of the final tea products (Ozturk et al., 2016).

Figure 1: Processing steps for different tea products starting from fresh green tea leaves to final products.

Analytical methods used to measure chemical constituents as quality indicators of plant-based products include high-performance liquid chromatography (HPLC) (Nieh et al., 2009), liquid chromatography/mass spectrometry (LC/MS) (Tan et al., 2016), gas chromatography/mass spectrometry (GC/MS) (Jing et al., 2017) and electrochemical systems (Kumar et al., 2016) (Domínguez et al., 2015). However, these methods have complex sample preparation, are time consuming, expensive and require a skilled analyst to carry out the experiments (Li et al., 2017). A non-destructive technique that has been used for analysis of tea processes and quality monitoring is e-nose (Yaroshenko et al., 2014) (Sharma et al., 2015). E-nose devices usually include an array of metal oxide sensors which respond to the amount of biochemical volatiles coming into contact with the corresponding sensor surface to explain the chemical profile (Bhattacharyya et al., 2007). However, a major disadvantage of e-nose sensors is that they are affected by environmental conditions such as temperature and humidity, which leads to sensor drift (Baldwin et al., 2011).

In recent years, there has been increasing interest in the use of optical spectroscopic techniques for rapid, non-destructive assessment of food products. NIR spectroscopy is particularly attractive for this purpose, where changes in the NIR spectral profiles can be correlated to perform qualitative and quantitative analysis of food products (Qu et al., 2015; Fu and Ying, 2016). NIR spectroscopy has been explored for discrimination (He et al., 2007; Chen et al., 2009), identification (Chen et al., 2007; Wang et al., 2015) and quality assessment (Panigrahi et al., 2016) of tea products. Also reported for non-destructive tea analysis are emerging studies utilising imaging techniques for the identification (Chen et al., 2008), classification (Wang et al., 2015) and for evaluation of sensory quality (Zhu et al., 2017) of tea products. Integration of spectroscopy and imaging is known as hyperspectral imaging (HSI) and use of NIR-HSI still seems unexplored in its application to the analysis of tea products.

HSI has been widely used in remote sensing for military applications (Goetz et al., 1985), but it is now popular in scientific domains such as forensics (Edelman et al., 2012), medical (Lu et al., 2014), food (Pu et al., 2015), pharmaceutical (Kandpal et al., 2016) and plants (Mishra et al., 2017). There are reports of the use of HSI for the understanding of different food products such as coffee (Nansen et al., 2016), tobacco (Garcia-Allende et al., 2008), and seeds of vegetable and fruits (Shrestha et al., 2016; Kandpal et al., 2016). Some applications of HSI of tea have been reported but these studies only considered a single variety of tea and measured the visible and very near infrared (VNIR) range (around 400-1000 nm), which is dominated by the pigments and physical characteristics of the samples (Zhao et al., 2009; Xie et al., 2015). In comparison to the VNIR region, the NIR region provides more detailed chemical information such as overtones resulting from the molecular vibration of O-H, C-H, N-H bonds and their combinations, which can support a better classification system based on the chemistry of the samples (Mishra et al., 2016).

The aim of the present work is to demonstrate the use of NIR (950-1760 nm) HSI for rapid, non-destructive classification of six different commercial tea products (oolong, green, yellow, white, black and Pu-erh). The study investigates and compares four different dimensionality reduction techniques (linear and non-linear) to visualise the high dimensional HSI tea data. Furthermore, multi-class support vector machine (SVM) modelling has been performed to generate spatial classification maps of tea products.

2. Materials and Methods

2.1. Samples

Six commercial tea samples were obtained from the local market (Glasgow, United Kingdom). The samples were obtained in airtight sealed packaging and stored at ambient temperature. All samples of tea were in loose-leaf form. Black, green and white tea were from Vahdam Teas (New Delhi, India), oolong tea was from Yamamotoyama (California, USA), Pu-erh tea was from The Tea Makers of London (London, United Kingdom) and yellow tea was of an unspecified Chinese origin. The six tea products can also be broadly grouped as minimally processed (green, white and yellow), oxidised (black and oolong tea) and fermented (Pu-erh tea). The samples for each imaging experiment were transferred on the day of analysis into a black plastic circular container (diameter = 3.3 cm, depth = 1.3 cm). A different cap was used for each tea to avoid any cross-contamination.

2.2. Hyperspectral imaging measurements

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 camera communicated with the computer via a gigabit Ethernet connection. 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). The speed of the translation stage was optimised before image acquisition to avoid any distortion in the shape of the image arising from the overlapping of the spectral information in the adjacent pixels. The image acquisition and management of settings (integration time) were performed using the software interface called SiCAP provided with the camera by INNO-SPEC. Images were first acquired of six different tea samples placed adjacent to each other in their respective sample containers in the field of view of the camera. An image was then acquired of black, Pu-erh and oolong teas where each tea occupied approximately a third of the volume of the sample container; the teas were not physically mixed. Finally, equal proportions of all six tea samples were mixed, by manually shaking the different tea products in a container, and an image of the mixture was acquired. One image was acquired of each sample, with each image comprising more than 2000 pixels (spectra) for the individual tea samples and more than 11200 pixels for the samples containing more than one type of tea. An illustration of the HSI setup configured for imaging of tea samples can be found in Figure 2.

Figure 2: Illustrative diagram for the hyperspectral imaging setup used to acquire the images of tea samples.

2.3. Data analysis

2.3.1. Pre-processing of HSI data

The data cubes not only contain information about the samples imaged but also consist of different unwanted influences in signal resulting from factors such as illumination intensity, the detector sensitivity and transmission properties of the optics. The effects resulting from these factors are both wavelength dependent and independent. To correct for these effects, radiometric calibration was performed using dark and white reference images acquired along with the samples. The correction was performed for every pixel in the HS image according to equation 1:

$$\boldsymbol{I}_{R(i,j)} = \frac{\boldsymbol{I}_{raw(i,j)} - \boldsymbol{I}_{dark(i,j)}}{\boldsymbol{I}_{white(i,j)} - \boldsymbol{I}_{dark(i,j)}} \qquad (1)$$

where, $\boldsymbol{I}_{R}$ is the calibrated reflectance image, $\boldsymbol{I}_{raw}$ is the raw intensity image measured from the test sample, $\boldsymbol{I}_{dark}$ is the intensity of the dark response, $\boldsymbol{I}_{white}$ is the intensity for the uniform white reference and $\boldsymbol{i}$ and $\boldsymbol{j}$ were spatial coordinates over the image.

Often, the radiometric correction is sufficient to remove the effects of illumination inhomogeneity from the spectral data, however, when the sample surfaces are not uniform, as in the case of samples of loose tea leaves, the light scattering during diffuse reflection causes additive and multiplicative effects (Mishra et al., 2016). These scattering effects lead to baseline shifts in the spectrum and variation in the global intensity, which is again dependent on the wavelength. Standard normal variate (SNV) is a very common technique used in NIR spectroscopy to remove these effects (Barnes et al., 1989). 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 (15-point width and second order polynomial) (Savitzky and Golay, 1964). The *savgol* and *snv* functions from PLS toolbox (version 8.11, Eigenvector Research Inc., USA) were used. All visualisation and classification analysis was performed on the pre-processed spectra. The pre-processed pure spectra of six pure tea samples were extracted using Matlab’s (R2016b, Mathworks, USA) *roipoly* function. The *roipoly* function provides a graphical user interface in Matlab to extract the information from each image over the manually selected locations.

2.3.2. Principal Component Analysis

Principal component analysis (PCA) introduced by Pearson in 1901 belongs to the family of linear methods for visualising high dimension data (Wold et al., 1987). In PCA, a set of observations containing correlated variables is orthogonally transformed to linearly uncorrelated variables defined as principal components (PCs). In PCA, the transformation is performed to retain the major amount of variability in the dataset.

The PCA decomposition model for a given observation data matrix X can be understood as equation 2:

$$\boldsymbol{X = TW^T} \qquad (2)$$

where T is the score in the lower dimension explained by the number of PCs specified and W is a *p* × *p (p denotes number of variables)* matrix whose columns are the eigenvectors of $X^T X$.

In the case of dimensionality reduction, the aim is to preserve the maximum amount of meaningful variation present in the dataset. The extracted PCs define a new orthonormal basis set which can be used to transform the data from a high dimension space to the lower space explained by the PCs. PCA from a dimensionality reduction perspective can be understood as minimising the squared reconstruction error as given in equation 3.

$$min||TW^T - T_rW_r^T||^2 \qquad (3)$$

where, *TW* and T_rW_r are the reconstructed original dataset in higher and lower dimensional space respectively. Minimisation of the reconstruction error results in the maximisation of the information that was present in the higher dimensional space when defined in the lower dimensional space given by the significant number of PCs. To interpret the data in two or three dimensional plots, the respective PCs can be selected and used for transformation to the orthogonal axes represented by the PCs. Transformation from a higher dimension to a lower dimension can be performed as in equation 4.

$$\hat{X}_r = XW \qquad (4)$$

2.3.3. Multi-Dimensional Scaling

Multi-dimensional scaling (MDS) is a linear method for visualising high dimensional data (Cox et al., 2000). MDS performs a transformation by preserving the between object distances from the higher dimension to lower dimension. The MDS utilises calculation of the Euclidean distances for each data point in the multidimensional space to capture the pattern. The distances are defined as a symmetric distance matrix (*D*). MDS attempts to find data points in a specified (d-dimensional) space such that the Euclidean distance between data points ($D\hat{}$) is similar to the distance in higher dimensional space. The minimisation function can be understood as equation 5:

$$\min \sum_i \sum_j ||d_{ij} - \hat{d_{ij}}||^2 \qquad (5)$$

where, $D = d_{ij} = // x_i{-}x_j //^{2}$and $\hat{D} = d_{ij} = // y_i{-}\ y_j //^{2}$ explaining the Euclidean distance between points in high (x_i, x_j) and low dimensional space (y_i, y_j), respectively. *i, j* denotes specific position of point.

2.3.4. Isometric Mapping

Isometric mapping (ISOMAP) belongs to the family of non-linear techniques for visualising high dimensional data (Tenenbaum, 1998; Balasubramanian and Schwartz, 2002). ISOMAP can be understood as a generalised non-linear form of MDS which utilises the geodesic space accounting for the non-linearity in the high dimensional data manifold. The geodesic distance is defined as the shortest distance between two data points on a curved surface of a non-linear manifold. As a first step, ISOMAP approximates a neighbourhood graph by identifying k nearest neighbours (kNNs) or selecting neighbourhood data points based on any other condition for every data points. The geodesic distance is then approximated for all the pairs of data points on the neighbourhood graph. Finally, the distance data obtained from the graph is embedded to a lower dimension Euclidean space using MDS as shown in equation (6).

$$\min \sum_{i} \sum_{j} \|\boldsymbol{D_G} - \boldsymbol{D_E}\|^2 \qquad (6)$$

where, $\boldsymbol{D_G}$ and $\boldsymbol{D_E}$ explaining the geodesic and Euclidean distance between points in high and low dimensional space, respectively.

2.3.5. t-Distributed Stochastic Neighbour Embedding

t-distributed stochastic neighbour embedding (t-SNE) is a non-linear technique used to visualise high dimensional data in two or three dimensional scatter plots (Maaten and HInton, 2008). The main objective of t-SNE is to model the similar points using nearby points (small pairwise distance) and the dissimilar points using distant points (large pairwise distances). As a first step, to represent the similarity, the t-SNE converts high-dimensional Euclidean distances between data points into conditional probabilities using a Gaussian distribution. The joint probability for a data point $\boldsymbol{x_j}$ to $\boldsymbol{x_i}$ can be calculated with equation (7):

$$p_{j|i} = \frac{\exp(-||x_i - x_j||^2/2\sigma_i^2)}{\sum_k \sum_{k\neq l} \exp(-||x_k - x_l||^2/2\sigma_i^2)}, \qquad (7)$$

The conditional probability represents the probability that $\boldsymbol{x_i}$ will pick $\boldsymbol{x_j}$ as a neighbour based on the proportion of probability density under a Gaussian centred at $\boldsymbol{x_i}$. If the points are near then the value of $\boldsymbol{p_{i|j}}$ will be higher compare to the points far away. Furthermore, the conditional probabilities are symmetrised to reduce the effects of outliers by setting (8):

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N} \qquad (8)$$

To represent joint probabilities in the low dimensional map $\boldsymbol{q_{ij}}$, t-SNE utilises a heavy tailed Student t-distribution. The benefit of using a heavy tailed distribution is that it makes the joint probabilities invariant to changes in the scale of the map. The joint probabilities $\boldsymbol{q_{ij}}$ can be estimated by (9):

$$q_{ij} = \frac{(1 + ||y_i - y_j||^2)^{-1}}{\sum_k \sum_{k \neq l}(1 + ||y_k - y_l||^2)^{-1}}, \qquad (9)$$

Finally, the t-SNE minimises a single Kullback-Leibler (*KL*) divergence between a joint probability distribution, P, in the high-dimensional space and a joint probability distribution, Q, in the low-dimensional space as can be understood from equation (10):

$$KL(P||Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \qquad (10)$$

The minimisation of the KL divergence is performed using a gradient descent algorithm with respect to the locations of the points in the map $\boldsymbol{y_i}$.

All the data visualisation methods (PCA, MDS, ISOMAP and t-SNE) were implemented in Matlab using the Toolbox for Dimensionality Reduction (https://lvdmaaten.github.io/drtoolbox/) (Maaten et al., 2009; Maaten and Hinton, 2008). The Mahalanobis distance (Mahalanobis, 1936) was used to assess the separation of the clusters identified with the different data visualisation methods.

2.3.6. Support vector machines for multi-class classification

Support vector machines (SVMs) are supervised non-probabilistic learning models which utilise hyperplanes to define the decision boundaries for performing classification (Vapnik and Vapnik, 1998). The SVM algorithms are usually developed to perform a binary classification, however, SVM 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. A wide range of applications of SVM to process HS images can be seen (García Allende et al., 2008) (Mountrakis et al., 2011).

In the present work, the ECOC-SVM algorithm available in Matlab's Statistics and Machine Learning Toolbox (R2016b) was implemented to perform the classification utilising the classification learner application. ECOC-SVM uses a one-versus-all coding design, in which for each binary learner one class is assigned a positive value and all others are assigned negative values. To map the data to the higher dimension, a radial basis function (RBF) kernel (scale parameter=10) was used. The RBF kernel has the benefit of non-linearly mapping the sample to the higher dimensional space for dealing with a non-linear relationship between observation and classes. For every pure tea sample, spectra (967 - 1700 nm) were extracted from 200 pixels, which were selected at random from the image collected, leading to 1200 spectra in total for calibration of the classification model. Validation of the model was performed with a 10-fold cross-validation method. Furthermore, to have confidence in the model accuracy, the model was recalibrated with 1200 iterations and the mean and standard deviation were noted. The trained classifier was further used to generate the classification maps of the HS images. The HSI cubes were first unfolded from a 3D map ($\boldsymbol{n}\times \boldsymbol{p}\times \boldsymbol{k}$) to a 2D matrix ($\boldsymbol{np}\times \boldsymbol{k}$) and then the class of every row of the matrix (representing the pixel) was predicted, where n, p, k defined the x, y and z dimension of data. After prediction, the matrix ($\boldsymbol{np} \times \mathbf{1}$) was reshaped to the original image dimension ($\boldsymbol{n} \times \boldsymbol{p}$).

3. Results

3.1. Spectral profiles of tea samples

Figure 3: Absorbance spectra of pure tea samples of yellow, oolong, green, black, white and Pu-erh. (a). Mean absorbance spectra (n = 200). (b) Mean spectra after pre-processing (SNV and Savitzky- Golay smoothing), and (c) standard deviation of the absorbance spectra and spectra after pre- processing. The vertical green lines denote the positions of the main peaks.

Figure 3 presents the spectral profiles of individual tea samples. Figure 3(a) presents the mean absorbance spectra calculated from the 200 spectra extracted for each of the six tea samples (yellow, oolong, green, black, white and Pu-erh), Figure 3(b) presents the mean spectra after pre-processing with Savitzky-Golay filtering followed by SNV, and Figure 3(c) presents the standard deviation of the spectra before and after pre-processing. From Figure 3(a), it can be seen that the absorbance spectra of different tea samples contain scattering effects leading to baseline shifts. These effects can also be seen in the standard deviation plot in Figure 3(c) for the absorbance spectra (red), where the standard deviation over the entire spectral range is approximately constant. These scattering effects can bias modelling of the data, therefore, they were removed via pre-processing. In Figure 3(b), it can be seen that after pre-processing, differences in spectra at various wavelengths have emerged, and so spectral differences corresponding to different teas can be noted. Scattering effects arise in the imaging experiments as the inhomogeneity in the size of the loose leaves does not get compensated for by the flat surface of the white reflectance standard used for radiometric calibration.

In Figure 3(c), it can be noted that the pre-processing reveals the spectral variation arising from differences in the tea, which was previously dominated by the effects of light scattering. In Figure 3(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 (Chen et al., 2006; Bian et al., 2010; Bian et al., 2013), 1361 nm is representative of moisture content (Panigrahi et al., 2016), 1093-1121 nm for thearubigin components of TRS1 (Panigrahi et al., 2016), 1492 nm corresponds to free amino acids (Bian et al., 2010), 1176 nm is a second overtone C-H (Tan et al., 2012) and 1390 nm for the CH_2 overtone (Lee et al., 2014).

3.2. Visualising high dimensional data

Figure 4: 2-Dimensional scatter plots for visualising high dimensional tea data. (a). Principal Component Analysis (PCA), (b). Multidimensional Scaling (MDS), (c). Isometric Mapping (ISOMAP), and (d). t-distributed Stochastic Neighbour Embedding (t-SNE). In all the plots, the first dimension is represented in the x-axis and the second in the y-axis, and the six tea products are coloured as follows: Pu-erh (pink), black (sky blue), oolong (yellow), green (green), white (blue) and yellow (red).

To visualise the high dimensional data in the lower dimension, the 256-dimensional HSI data were transformed to 2-dimensional plots using PCA, MDS, ISOMAP and t-SNE as shown in Figure 4. It can be seen clearly in Figure 4 that the t-SNE (Figure 4(d)) outperforms PCA, MDS and ISOMAP (Figures 4(a), 4(b) and 4(c), respectively) regarding identification of the maximum number of separate clusters. These separate clusters correspond to different tea products and their representation as separate clusters in the plots signifies that the visualisation method is able to preserve the structure of the data on transformation from a high dimensional space to a lower dimensional space. In general, all the methods were able to separate the Pu-erh tea (pink) from all other tea samples. The reason for this can be seen in Figure 3(b) where Pu-erh tea (sky blue) has a very different spectral signature compared to the other tea samples. This is likely to be because the

Pu-erh tea undergoes very different processing, which includes microbial fermentation of sun dried leaves (Lv et al., 2013), compared to the other teas.

It can be seen in Figure 4((a), (b) and (c)) that with the exception of Pu-erh tea, all other types of tea samples are mixed and their clear distinction is not possible. In comparison, black and oolong tea are identified as separate clusters with t-SNE. However, while t-SNE was not able to separate the green, yellow and white tea, it still provided better separation of these three teas as shown in Figure 4(d). Green, yellow and white teas appear in the same cluster as they have similar spectral signatures (see Figure 3(b)). This may arise from the fact that these teas are most similar in terms of processing conditions; they are subjected to either limited or no oxidation. In comparison, oolong and black teas undergo oxidation during their manufacturing. This may be why these two teas lie in two adjacent clusters that are far away from the cluster containing green, yellow and white teas. However, further information is required to identify the exact source of the spectral differences observed.

Figure 5: Mahalanobis distances between the three different cluster groups obtained using PCA (dark-blue), MDS (sky-blue), ISOMAP (light- green) and t-SNE (yellow).

To assess further the separation of clusters with each method, the Mahalanobis distance between the clusters was calculated. Figure 5 presents the Mahalanobis distance estimated for the three major clusters identified in Figure 4. The three major cluster can be understood as the group of minimally processed tea products available on the market (denoted the green group), the teas subjected to oxidation (oxidised group) and those that have been subjected to microbial fermented (fermented group). The x-axis in Figure 5 presents the pairwise groups used for estimating the distance and the y-axis gives the respective Mahalanobis distance obtained from the different data visualisation methods. It can be seen that the t-SNE (yellow) was superior to all other methods followed by the ISOMAP (light green), and then PCA (dark blue) and MDS (sky blue) for separating all three groups in the data-visualisation plots.

From a statistical perspective, a better visualisation of separate clusters corresponding to different tea products with t-SNE could be due to its ability to capture the non-linearity present in the data set and consideration of neighbourhood information. This supports the modelling of both distant and nearby points (Maaten and Hinton, 2008). Often, in high dimensional space when the data lies near, or in a non-linear manifold, linear methods like PCA and MDS fail to preserve the structure of data in the lower dimension space. This is because with linear methods like PCA and MDS, the aim is to keep the distant object far apart; no consideration is given to utilising the information about the neighbouring data points (Maaten et al., 2009).

It can be seen in Figure 4(c) that ISOMAP provides a little insight on differences in the classes belonging to black and oolong teas compared to what was achieved with PCA (Figure 4(a)) and MDS (Figure 4(b)). However, ISOMAP was not able to provide a clear separation of the two teas as was obtained with t-SNE. A reason for the poor performance of ISOMAP compared to t-SNE could be due to its weakness in dealing with the holes and non-convex nature of the data manifold in the higher dimension (Tenenbaum, 1998). Another important weakness of ISOMAP is its topological instability, which leads to a short-circuiting problem in the neighbourhood graph and results in its poor performance (Balasubramanian and Schwartz, 2002).

3.3. Support vector machine classification

Figure 6: (a) Greyscale image constructed from the spectral plane extracted from the hypercube at 1424 nm, (b) Classification maps obtained from the application of the ECOC-SVM model. From left to right the samples can be understood as yellow (dark blue), oolong (light blue), green (cyan), black (light green), white (orange) and Pu-erh (yellow). (c) Histograms showing the proportion of pixels attributed to the different tea products for the classification maps in (b).

The results from the application of the ECOC-SVM multi-class classification model are presented as classification maps in Figures 6 and 7. Figure 6(b) presents the classification maps of pure tea samples, from left to right, the samples can be understood as yellow, oolong, green, black, white and Pu-erh. For comparison, a greyscale image was also produced (Figure 6(a)) using the spectral plane corresponding to 1424 nm; this wavelength was selected merely to allow visualisation of the data hypercube. It can be seen from Figure 6(b) that all six teas were classified into their respective individual classes. However, there are some pixels that were misclassified; Figure 6(c) shows the proportion of pixels attributed to the different tea products for the classification maps in Figure 6(b). The misclassification was most dominant at the edges owing to signal from the circular sample container; such pixels (approximately 20%) were misclassified as Pu-erh. When these pixels were excluded, an overall accuracy of 97.41±0.16 % was obtained for cross-validated samples using 1200 iterations.

Apart from the edges, a reason for the misclassification between different teas can be attributed to their spectral similarity. When visualising the data with t-SNE (see Figure 4(d)), green, white and yellow tea were found to be lying near in the same cluster, and black and oolong were near to each other due to their spectral similarity. Hence, the classification map for the yellow tea (dark blue) has some misclassified pixels that have been attributed to either white (orange) or green tea (cyan). For black and oolong teas, it can be noted that there are some pixels in the classification map for black tea (light green) that were misclassified as oolong (light blue class) and vice-versa. Another possible reason for misclassification could arise from the purity of the tea; for example, a minimally processed tea (e.g. white) may contain small amounts of oxidised product (e.g. black tea).

Figure 7: (a). Greyscale image at 1424 nm for the sample comprising oolong, black and Pu-erh tea, (b). The classification map for the sample comprising oolong, black and Pu-erh tea, (c). Pie chart representing the proportion of pixels belonging to a particular class for the classification map presented in (b), (d). Greyscale image at 1424 nm for a sample containing a mixture of all teas, (e). The classification map for a sample containing a mixture of all teas, and (f) Pie chart representing the proportion of pixels belonging to a particular class for the classification map presented in (e).

Figure 7 presents the classification maps for the HS images acquired for samples comprising mixtures of teas. This analysis was performed to assess the feasibility of using the methodology developed to classify different tea samples when more than one tea is present. Figure 7(a) presents the spectral plane corresponding to 1424 nm for a sample containing oolong, black and Pu-erh teas (not mixed) in roughly equal portions. These three teas were selected as there is an oxidation stage in their manufacturing. The location of the different teas in Figure 7(a) can be identified with the red markers. As can be seen from Figure 7(b), the model provided a clear classification of the three teas into their respective classes. However, some misclassification can be seen at the interface between different types of teas; individual pixels will detect the presence of more than one type of tea at these locations. Furthermore, Figure 7(c) provides insight into the proportion of pixels belonging to each class. It can be seen that the pie chart is mainly dominated by the proportion of oolong, black and Pu-erh tea and contains a very small portion (<1 %) of pixels classified as green, white and yellow.

The methodology developed was also tested for a mixture of all six tea samples. The result for classification of the sample containing a mixture of all six types of tea is presented in Figure 7(e). The classification map shown in figure 7(e) can be interpreted in conjunction with the pie chart (Figure 7(f)) representing the proportion of pixels classified belonging to different classes. The pie chart shows that the presence of all the classes can be detected with the classification model and the portion of each type of tea ranged from 10 – 26%. However, it was not possible to validate the classification result of the mixture image because it is not known if the sample was a homogenous mixture of the six types of teas and hence, the exact composition of the upper surface of the sample is unknown. In addition, there may be some misclassification of pixels that detect more than one type of tea.

4. Conclusions

NIR HSI has been used to classify six different types of commercial tea samples. Before any data modelling, the spectral imaging data from tea products should be pre-processed to reduce the effects of light scattering arising from the inhomogeneous and uneven leaf surface. Four different types of linear and non-linear dimensionality reduction methods were compared for visualisation of imaging data. The non-linear method, t-SNE, gave better separation of the different tea products than classical linear techniques such as PCA and MDS. This is because t-SNE uses information from neighbouring data points in the high dimensional space to preserve the structure in the low dimensional representation. It was possible to classify the tea according to product type using a ECOC-SVM multi-class classification model constructed using the NIR HSI data. Therefore, NIR HSI in conjunction with machine learning could be a potential tool for classification of different types of tea products. The source of spectral differences is assumed to arise from the different processing steps that are involved in the manufacture of various types of tea. However, there could be other sources, e.g. geographical, that contribute to spectral differences and hence, this requires further investigation.

5. 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 Sklodowska-Curie grant agreement number 675251.

6. References

Balasubramanian, M., Schwartz, E.L., 2002. The Isomap Algorithm and Topological Stability. Science (80). 295, 7 LP –7.

Baldwin, E.A., Bai, J., Plotto, A., Dea, S., 2011. Electronic Noses and Tongues: Applications for the Food and Pharmaceutical Industries. Sensors (Basel). 11(5), 4744-4766. https://doi.org/10.3390/s110504744

Barnes, R.J., Dhanoa, M.S., Lister, S.J., 1989. Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. Appl. Spectrosc. 43, 772–777. https://doi.org/10.1366/0003702894202201

Bhattacharyya, N., Seth, S., Tudu, B., Tamuly, P., Jana, A., Ghosh, D., Bandyopadhyay, R., Bhuyan, M., 2007. Monitoring of black tea fermentation process using electronic nose. J. Food Eng. 80, 1146–1156. https://doi.org/https://doi.org/10.1016/j.jfoodeng.2006.09.006

Bian, B.M., Skidmore, A.K., Schlerf, M., Fei, T., Liu, Y.F., Wang, T., 2010. Reflectance spectroscopy of biochemical components as indicators of tea, Camellia Sinensis, quality 76, 1385–1392.

Bian, M., Skidmore, A.K., Schlerf, M., Wang, T., Liu, Y., Zeng, R., Fei, T., 2013. Predicting foliar biochemistry of tea (Camellia sinensis) using reflectance spectra measured at powder, leaf and canopy levels. ISPRS J. Photogramm. Remote Sens. 78, 148–156. https://doi.org/https://doi.org/10.1016/j.isprsjprs.2013.02.002

Chang, K, 2015. World tea production and trade: current and future development, Food and Agriculture Organisation (FAO), United Nations (UN), Rome. http://www.fao.org/3/a-i4480e.pdf (last accessed: 14 September 2017)

Chen, Q., Zhao, J., Cai, J., 2008. Identification of Tea Varieties Using Computer Vision. Transactions of the ASABE. 51(2), 623-628. https://doi.org/10.13031/2013.24363

Chen, Q., Zhao, J., Fang, C.H., Wang, D., 2007. Feasibility study on identification of green, black and Oolong teas using near-infrared reflectance spectroscopy based on support vector machine (SVM). Spectrochim. Acta Part A Mol. Biomol. Spectrosc. 66, 568–574. https://doi.org/https://doi.org/10.1016/j.saa.2006.03.038

Chen, Q., Zhao, J., Huang, X., Zhang, H., Liu, M., 2006. Simultaneous determination of total polyphenols and caffeine contents of green tea by near-infrared reflectance spectroscopy. Microchem. J. 83, 42–47. https://doi.org/http://dx.doi.org/10.1016/j.microc.2006.01.023

Chen, Q., Zhao, J., Lin, H., 2009. 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, 845–850. https://doi.org/http://dx.doi.org/10.1016/j.saa.2008.12.002

Cox, T.F., Cox, M.A.A., 2000. Multidimensional scaling. CRC press.

Domínguez, I., Doménech-Carbó, A., 2015. Screening and authentication of tea varieties based on microextraction-assisted voltammetry of microparticles. Sensors Actuators B Chem. 210, 491–499. https://doi.org/https://doi.org/10.1016/j.snb.2015.01.009

Edelman, G.J., Gaston, E., van Leeuwen, T.G., Cullen, P.J., Aalders, M.C.G., 2012. Hyperspectral imaging for non-contact analysis of forensic traces. Forensic Sci. Int. 223, 28–39. https://doi.org/https://doi.org/10.1016/j.forsciint.2012.09.012

Fu, X., Ying, Y., 2016. Food Safety Evaluation Based on Near Infrared Spectroscopy and Imaging: A Review. Crit. Rev. Food Sci. Nutr. 56, 1913–1924. https://doi.org/10.1080/10408398.2013.807418

García Allende, P.B., Anabitarte García, F., Conde Portilla, O.M., Mirapeix Serrano, J.M., Madruga Saavedra, F.J., López Higuera, J.M., 2008. Support vector machines in hyperspectral imaging spectroscopy with application to material identification. Proc. SPIE 6966, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIV, 69661V. http://dx.doi.org/10.1117/12.770306

Garcia-Allende, P.B., Conde, O.M., Mirapeix, J., Cobo, A., Lopez-Higuera, J.M., 2008. Quality control of industrial processes by combining a hyperspectral sensor and Fisher's linear discriminant analysis. Sensors Actuators B Chem. 129, 977–984. https://doi.org/https://doi.org/10.1016/j.snb.2007.09.036

Goetz, A.F., Vane, G., Solomon, J.E., Rock, B.N., 1985. Imaging spectrometry for Earth remote sensing. Science 228, 1147–1153. https://doi.org/10.1126/science.228.4704.1147

He, Y., Li, X., Deng, X., 2007. Discrimination of varieties of tea using near infrared spectroscopy by principal component analysis and BP model. J. Food Eng. 79, 1238–1242. https://doi.org/https://doi.org/10.1016/j.jfoodeng.2006.04.042

Jing, J., Shi, Y., Zhang, Q., Wang, J., Ruan, J., 2017. Prediction of Chinese green tea ranking by metabolite profiling using ultra-performance liquid chromatography–quadrupole time-of-flight mass spectrometry (UPLC–Q-TOF/MS). Food Chem. 221, 311–316. https://doi.org/https://doi.org/10.1016/j.foodchem.2016.10.068

Kandpal, L.M., Lohumi, S., Kim, M.S., Kang, J.-S., Cho, B.-K., 2016. Near-infrared hyperspectral imaging system coupled with multivariate methods to predict viability and vigor in muskmelon seeds. Sensors Actuators B Chem. 229, 534–544. https://doi.org/https://doi.org/10.1016/j.snb.2016.02.015

Kandpal, L.M., Tewari, J., Gopinathan, N., Boulas, P., Cho, B.-K., 2016. In-Process Control Assay of Pharmaceutical Microtablets Using Hyperspectral Imaging Coupled with Multivariate Analysis. Anal. Chem. 88, 11055–11061. https://doi.org/10.1021/acs.analchem.6b02969

Kumar, A.S., Shanmugam, R., Nellaiappan, S., Thangaraj, R., 2016. Tea quality assessment by analyzing key polyphenolic functional groups using flow injection analysis coupled with a dual electrochemical detector. Sensors Actuators B Chem. 227, 352–361. https://doi.org/https://doi.org/10.1016/j.snb.2015.12.072

Lee, M.-S., Hwang, Y.-S., Lee, J., Choung, M.-G., 2014. The characterization of caffeine and nine individual catechins in the leaves of green tea (Camellia sinensis L.) by near-infrared reflectance spectroscopy. Food Chem. 158, 351–357. https://doi.org/10.1016/j.foodchem.2014.02.127

Li, J., Fu, B., Huo, D., Hou, C., Yang, M., Shen, C., Luo, H., Yang, P., 2017. Discrimination of Chinese teas according to major amino acid composition by a colorimetric {IDA} sensor. Sensors Actuators B Chem. 240, 770–778. https://doi.org/http://dx.doi.org/10.1016/j.snb.2016.09.019

Lu, G., Fei, B., 2014. Medical hyperspectral imaging: a review. J. Biomed. Opt. 19, 10901. https://doi.org/10.1117/1.JBO.19.1.010901

Lv, H., Zhang, Y., Lin, Z., Liang, Y., 2013. Processing and chemical constituents of Pu-erh tea: A review. Food Res. Int. 53, 608–618. https://doi.org/https://doi.org/10.1016/j.foodres.2013.02.043

Maaten, L. van der, Hinton, G., 2008. Visualizing high-dimensional data using t-SNE. J. Mach. Learn. Res. 9, 2579–2605.

Maaten, L. van der, Postma, E., Van den Herik, J., 2009. Dimensionality reduction: a comparative review. Tilburg University Technical Report, TICC-TR 2009-005.

Mahalanobis, P.C., 1936. On the generalised distance in statistics, in: Proceedings National Institute of Science, India. pp. 49–55.

Mishra, P., Asaari, M.S.M., Herrero-Langreo, A., Lohumi, S., Diezma, B., Scheunders, P., 2017. Close range hyperspectral imaging of plants: A review. Biosyst. Eng. 164, 49–67. https://doi.org/https://doi.org/10.1016/j.biosystemseng.2017.09.009

Mishra, P., Cordella, C.B.Y., Rutledge, D.N., Barreiro, P., Roger, J.M., Diezma, B., 2016. Application of independent components analysis with the JADE algorithm and NIR hyperspectral imaging for revealing food adulteration. J. Food Eng. 168, 7–15.

Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: A review. ISPRS J. Photogramm. Remote Sens. 66, 247–259. https://doi.org/https://doi.org/10.1016/j.isprsjprs.2010.11.001

Nansen, C., Singh, K., Mian, A., Allison, B.J., Simmons, C.W., 2016. Using hyperspectral imaging to characterize consistency of coffee brands and their respective roasting classes. J. Food Eng. 190, 34–39. https://doi.org/https://doi.org/10.1016/j.jfoodeng.2016.06.010

Nieh, C.-H., Hsieh, B.-C., Chen, P.-C., Hsiao, H.-Y., Cheng, T.-J., Chen, R.L.C., 2009. Potentiometric flow-injection estimation of tea fermentation degree. Sensors Actuators B Chem. 136, 541–545. https://doi.org/https://doi.org/10.1016/j.snb.2008.09.024

Ozturk, B., Seyhan, F., Ozdemir, I.S., Karadeniz, B., Bahar, B., Ertas, E., Ilgaz, S., 2016. Change of enzyme activity and quality during the processing of Turkish green tea. LWT - Food Sci. Technol. 65, 318–324. https://doi.org/https://doi.org/10.1016/j.lwt.2015.07.068

Panigrahi, N., Bhol, C.S., Das, B.S., 2016. Rapid assessment of black tea quality using diffuse reflectance spectroscopy. J. Food Eng. 190, 101–108. https://doi.org/https://doi.org/10.1016/j.jfoodeng.2016.06.020

Plaza, A., Benediktsson, J.A., Boardman, J.W., Brazile, J., Bruzzone, L., Camps-Valls, G., Chanussot, J., Fauvel, M., Gamba, P., Gualtieri, A., Marconcini, M., Tilton, J.C., Trianni, G., 2009. Recent advances in techniques for hyperspectral image processing. Remote Sens. Environ. 113, S110–S122. https://doi.org/https://doi.org/10.1016/j.rse.2007.07.028

Pu, Y.-Y., Feng, Y.-Z., Sun, D.-W., 2015. Recent Progress of Hyperspectral Imaging on Quality and Safety Inspection of Fruits and Vegetables: A Review. Compr. Rev. Food Sci. Food Saf. 14, 176–188. https://doi.org/10.1111/1541-4337.12123

Qu, J.-H., Liu, D., Cheng, J.-H., Sun, D.-W., Ma, J., Pu, H., Zeng, X.-A., 2015. Applications of Near-infrared Spectroscopy in Food Safety Evaluation and Control: A Review of Recent Research Advances. Crit. Rev. Food Sci. Nutr. 55, 1939–1954. https://doi.org/10.1080/10408398.2013.871693

Ruan, J., Haerdter, R., Gerendás, J., 2010. Impact of nitrogen supply on carbon/nitrogen allocation: a case study on amino acids and catechins in green tea [Camellia sinensis (L.) O. Kuntze] plants*. Plant Biol. 12, 724–734. https://doi.org/10.1111/j.1438-8677.2009.00288.x

Sabale, S.P., Jadhav, C.R., 2015. Hyperspectral Image Classification Methods in Remote Sensing - A Review, in: 2015 International Conference on Computing Communication Control and Automation. pp. 679–683. https://doi.org/10.1109/ICCUBEA.2015.139

Sang, S., 2016. Tea: Chemistry and Processing BT - Encyclopedia of Food and Health. Academic Press, Oxford, pp. 268–272. https://doi.org/https://doi.org/10.1016/B978-0-12-384947-2.00685-1

Savitzky, A., Golay, M.J.E., 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Anal. Chem. 36, 1627–1639. https://doi.org/10.1021/ac60214a047

Sharma, P., Ghosh, A., Tudu, B., Sabhapondit, S., Baruah, B.D., Tamuly, P., Bhattacharyya, N., Bandyopadhyay, R., 2015. Monitoring the fermentation process of black tea using QCM sensor based electronic nose. Sensors Actuators B Chem. 219, 146–157. https://doi.org/https://doi.org/10.1016/j.snb.2015.05.013

Shrestha, S., Knapič, M., Žibrat, U., Deleuran, L.C., Gislum, R., 2016. Single seed near-infrared hyperspectral imaging in determining tomato (Solanum lycopersicum L.) seed quality in association with multivariate data analysis. Sensors Actuators B Chem. 237, 1027–1034. https://doi.org/https://doi.org/10.1016/j.snb.2016.08.170

Tan, J., Dai, W., Lu, M., Lv, H., Guo, L., Zhang, Y., Zhu, Y., Peng, Q., Lin, Z., 2016. Study of the dynamic changes in the non-volatile chemical constituents of black tea during fermentation processing by a non-targeted metabolomics approach. Food Res. Int. 79, 106–113. https://doi.org/https://doi.org/10.1016/j.foodres.2015.11.018

Tan, S.-M., Luo, R.-M., Zhou, Y.-P., Gong, H., Tan, Z., 2012. Rapid and non-destructive discrimination of tea varieties by near infrared diffuse reflection spectroscopy coupled with classification and regression trees. African J. Biotechnol. 11, 2303–2312.

Tenenbaum, J.B., 1998. Mapping a Manifold of Perceptual Observations, in: Jordan, M.I., Kearns, M.J., Solla, S.A. (Eds.), Advances in Neural Information Processing Systems 10. MIT Press, pp. 682–688.

Vapnik, V.N., Vapnik, V., 1998. Statistical learning theory. Wiley New York.

Wang, S., Yang, X., Zhang, Y., Phillips, P., Yang, J., Yuan, T.-F., 2015. Identification of Green, Oolong and Black Teas in China via Wavelet Packet Entropy and Fuzzy Support Vector Machine. Entropy. 17, 6663-6682. https://doi.org/10.3390/e17106663

Wang, X., Huang, J., Fan, W., Lu, H., 2015. Identification of green tea varieties and fast quantification of total polyphenols by near-infrared spectroscopy and ultraviolet-visible spectroscopy with chemometric algorithms. Anal. Methods 7, 787–792. https://doi.org/10.1039/C4AY02106A

Wold, S., Esbensen, K., Geladi, P., 1987. Principal component analysis. Chemom. Intell. Lab. Syst. 2, 37–52. https://doi.org/https://doi.org/10.1016/0169-7439(87)80084-9

Xie, C., Li, X., Shao, Y., He, Y., 2015. Color Measurement of Tea Leaves at Different Drying Periods Using Hyperspectral Imaging Technique. PLoS One 9, 1–15. https://doi.org/10.1371/journal.pone.0113422

Yaroshenko, I., Kirsanov, D., Kartsova, L., Bhattacharyya, N., Sarkar, S., Legin, A., 2014. On the application of simple matrix methods for electronic tongue data processing: Case study with black tea samples. Sensors Actuators B Chem. 191, 67–74. https://doi.org/https://doi.org/10.1016/j.snb.2013.09.093

Zhao, J., Chen, Q., Cai, J., Ouyang, Q., 2009. Automated tea quality classification by hyperspectral imaging. Appl. Opt. 48, 3557–3564. https://doi.org/10.1364/AO.48.003557

Zhu, H., Ye, Y., He, H., Dong, C., 2017. Evaluation of green tea sensory quality via process characteristics and image information. Food Bioprod. Process. 102, 116–122. https://doi.org/https://doi.org/10.1016/j.fbp.2016.12.004

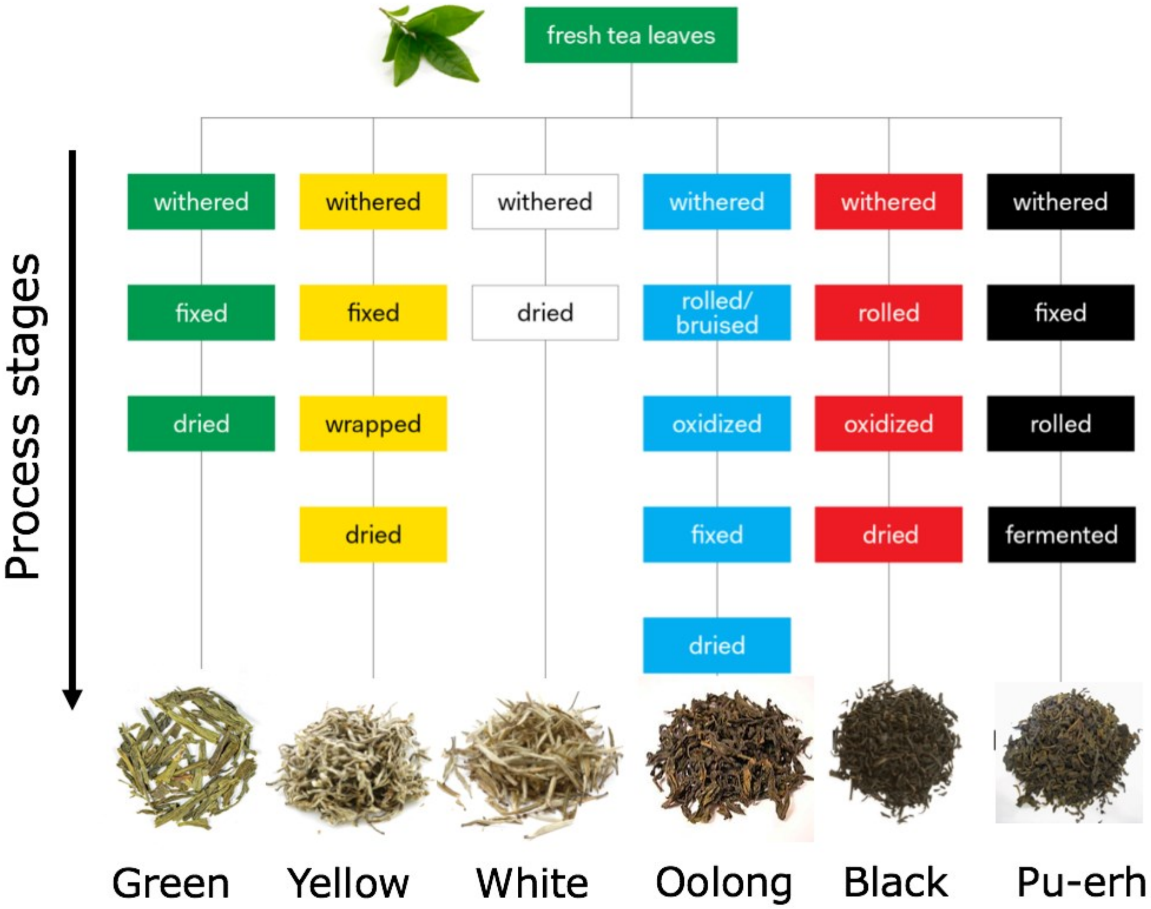


Figure 2: Processing steps for different tea products starting from fresh green tea leaves to final products.

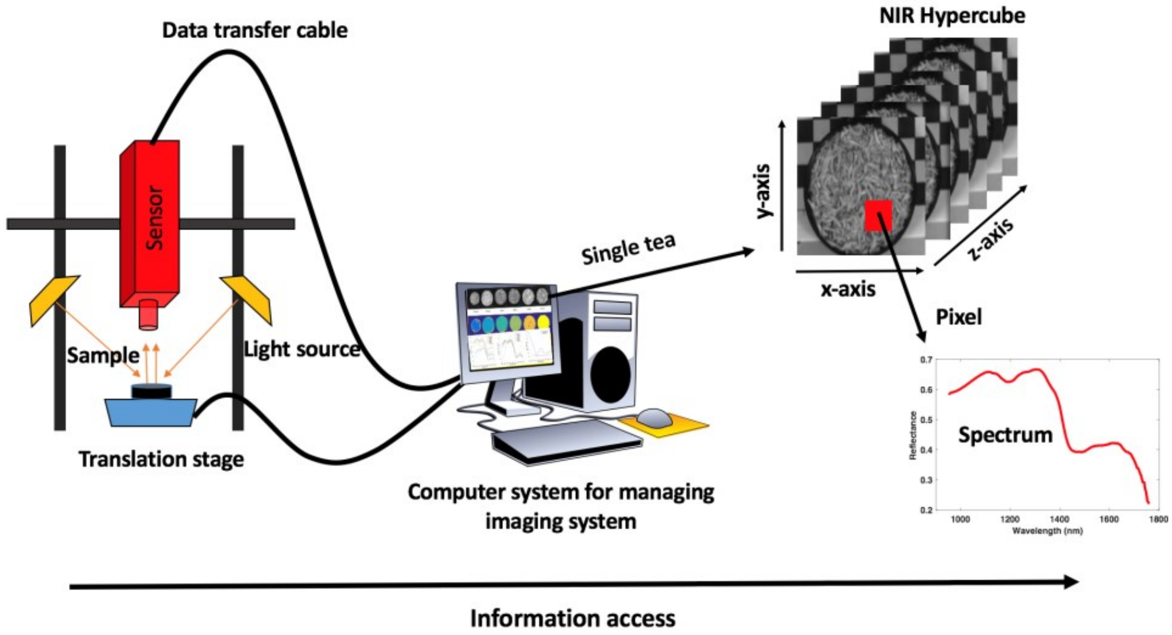
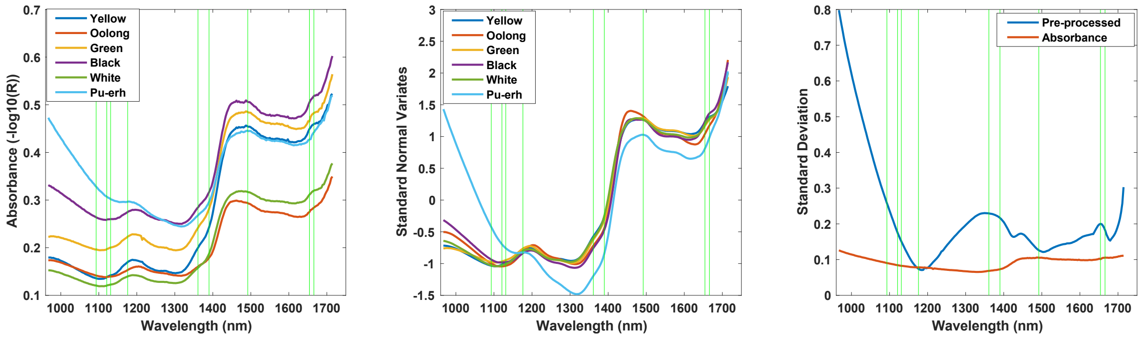


Figure 2: Illustrative diagram for the hyperspectral imaging setup used to acquire the images of tea samples.



(a) ***(b)*** ***(c)***

Figure 3: Absorbance spectra of pure tea samples of yellow, oolong, green, black, white and Pu-erh. (a). Mean absorbance spectra (n = 200). (b) Mean spectra after pre-processing (SNV and Savitzky- Golay smoothing), and (c) standard deviation of the absorbance spectra and spectra after pre- processing. The vertical green lines denote the positions of the main peaks.

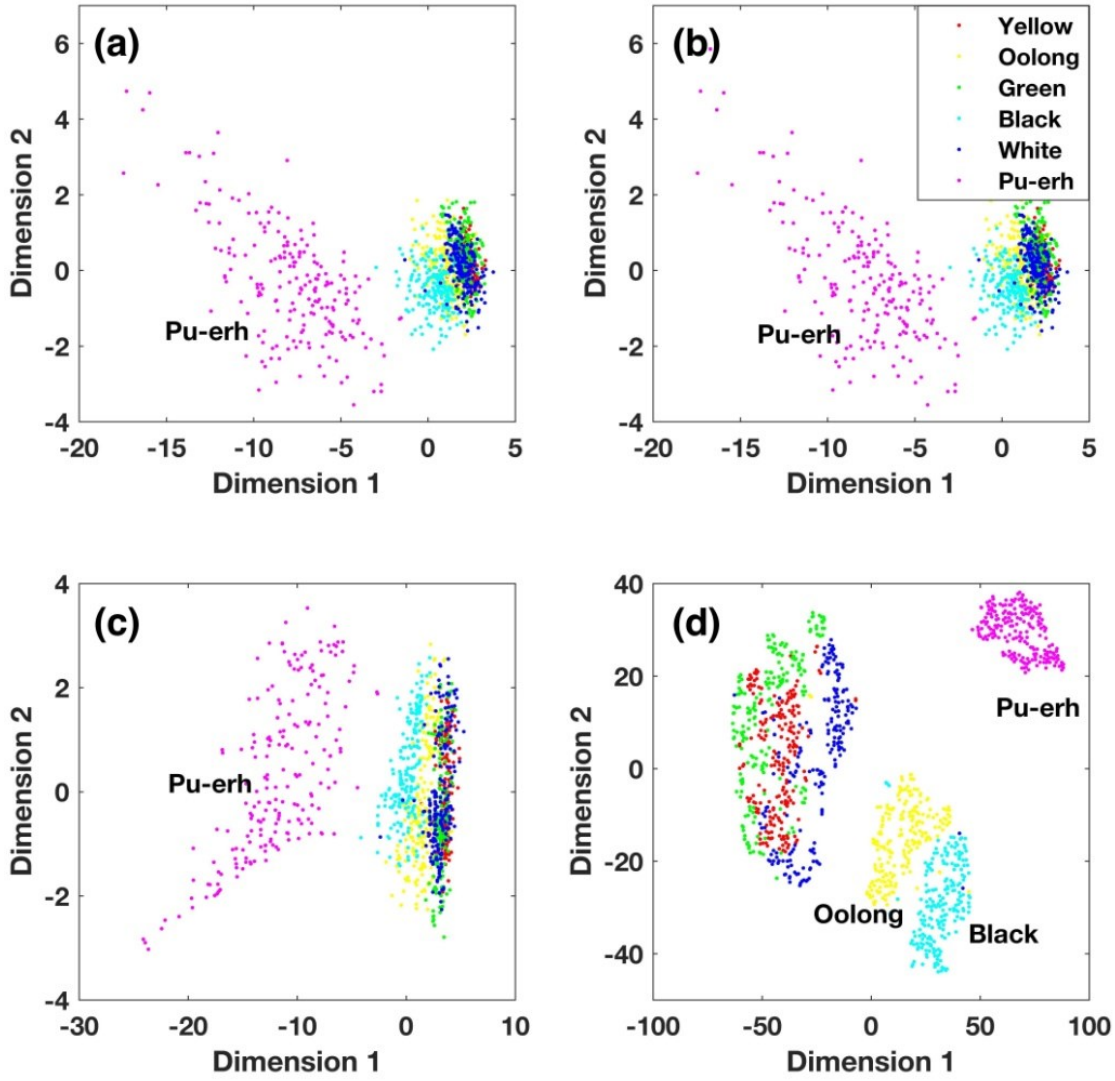


Figure 4: 2-Dimensional scatter plots for visualising high dimensional tea data. (a). Principal Component Analysis (PCA), (b). Multidimensional Scaling (MDS), (c). Isometric Mapping (ISOMAP), and (d). t-distributed Stochastic Neighbour Embedding (t-SNE). In all the plots, the first dimension is represented in the x-axis and the second in the y-axis, and the six tea products are coloured as follows: Pu-erh (pink), black (sky blue), oolong (yellow), green (green), white (blue) and yellow (red).

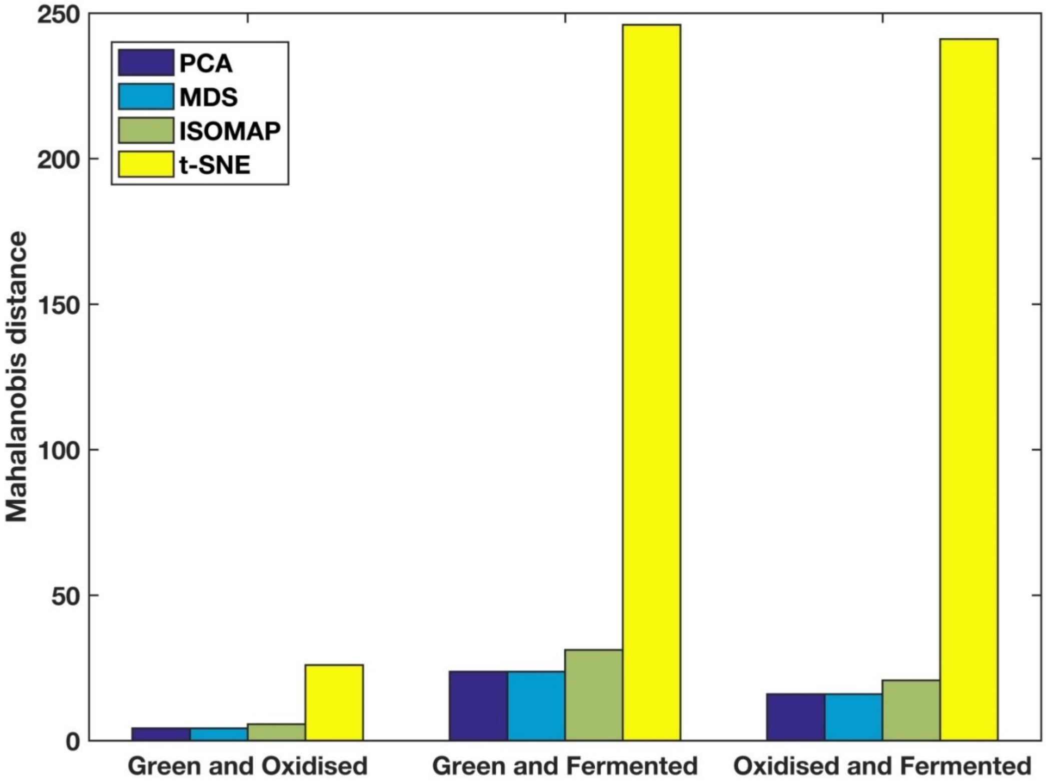


Figure 5: Mahalanobis distances between the three different cluster groups obtained using PCA (dark-blue), MDS (sky-blue), ISOMAP (light- green) and t-SNE (yellow).

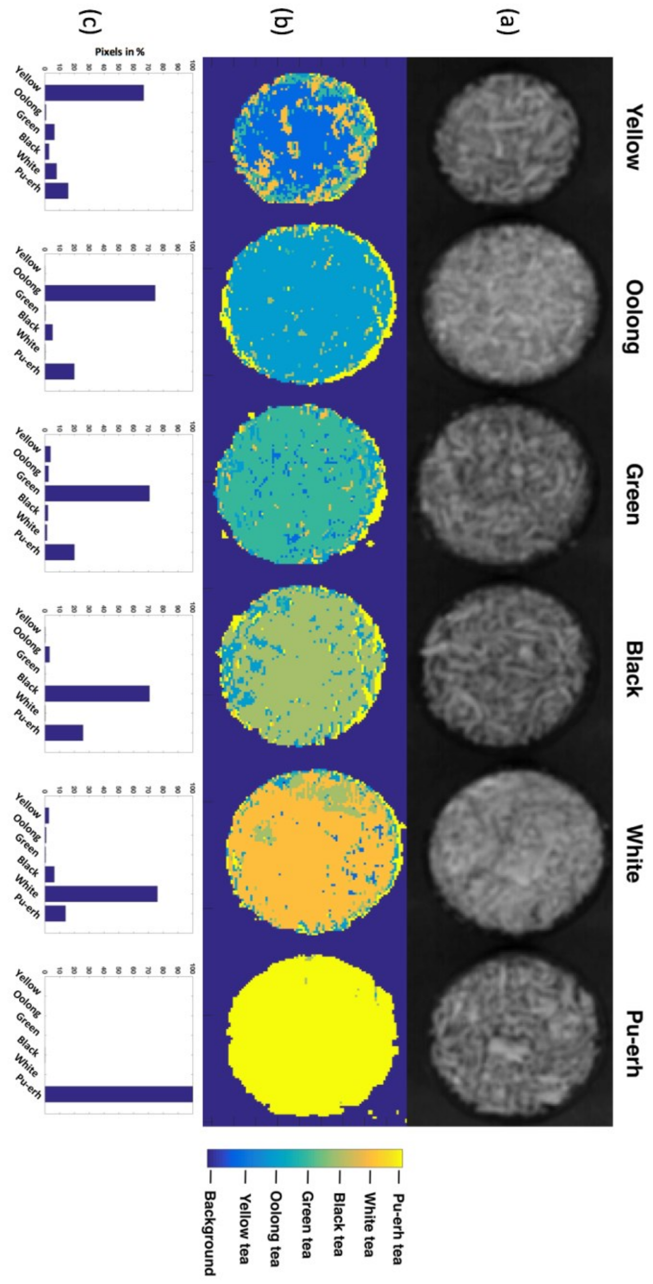
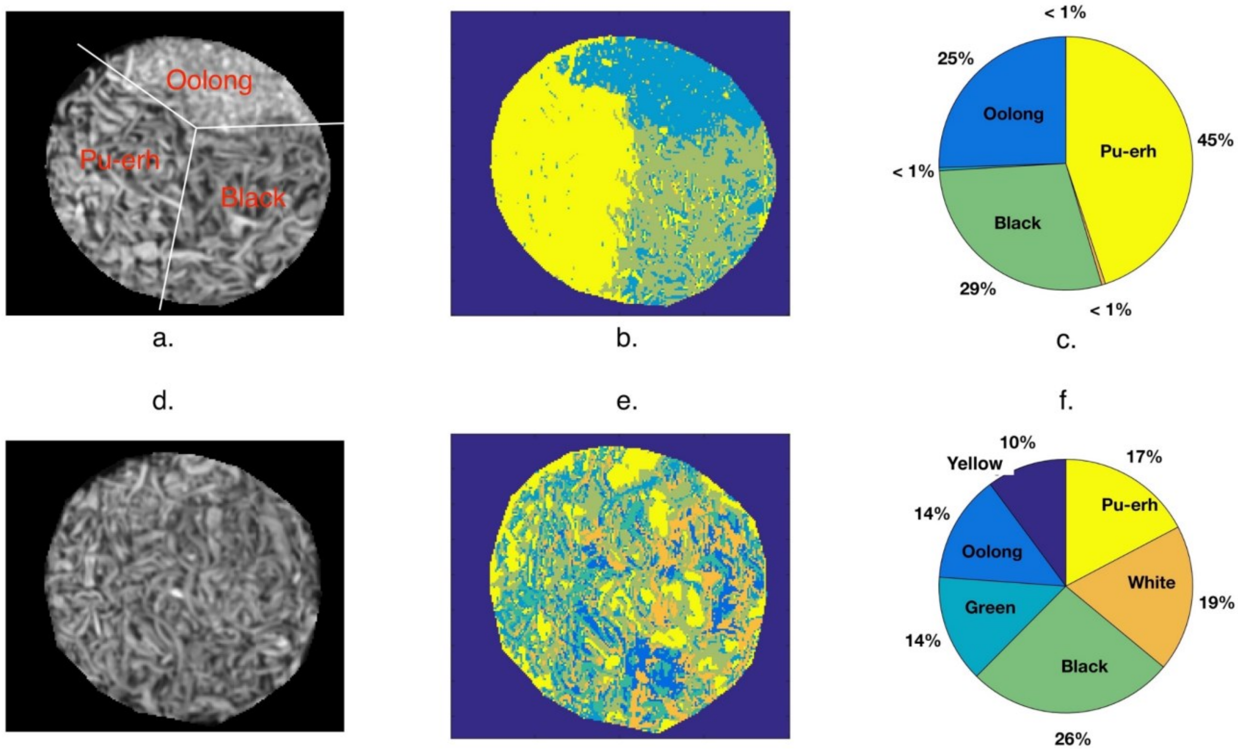


Figure 6: (a) Greyscale image constructed from the spectral plane extracted from the hypercube at 1424 nm, (b) Classification maps obtained from the application of the ECOC-SVM model. From left to right the samples can be understood as yellow (dark blue), oolong (light blue), green (cyan), black (light green), white (orange) and Pu-erh (yellow). (c) Histograms showing the proportion of pixels attributed to the different tea products for the classification maps in (b).



a. b. c.

d. e. f.

Figure 7: (a). Greyscale image at 1424 nm for the sample comprising oolong, black and Pu-erh tea, (b). The classification map for the sample comprising oolong, black and Pu-erh tea, (c). Pie chart representing the proportion of pixels belonging to a particular class for the classification map presented in (b), (d). Greyscale image at 1424 nm for a sample containing a mixture of all teas, (e). The classification map for a sample containing a mixture of all teas, and (f) Pie chart representing the proportion of pixels belonging to a particular class for the classification map presented in (e).