J.Natn.Sci.Foundation Sri Lanka 2026 54 (1): 129- 142
DOI: http://dx.doi.org/10.4038/jnsfsr.v54i1.12284

RESEARCH ARTICLE

Imaging Technology

Detection of sugar adulteration in black tea using multispectral imaging

WAND Wickramasinghe[1], G. Thilakarathne[2*], EMSLB Ekanayake[3], AD Wijesinghe[1], KAST Senarath[1], HMVR Herath[1], GMRI Godaliyadda[1], MPB Ekanayake[1], T Madhujith[4] and KMM Mohotti[5]

[1] *Department of Electrical and Electronic Engineering, Faculty of Engineering, University of Peradeniya, Peradeniya, Sri Lanka.*
[2] *Department of Electrical and Information Engineering, Faculty of Engineering, University of Ruhuna, Hapugala, Galle, Sri Lanka.*
[3] *School of Engineering and Technology, Sri Lanka Technological Campus, Meepe, Padukka, Sri Lanka.*
[4] *Department of Food Science and Technology, University of Peradeniya, Sri Lanka.*
[5] *Tea Research Institute, Thalawakelle, Sri Lanka.*

Submitted: 11 July 2024; Revised: 01 October 2025; Accepted: 14 March 2026

Abstract: Black tea, valued globally for its flavor, aroma, and nutritional benefits, is a major export commodity for countries such as China, India, and Sri Lanka. Rising demand has led to sugar adulteration to enhance color, twist, and weight, compromising quality and posing health risks, highlighting the need for rapid and reliable verification methods. This study presents a multispectral imaging (MSI) based approach for detecting and quantifying sugar adulteration in black tea. A custom-built system with thirteen narrow band LEDs (365 nm to 940 nm) sequentially illuminated powdered and brewed samples, capturing 26 spectral images in reflectance and transmittance modes, respectively. Spectral features corresponding to sugar induced color changes were extracted independently of natural tea variability. Preprocessing steps, including dark current subtraction, cropping, histogram equalization, and dimensionality reduction via linear discriminant analysis (LDA), ensured high quality data for analysis. Classification was performed using linear discriminant analysis (LDA), K-nearest neighbors (K-NN), support vector machine (SVM), feed forward neural network (FFNN), and convolutional neural network (CNN), achieving accuracies above 93% across the tested models. These methods showed high sensitivity in detecting adulteration at levels as low as 5% (w/w) and strong specificity in distinguishing pure from adulterated brewed samples. Polynomial regression was applied to quantify sugar content, yielding R^2 values above 0.97 for polynomial orders from the first to the fifth. A third order polynomial was selected as it provided a slightly improved fit (R^2 = 0.9739) while maintaining low model complexity. These results demonstrate that multispectral imaging combined with machine learning enables reliable detection of sugar adulteration and continuous estimation of adulteration levels between 5% and 25%, supporting rapid and non-destructive monitoring of black tea authenticity.

Keywords: Black tea, food quality assessment, machine learning, multispectral imaging, sugar adulteration of tea.

INTRODUCTION

Tea, or "cha," is derived from the leaves of *Camellia sinensis* and related varieties. Originally discovered in southwestern China, its global dissemination was particularly driven by the British, alongside other Western powers, and facilitated via the Silk Road (Meng et al., 2019). As the world's second most consumed beverage after water (Cabrera et al., 2006), tea production encompasses post-fermented (black), fully fermented (oolong), and unfermented (green or pu-erh) categories (Fraser et al., 2014). This study focuses on black tea, particularly popular in Asia, the European Union, and the Middle East (XU et al., 2022).

With annual global production reaching approximately 6.5 billion tons (Markets and Trade Division – Economic and Social Development Stream, 2022), tea represents a major source of foreign exchange. Leading producers

* Corresponding author (gayathrit@eie.ruh.ac.lk; https://orcid.org/0009-0004-0306-2116)

China, India, and Kenya derive USD 2.08 billion, USD 750.63 million, and USD 109 million respectively (IBEF, 2022; Kamer, 2022; Ridder, 2023), while Sri Lankan tea exports contribute 2% to the GDP, generating USD 1.3 billion annually (Nadeera, 2022).

Tea is among the most widely consumed beverages worldwide, necessitating stringent quality standards established by national and international bodies (Chen et al., 2006). Quality assessment typically evaluates caffeine, water extract, total polyphenols, and free amino acids, ensuring consistency, safety, and market acceptance (Polat et al., 2022). Tea's bioactive compounds, including polyphenols, amino acids, minerals, caffeine, and catechins, shape its flavor, aroma, and astringency while also providing potential health benefits in antioxidative activity and diabetes prevention (Meng et al., 2019; Samanta, 2020).

Pure black tea contains up to 6.5% natural sugars (fructose, glucose, sucrose, m-inositol, maltose, and raffinose), while higher levels typically indicate external addition (Engelhardt, 2010, pp. 999–1032). Adulteration with sugar or glucose syrup becomes visually detectable above 20% concentration (Luqing et al., 2015), representing the realistic adulteration range targeted for selective detection.

Black tea production involves harvesting, withering, rolling, fermenting, drying, and grading, all requiring precise control (Aaqil et al., 2023). Quality declines often occur due to sugar addition in low grade tea to improve appearance (Wedagedara et al., 2019; Wang et al., 2021), which promotes bacterial growth, moisture absorption (Luqing et al., 2015), and health risks such as obesity, cardiovascular disease, and dental caries. Therefore, accurate discrimination between pure and adulterated tea is essential.

Figure 1: The key phases of black tea production

Traditional detection methods such as, Benedict's tests (Hernández-López et al., 2020), strip methods (el Moctar et al., 2021, pp. 59–86), and sensory evaluations (Jia et al., 2022) have proven insufficient for reliable, real-time monitoring. Classical analytical techniques, including GC, CE, HPLC, and SPE-LC-MS/MS, accurately measure sugars (Liu et al., 2008; Fraser et al., 2014; Alwis et al., 2020; Piyasena et al., 2022; Przybylska et al., 2021; Wang et al., 2021) but remain time-intensive, laborious, and unsuitable for high throughput screening (Ren et al., 2013). Moreover, single point optical detectors cannot capture spatial heterogeneity in tea samples.

While hyperspectral imaging (HSI) and NIR spectroscopy have been applied to tea analysis (Chen et al., 2006; Luqing et al., 2015; Ahmad et al., 2021; Wang, Liu, et al., 2021; Hu et al., 2023), most studies rely primarily on spectral information, overlooking textural characteristics. This reduces robustness in heterogeneous samples, highlighting the need to integrate spatial and spectral data for rapid, reliable evaluation.

This study presents a nondestructive, algorithm driven multispectral imaging (MSI) system integrating reflectance and transmittance modes, employing a monochrome FLIR Blackfly USB3 camera, thirteen narrow-band LEDs spanning 365 nm to 940 nm, an integrating hemisphere, and dedicated switching circuitry. Following median filtering and Fisher's Discriminant Analysis preprocessing, classification via LDA, K-NN, SVM, FFNN, and CNN (Ekanayake et al., 2022; Wickramasinghe et al., 2021) and polynomial regression for quantification (Ostertagová, 2012) achieved over 93% classification accuracy in brewed tea and R^2 greater than 0.97 for sugar content prediction, validating multispectral imaging as a rapid, noninvasive, cost-effective solution for black tea adulteration detection.

MATERIALS AND METHODS

Sample preparation

Black tea samples were provided by the Hantana Tea Research Institute (TRI) of Sri Lanka in sealed packages to ensure integrity. Samples originated from mid-country tea plantations in the Kandy and Matale districts (Tea Research Institute, 2025). The sample set comprised a pure tea variant with natural sugars and pre-adulterated variants at 5%, 10%, 15%, 25%, and 35% (w/w). Sugar was uniformly blended during the first rolling stage using a miniature environment-controlled manufacturing (ECM) system (National Science Foundation, 2025).

Black tea grades are determined by leaf shape and size (Kamal et al., 2008). Since the proposed method relies on spectral image analysis, samples were ground into powder to eliminate grade variations and standardized particle size. Grinding and sieving minimize leaf morphology influence and enhance light interaction consistency, improving the accuracy of multispectral analysis. The grinding process is conducted by professionals at the TRI, resulting in tea particles filtered through a 1 mm mesh (Tea Research Institute, Sri Lanka, 2003). Six adulteration levels (5% to 35%), including the pure sample, were prepared with five replicates each, acquiring a total of 30 multispectral images per sample type.

Brewed tea was prepared using 20 g of tea powder per 250 ml of boiling water (~100 °C), comparable tea-to-water ratio to ISO 3103 (ISO 3103, 2019) and brewed for 5 to 6 minutes. After cooling to room temperature (21 °C) to prevent vapor condensation on the camera. A 150 ml sample was obtained by filtering through a 0.5 mm mesh to reduce light scattering, following gentle stirring for 2 minutes to ensure homogeneous sugar distribution while minimizing bubble formation. Brewed samples were imaged following the same procedure as powdered samples.

Multispectral image acquisition

The MSI system was calibrated prior to each acquisition to ensure reproducible and high-quality spectral data. Imaging parameters were standardized, with zoom adjusted to minimize background interference, aperture optimized for brightness and depth of field and focus fine-tuned to capture surface details. LED intensity, zoom, aperture, and focus were fixed across all samples, ensuring consistent, high fidelity image collection.

Monochrome images were sequentially captured at thirteen narrow band wavelengths spanning NUV to NIR regions, recording spectral and textural variations critical for detecting sugar adulteration. The system comprised a FLIR BFS-U3-13Y3M CMOS monochrome camera (1.3 MP, USB3 Vision v1.0, 1280 × 1024 resolution, 10-bit ADC) with an onsemi PYTHON1300 sensor offering 365 nm to 940 nm spectral sensitivity aligned with the selected LEDs. Figure 2 illustrates the MSI setup, and Table 1 lists the spectral bands.

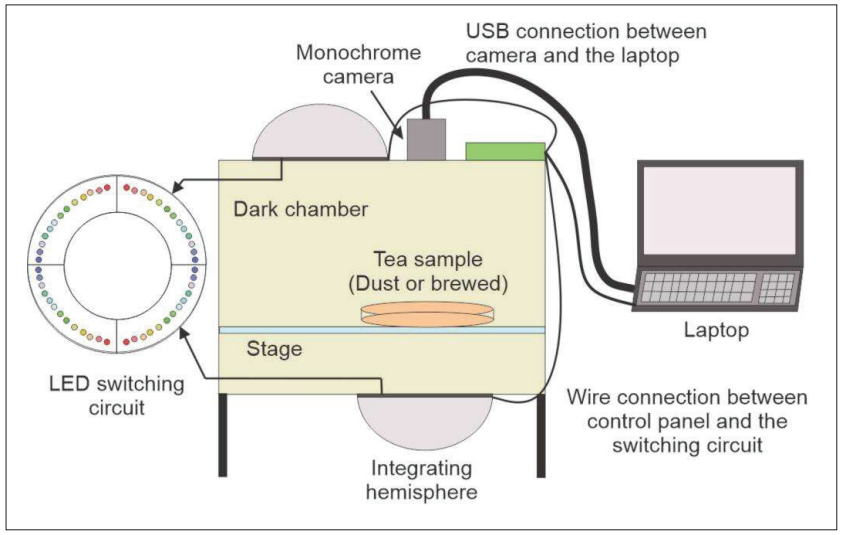

Figure 2: Schematic representation of the multispectral imaging setup

Table 1: An overview of narrow band LEDs used in the integrating hemisphere

Standard LED Part Number Dominant center	Wavelength (nm)
1080-ELUA2016OGB-P6070Q43040020-VA1MCT-ND	365
VLMU3100-GS08CT-ND	405
1830-1108-1-ND	473
1830-1109-1-ND	530
516-2827-1ND	575
516-HSMJ-A100-Q00J1CT-ND	621
754-1614-1-ND	660
1080-67-21S/NFR2CP2050A3B21522Z6/2TCT-ND	735
1125-MTSM 0077-843-IRCTC-ND	770
751-1214-ND	830
VSMY3850-GS08CT-ND	850
751-VSMY5890X01CT-ND	890
1830-IN-P32ZTHIRCT-ND	940

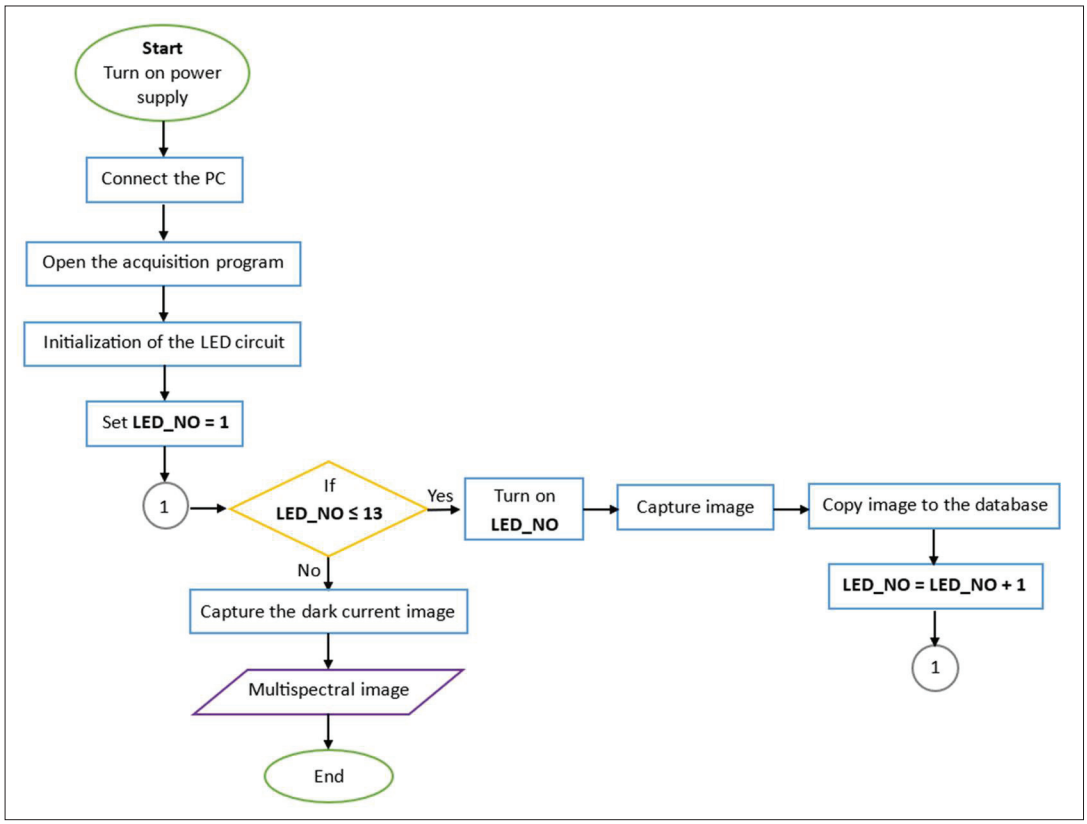

Figure 3: Flow chart of the multispectral imaging process

A custom Graphical User Interface (GUI) synchronized LED activation and image capture, eliminating operator variability. LED switching was controlled via an Arduino Due microcontroller, while the host computer managed acquisition and preprocessing. Powdered samples were imaged in reflectance mode on a wooden stage and cooled brewed samples in transmittance mode on a glass stage. Each sample was imaged across all thirteen spectral bands with dark current correction. The workflow is presented in Figure 3.

Spectral data preprocessing

Image preprocessing

The MSI device captures images of 1280 × 1024 pixels, but only a 100 × 100 pixel² window is used to focus on the brewed or powdered tea, excluding background elements such as the stage or petri dish. This window size balances sufficient data points with computational efficiency, as smaller windows reduce accuracy while larger windows offer minimal improvement.

Multispectral images obtained may contain erroneous pixel values owing to the diverse types of noise within the MSI system, predominantly originating from the camera during pixel acquisition. The photon absorption stimulates electrons within the photodetector (Skauli, 2011), which are converted into electrical signals representing the pixel values.

$$N[i,j] = \eta[i,j]N_{ph}[i,j] + I_d[i,j]t + \delta[N] \quad ...(1)$$

where $N[i,j]$ is the measured photoelectron count at the (i, j) position of the image, $\eta[i,j]$ is the quantum efficiency of the photodetector (the probability of generating a photoelectron from an incoming photon), $N_{ph}[i,j]$ is the incident photon count, $I_d[i,j]$ represents the dark current, t is the measurement time, and $\delta[N]$ denotes the readout noise associated with amplifying the photoelectric signal and converting it to digital values. The last two terms of Equation (1) are noise.

Charges can be generated in the photodetector due to thermal radiation over a unit of time, even without incident light, a phenomenon known as dark current noise (Shieh & Djordjevic, 2010, pp. 83–87). As the enclosure of the MSI device is not completely sealed during image acquisition, environmental radiation may also activate the photodiodes during image acquisition. These errors are mitigated by subtracting the dark current image from each spectral band (Bandara et al., 2020).

$$A[\lambda] = B[\lambda] - D \quad (2)$$

where $B[\lambda]$ is the raw pixel values at λ wavelength, D is the dark current image, and $A[\lambda]$ is the dark corrected pixel value.

A 2D median filter was applied to dark current corrected multispectral images to suppress high frequency electronic fluctuations and illumination nonuniformity. The filter was employed across both rows and columns simultaneously, with each pixel replaced by the median intensity of its 50 × 50 pixel² neighborhood, chosen through trial-and-error by evaluating classification accuracy across window sizes from 5 to 90 pixels. This configuration balances noise reduction and feature preservation, as larger windows may remove spectral details while smaller windows insufficiently improve the signal-to-noise ratio (SNR), maintaining spatial and spectral integrity.

Spectral signature

Each multispectral image consisted of 13 spectral bands, with five acquisitions (replicates) per sugar adulteration level, constrained by available resources at the TRI. Notably, the spatial arrangement (order) of these pixels within the cropped region was disregarded, as the analysis treated individual pixel intensities as independent data points rather than preserving their 2D locations. This pixel level approach substantially augments the effective dataset size for training and testing, minimizing the need for a large number of full images while enabling robust statistical modeling.

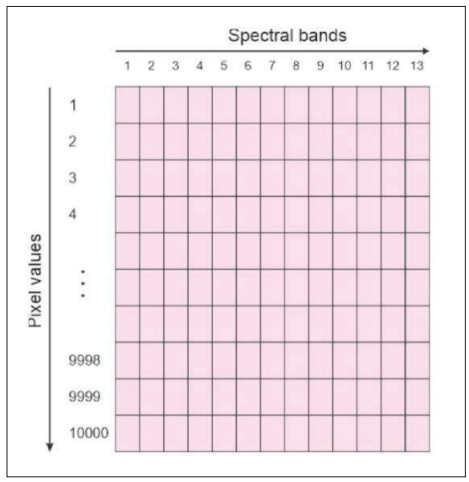

Figure 4: The generated data matrix for one sample

For data organization, a 100 × 100 pixel² region was cropped from each image, resulting in 10,000 pixels per band. These pixels from each of the 13 bands were flattened into column vectors and systematically concatenated in a controlled manner to form a 10,000 × 13 pixel² matrix per sample, where R represents the number of pixels (10,000) and B represents the number of spectral bands (13), defining the general matrix size as $R \times B$ as shown in Figure 4. This process was repeated for all 30 images (5 replicates × 6 sugar levels: 0%, 5%, 10%, 15%, 25%, 35%).

To generate spectral signatures, the mean pixel intensity was calculated across all pixels and replicates for each sugar level and spectral band. The resulting mean values were plotted against the dominant peak wavelength of each band to visualize the spectral response across adulteration levels, as shown in Figure 5.

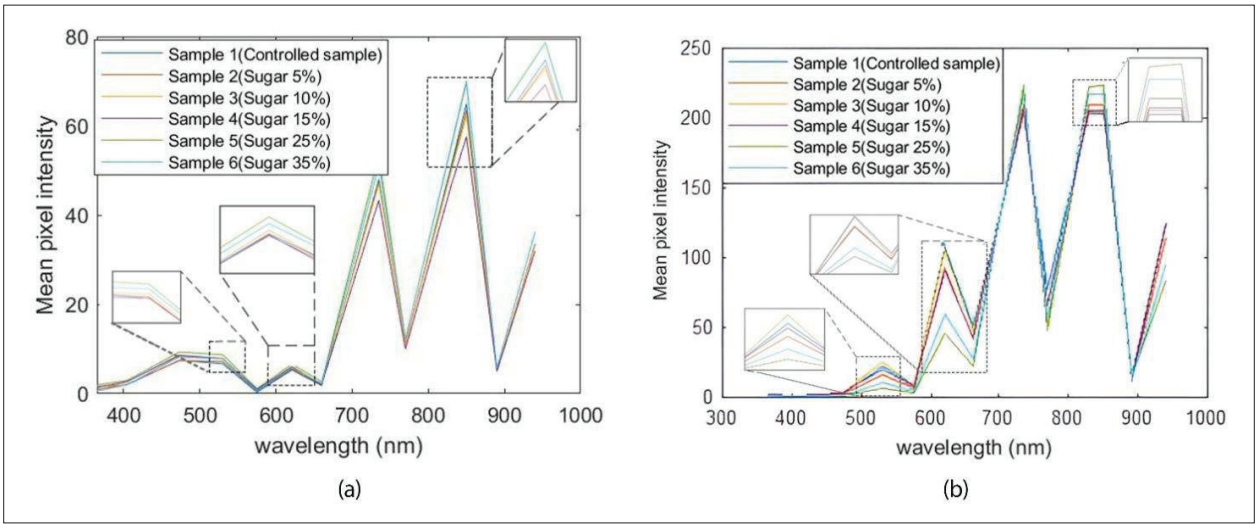

Figure 5: The spectral signature of the (a) powdered and (b) brewed black tea samples

The spectral signatures reveal differentiation among sugar levels at various wavelengths. Notably, brewed black tea shows clearer separations than powdered tea, as illustrated in Figure 5, particularly at peak wavelengths of 530 nm, 621 nm, 830 nm, and 850 nm. This suggests that brewed tea may yield higher classification accuracy. For dataset splitting, an 80:20 training-to-testing ratio was adopted, as recent studies indicate it optimizes performance in machine learning applications (Nguyen et al., 2021). Accordingly, four replicates per sugar level were allocated to training, and one replicate to testing.

The training matrix was formed by vertically stacking the 10,000 × 13 matrices from the training replicates across all classes, yielding a 240,000 × 13 matrix (4 replicates × 10,000 pixels × 6 classes). Similarly, the testing matrix comprised a 60,000 × 13 matrix (1 replicate × 10,000 pixels × 6 classes). Then, the training data matrix and testing data matrix are formed by vertically stacking corresponding $R \times B$ pixel² matrices one below the other. Ultimately, 2D matrices $R_{Tr} \times B$ pixel² (where $R_{Tr} = R \times$ training multispectral images per each class × number of classes = 240,000) and $R_{Te} \times B$ pixel² (where $R_{Te} = R \times$ testing multispectral images per each class × number of classes = 60,000) are created as the training data set and the testing data set, respectively.

Fisher's discriminant analysis

Fisher's Discriminant Analysis (FDA) was applied to enhance class separability before classification (Shashoa et al., 2016). FDA projects the data into a lower-dimensional space by maximizing between-class variation while minimizing within-class variation.

Let the training data set be represented as

$$X = [x_0, x_5, x_{10}, x_{15}, x_{25}, x_{35}]^T$$

where $x_k x_k$ (k = 0, 5, 10, 15, 25, and 35) denotes the dataset for each sugar adulteration level with a dimensionality of $R_{cl} \times B$ (number of rows per each sugar adulterant level, $R_{cl} = R \times$ number of training multispectral images = 40000) (Herath et al., 2023).

Step 01: Calculation of the mean vector of each sugar level.

$$\mu_k = \frac{1}{R_{cl}} \sum_{i=1}^{R_{cl}} x_k(i) \quad ...(3)$$

where $x_k(i)$ is the i^{th} row of x_k sugar adulteration level matrix.

Step 02: Calculation of the total mean vector.

$$\mu = \frac{1}{R_{Tr}} \sum_{i=1}^{R_{Tr}} X(i) \quad ...(4)$$

where $X(i)$ is the i^{th} row of the training data set (X).

Step 03: Calculation of the between class scatter matrix.

$$S_b = \sum_{i=1}^{L} (\mu_i - \mu)(\mu_i - \mu)^T \quad ...(5)$$

where L is the number of sugar adulterant levels, in this particular case $L = 6$ and μ_i is the mean vector of each sugar level.

Step 04: Calculation of the within class scatter matrix.

$$S_w = \sum_{i=1}^{L} C_i \quad ...(6)$$

$$C_k = \sum_{i=1}^{L} (x_k(i) - \mu_k)(x_k(i) - \mu_k)^T \quad ...(7)$$

where C_k is the covariance matrix of each sugar level (k = 0, 5, 10, 15, 25, and 35).

Step 05: Solve the following eigenvalue problem.

$$S_b w_p = \lambda_p S_w w_p \quad ...(8)$$

where w_p is the eigenvector corresponding to the λ_p eigenvalue (p = 1, 2, . . ., B) and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \geq \cdots \geq \lambda_{13}$

Step 06: Generation of the projection matrix.

$$W = [w_1, w_2, \ldots, w_k, \ldots, w_p] \quad ...(9)$$

where W is generated by extracting and sorting the eigenvectors corresponding to the eigenvalues in descending order. Here, the dimensionality of W is $B \times B$, and it is possible to reduce the dimensionality of W to $B \times m$ ($m \leq B$).

Step 07: Projection of the training data set into the new space.

$$Z(X) = XW \quad ...(10)$$

where Z is the projected training data set with the dimension of $R_{Tr} \times B$.

The FDA procedure involved computing class and overall means, estimating within and between class scatter, and solving the generalized eigenvalue problem to derive discriminant vectors. These vectors were used to project the datasets into a discriminant subspace, enhancing the visibility of class level differences. The $Z(X)$ illustration presents multiple rotated views to highlight the spatial distribution, as shown in Figures 6 and 7.

The 3D scatter plots project powdered and brewed black tea samples onto the first three FDA components, representing different sugar levels. Figure 6 shows partial separation among classes, with overlaps between intermediate levels (10% to 25%), indicating moderate discriminative capability. In contrast, Figure 7 reveals clearer separation in brewed tea, with distinct clusters along the principal axes, demonstrating FDA's effectiveness in maximizing between-class variance. Overall, brewed tea exhibits a more pronounced separation among sugar levels than powdered tea.

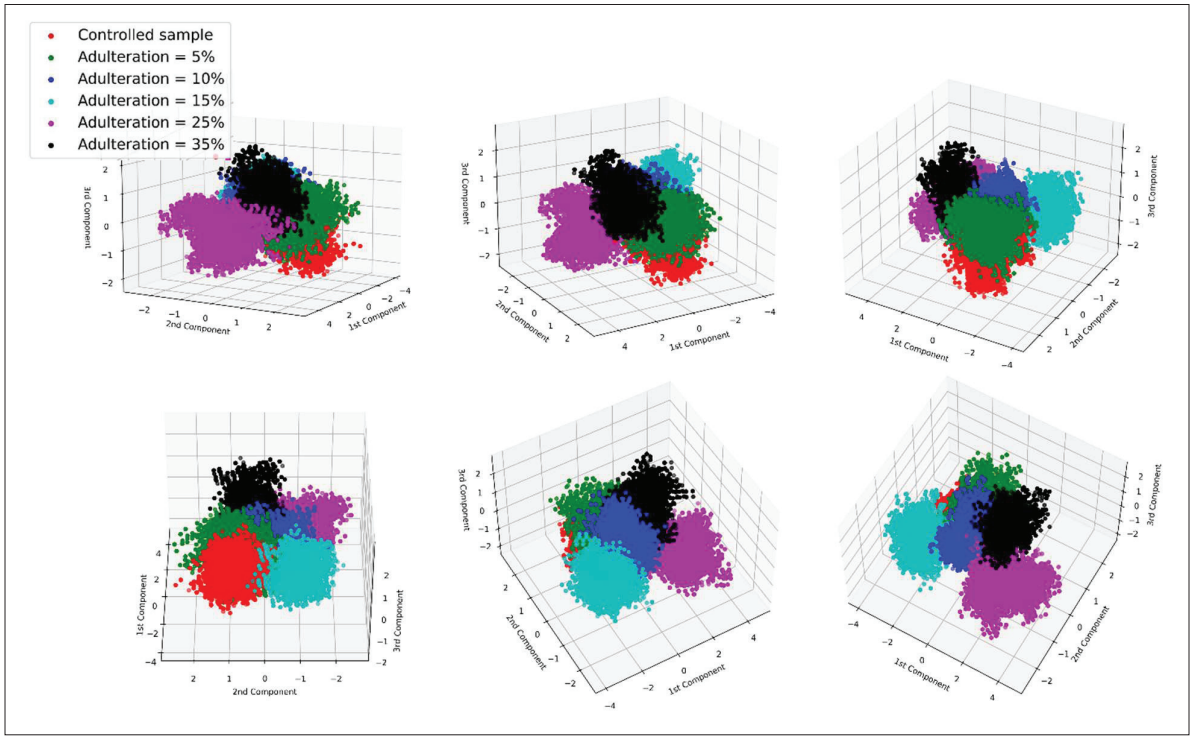

Figure 6: 3D visualization of FDA projected powdered black tea samples across the first three components with multiple rotated views

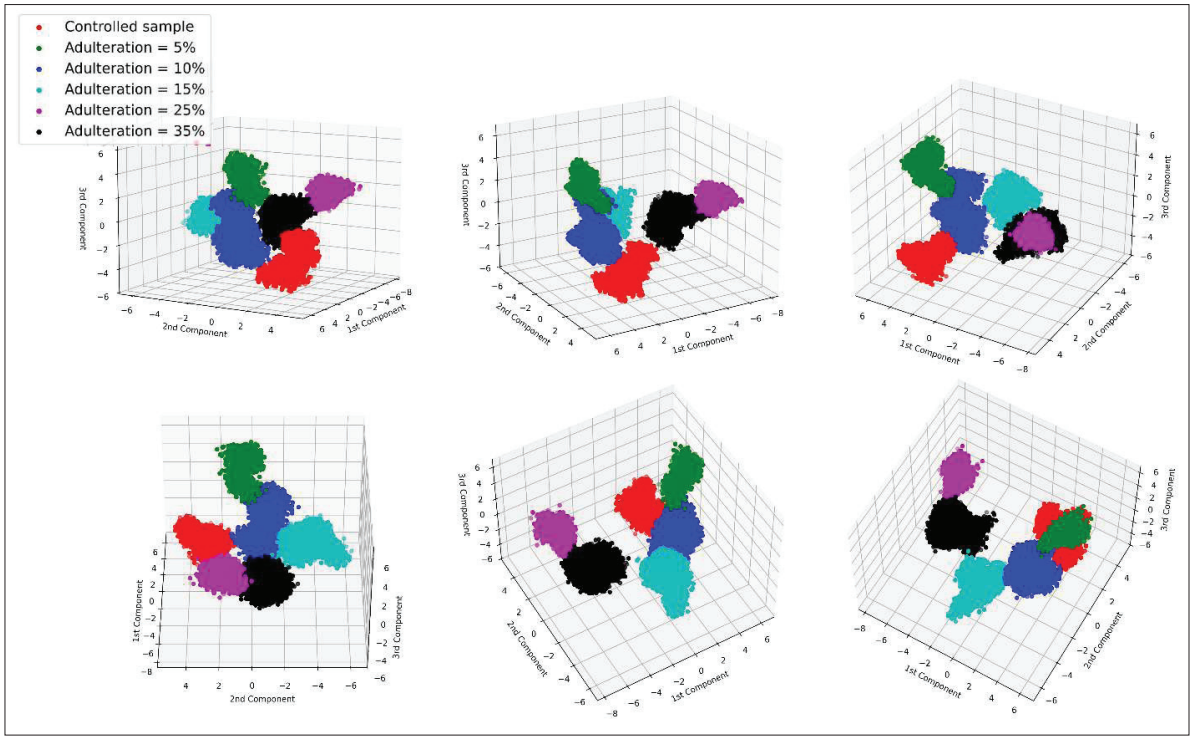

Figure 7: 3D visualization of FDA projected brewed black tea samples across the first three components with multiple rotated views

Classification models

Tea samples were classified based on sugar adulteration levels using machine learning algorithms, such as Linear Discriminant Analysis (LDA), K-Nearest Neighbors (K-NN), Support Vector Machine (SVM), Feed-Forward Neural Network (FFNN), and Convolutional Neural Network (CNN). LDA provided dimensionality reduction while preserving class separability in high-dimensional spectral datasets. K-NN, a non-parametric method, enabled robust detection of local spectral patterns without distributional assumptions, with MATLAB hyperparameter optimization identified optimal performance at k = 1 with cosine distance (Guo et al., 2003). SVM with a Gaussian kernel separated classes in higher dimensional space using one-vs-rest (OVR) multiclass classification (Zhao et al., 2009; Saha et al., 2012).

FFNN captured complex nonlinear relationships between multispectral features and adulteration concentrations using five hidden layers (1024 to 32 neurons) and the Adam optimizer (learning rate 0.0001, selected through trial-and-error) (Murat, 2006). CNN autonomously extracted spatial spectral features from cropped multispectral images (100 × 100 × 13) using a single convolutional layer (3 × 3), max pooling, and one fully connected hidden layer (Alzubaidi et al., 2021; Gu et al., 2018). The CNN was trained over 100 epochs with a batch size of 16, employing Adaptive Moment Estimation (Adam) optimization and a categorical cross entropy loss function for multiclass probabilistic discrimination.

Regression model

Although the classification algorithms showed high accuracy, they were limited in interpolating sugar concentrations between discrete trained levels due to substantial gaps in the data. To address this issue, a polynomial curve fitting model was implemented to enable continuous estimation of sugar content, establishing a functional relationship between independent and dependent variables (Ostertagová, 2012).

Analysis of FDA projected data demonstrated clear class separability, prompting the use of histograms to evaluate sample distributions across sugar adulteration levels as shown in Figure 8. The bell-shaped distributions displayed varying degrees of overlap, with the controlled sample (2% natural sugar) and 35% samples showing significant dispersion. This led to their exclusion to concentrate on the 5% to 25% range. A scatter plot of sugar levels against the first FDA component, shown in Figure 9, indicated horizontal clusters at 5%, 10%, 15%, 20%, and 25%, with component values increasing with higher adulteration, suggesting a predominantly linear yet slightly nonlinear trend amenable to polynomial fitting.

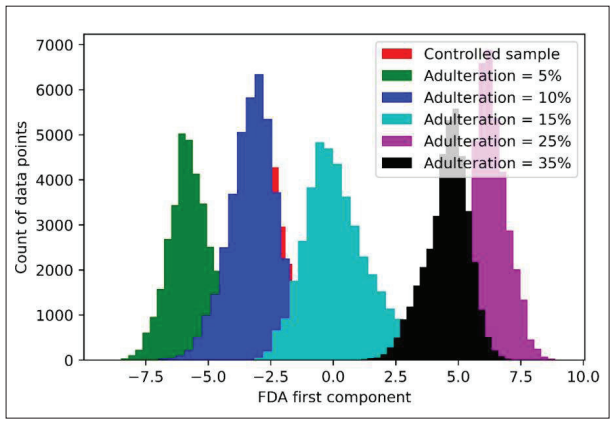

Figure 8: Histogram of the training dataset projected onto the first FDA component

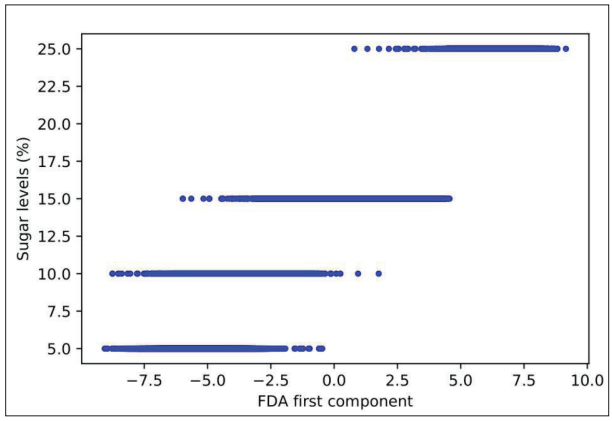

Figure 9: Scatter plot of sugar adulteration levels (5% to 25%) versus the first FDA component, demonstrating clustered data points and the underlying trend for polynomial fitting

The general form of a polynomial function is as follows, with the independent variable X and the dependent variable S.

$$S = \alpha_0 + \alpha_1 X^1 + \alpha_2 X^2 + \cdots + \alpha_n X^n \quad \ldots(18)$$

where n is the order of the polynomial

The mean squared error (MSE) is utilized to evaluate the accuracy of the developed curve, defined as:

$$MSE = \frac{1}{R_{Tr}} \sum_{i=1}^{R_{Tr}} (S_i - \hat{S}_i)^2 \quad ...(19)$$

where S_i is the actual sugar class label and $\hat{S}_i$ is the predicted sugar label.

RESULTS AND DISCUSSION

In this section, the results of both the classification algorithms and the regression algorithm are discussed individually.

Classification algorithms

These algorithms aim to categorize a given data point into a discrete sugar level. Utilizing the data captured by the MSI device and the first five components of the FDA applied data, machine learning algorithms and subsequent neural networks were employed. Results were obtained for both powdered and brewed black tea data to identify the most suitable sample preparation method.

The testing and training accuracies for the K-NN, LDA, SVM, FFNN, and CNN algorithms were evaluated initially. The results for powdered black tea are detailed in Table 2, whereas the results for brewed black tea are presented in Table 3.

Table 2: The training and testing accuracies for the powdered black tea original MSI data

Algorithm	Training accuracy (%)	Testing accuracy (%)
K-NN	99.928	68.470
LDA	80.477	60.870
SVM	89.060	64.889
FFNN	97.800	76.600
CNN	98.750	98.750

Table 3: The training and testing accuracies for the brewed black tea original MSI data

Algorithm	Training accuracy (%)	Testing accuracy (%)
K-NN	96.639	84.695
LDA	99.910	98.603
SVM	99.718	98.852
FFNN	99.771	96.967
CNN	93.229	92.188

Considerably improved validation accuracies were obtained for brewed black tea samples, reflecting the efficacy of multispectral imaging in detecting sugar adulteration signatures. Both LDA and SVM achieved validation accuracies above 98%, emphasizing their robustness for this classification task. However, since SVM demands greater computational resources, LDA emerges as the more efficient choice, particularly under computational constraints.

For neural network-based approaches, FFNN achieved strong validation accuracy, whereas CNN attained 98.75% testing accuracy, closely matching training, demonstrating superior generalization. CNN's convolutional layers more effectively capture complex spectral features in powdered and brewed samples than FFNN's pixel-based approach. This performance suggests robust invariant feature learning, particularly in homogeneous brewed samples. While FFNN outperformed traditional models, CNN's superior handling of high-dimensional data establishes it as the most reliable for unseen sample detection, with LDA offering an efficient alternative when resources are limited.

In theory, the FDA enhances the class separation of the data. To leverage this theoretical advantage, the same algorithms are assessed for training and testing accuracies. The results for powdered black tea and brewed black tea are tabulated in Table 4 and Table 5, respectively.

Table 4: The training and testing accuracies for the powdered black tea FDA applied data

Algorithm	Training accuracy (%)	Testing accuracy (%)
K-NN	99.928	68.498
LDA	88.673	73.200
SVM	69.449	60.510
FFNN	88.900	70.300

Table 5: The training and testing accuracies for the brewed black tea FDA applied data

Algorithm	Training accuracy (%)	Testing accuracy (%)
K-NN	99.877	98.495
LDA	98.695	93.125
SVM	97.146	96.193
FFNN	99.578	93.437

The implementation of the FDA has resulted in decreased testing accuracy, except for LDA in powdered tea and K-NN in brewed black tea. Nonetheless, the overall validation accuracies for brewed black tea remain above 93%.

Brewed black tea images proved more effective for detecting sugar adulteration, as brewing extracts soluble components, including sugar, into the liquid phase, yielding clearer and more uniform spectral changes. In contrast, powdered tea retains heterogeneous particle structures and natural leaf variations, which can obscure subtle adulteration effects. Thus, brewed samples provide more consistent spectral signatures, enabling more accurate classification and highlighting their advantage for multispectral imaging-based detection.

Table 6: The validation accuracy of the fitted polynomial functions of the orders of one to five

Order	R^2 *Score*	Fitted function (P=)
1st order	0.9734	$+1.582641X + 14.63158648$
2nd order	0.9700	$+0.007222X^2 + 1.575401X + 14.476192$
3rd order	0.9739	$-0.003618X^3 + 0.007136X^2 + 1.704775X + 14.582717$
4th order	0.9748	$-0.000200X^4 - 0.003618X^3 + 0.016530X^2 + 1.708238X + 14.534085$
5th order	0.9802	$-0.000652X^5 + 0.000572X^4 + 0.040239X^3 - 0.008211X^2 + 1.064254X + 14.333792$

Regression analysis

Continuous prediction of sugar levels in brewed black tea was performed using regression, with the first FDA principal component as the independent variable (X) and sugar concentration (in percentage) as the dependent variable (P). Polynomial regression models of orders 1st to 5th were trained to capture nonlinear trends, and performance was evaluated using the coefficient of determination (R^2) (Enders, 2022; Zhang et al., 2013).

As shown in Table 6, the linear model (1st order) yielded an $R^2 = 0.9734$, indicating a predominantly linear relationship. Higher order models slightly improved fit, with the 5th order polynomial achieving the highest R^2 of 0.9802. However, coefficients beyond the 3rd order were negligible and added complexity. Given that escalating polynomial order did not yield substantial gains in R^2 and risked overfitting, particularly with data clustered around discrete sugar levels (5%, 10%, 15%, 20%, and 25%) the 3rd order polynomial was selected as the optimal balance between accuracy and parsimony.

Subsequently, the polynomial fitting function of the 3rd order was validated using unseen data from the polynomial fitting process, yielding an R^2 score of 0.9739. The polynomial curve is shown in Figure 10, where the red curve represents the fitted 3rd order polynomial, and the green markings represent the predictions of the validation data. The results demonstrate that the 3rd order polynomial accurately models the relationship between the input spectral features and sugar content, providing reliable predictions for unseen samples.

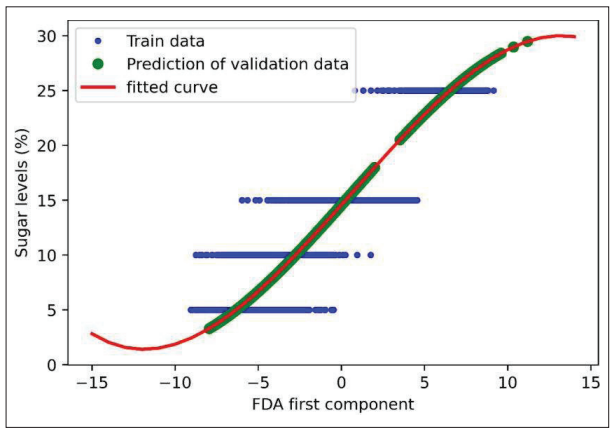

Figure 10: The 3rd order polynomial modeling of sugar content in brewed black tea

The selected 3rd order model achieved robust performance, explaining approximately 97% of the variance in sugar levels. Its high accuracy makes it a reliable tool for quantifying sugar adulteration in brewed black tea, offering improved sensitivity over categorical

classification by enabling continuous predictions within the specified range.

Beyond this study, the MSI setup has demonstrated versatility in detecting turmeric adulteration (Chaminda Bandara et al., 2020), monitoring fungal growth (Darsha Udayanga et al., 2024), quantifying aflatoxin contamination (Madushan et al., 2024), differentiating reheated coconut oil (Ranasinghe et al., 2022), and evaluating soil moisture (Ranasinghe et al., 2025), highlighting its adaptability for interdisciplinary research in food science, agriculture, and construction.

CONCLUSION

The present study evaluated the feasibility of using multispectral imaging to estimate sugar adulteration levels in black tea, using sugar adulterated black tea samples as a case study. Images were captured following standard laboratory conditions of room temperature (21 °C) and atmospheric pressure (0.999 atm) and procedures included controlled sample handling, uniform preparation of powdered and brewed tea, consistent imaging, and proper calibration of instruments to ensure reproducibility and accuracy. Thirteen narrowband LEDs from 365 nm to 940 nm were used for image acquisition. The images were preprocessed via dark current removal and median filtering, followed by FDA based dimensionality reduction. Classification with K-NN, LDA, SVM, and FFNN achieved testing accuracies above 93%, with brewed tea samples showing clearer class separation than powdered tea. Continuous prediction of sugar content in brewed tea was enabled through a 3rd order polynomial regression model, achieving an R^2 of 0.97, with high precision for intermediate sugar levels of 5% to 25%.

This work is novel in integrating a cost effective MSI platform with advanced multistage machine learning and regression models to provide both classification and continuous estimation of sugar adulteration. Unlike conventional chemical methods, the approach is rapid, nondestructive, and requires minimal sample preparation. It has strong applicability in the tea industry and across all sciences and technologies, enabling regulatory bodies, quality assurance laboratories, and exporters to efficiently ensure product authenticity, enhance quality control, and protect consumer trust. The scalable framework can be further refined with higher resolution sensors and expanded spectral ranges, supporting industrial deployment and broader applications in food quality monitoring.

Acknowledgement

The authors gratefully acknowledge the Tea Research Institute (TRI) and the Sri Lanka Tea Board for generously providing the tea samples, laboratory expertise, and support.

Special appreciation is extended to Dr. Keerthi M. Mohotti, Deputy Director of TRI, and Dr. Kumara Kithsiri, Deputy Tea Commissioner of the Sri Lanka Tea Board, for their invaluable guidance and support throughout the research process, which significantly contributed to achieving favorable outcomes.

REFERENCES

Aaqil, M., Peng, C., Kamal, A., Nawaz, T., Zhang, F., & Gong, J. (2023). Tea harvesting and processing techniques and its effect on phytochemical profile and final quality of black tea: A review. *Foods*, *12*(24), 4467. https://doi.org/10.3390/foods12244467

Ahmad, H., Sun, J., Nirere, A., Shaheen, N., Zhou, X., & Yao, K. (2021). Classification of tea varieties based on fluorescence hyperspectral image technology and ABC-SVM algorithm. *Journal of Food Processing and Preservation*, *45*(3). https://doi.org/10.1111/jfpp.15241

Alwis, J. A. P. N., Perera, G. A. A. R., & Abeywickrama, K. R. W. (2020). Determination of Sucrose, Glucose and Fructose Level in Low Grown Black Tea Using High Performance Liquid Chromatography . *International Research Conference of Uva Wellassa University*, 281.

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*(1), 1–74. https://doi.org/10.1186/s40537-021-00444-8

Cabrera, C., Artacho, R., & Giménez, R. (2006). Beneficial effects of green tea—A review. *Journal of the American College of Nutrition*, *25*(2), 79–99. https://doi.org/10.1080/07315724.2006.10719518

Chaminda Bandara, W. G., Kasun Prabhath, G. W., Sahan Chinthana Bandara Dissanayake, D. W., Herath, V. R., Roshan Indika Godaliyadda, G. M., Bandara Ekanayake, M. P., Demini, D., & Madhujith, T. (2020). Validation of multispectral imaging for the detection of selected adulterants in turmeric samples. *Journal of Food Engineering*, *266*, 109700. https://doi.org/10.1016/j.jfoodeng.2019.109700

Chen, Q., Zhao, J., Zhang, H., & Wang, X. (2006). Feasibility study on qualitative and quantitative analysis in tea by near infrared spectroscopy with multivariate calibration. *Analytica Chimica Acta*, *572*(1), 77–84. https://doi.org/10.1016/j.aca.2006.05.007

Ekanayake, E. M. S. L. B., Wickramasinghe, W. A. N. D., Wijesinghe, A. D., Wijedasa, M. A. C. S., Weerasooriya,

H. M. H. K., Ranasinghe, D. Y. L., Ekanayake, M. P. B., Herath, H. M. V. R., Godaliyadda, G. M. R. T., & Madhujith, T. (2021). Multispectral Imaging System to Estimate Sugar Adulteration Level of Black Tea. *2021 IEEE 16th International Conference on Industrial and Information Systems (ICIIS)*, 17–22.

Fraser, K., Harrison, S. J., Lane, G. A., Otter, D. E., Hemar, Y., Quek, S.-Y., & Rasmussen, S. (2014). Analysis of low molecular weight metabolites in tea using mass spectrometry-based analytical methods. *Critical Reviews in Food Science and Nutrition*, *54*(7), 924–937. https://doi.org/10.1080/10408398.2011.619670

Kamer, L. (2022, August 1). Tea industry in Kenya-statisticsfacts. Statista.https://www.statista.com/topics/8363/tea-industry-in-kenya/

el Moctar, B. O., Schellin, T. E., & Söding, H. (2021). *Numerical Methods for Seakeeping Problems* (pp. 59–86). Springer Nature.

Enders, F. B. (2022). Coefficient of Determination | Interpretation & Equation | Britannica. In *Encyclopædia Britannica*.

Engelhardt, U. H. (2010). *Comprehensive Natural Products II* (pp. 999–1032). Elsevier. https://doi.org/10.1016/B978-008045382-8.00089-7.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., & Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, *77*, 354–377. https://doi.org/10.1016/j.patcog.2017.10.013

Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2003). KNN Model-Based Approach in Classification. *On the Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, *2888*, 986–996. https://doi.org/10.1007/978-3-540-39964-3_62

Herath, S., Weerasooriya, H. K., Ranasinghe, D. Y. L., Bandara, W. G. C., Herath, V. R., Godaliyadda, R. I., Ekanayake, M. P. B., & Madhujith, T. (2023). Quantitative assessment of adulteration of coconut oil using transmittance multispectral imaging. *Journal of Food Science and Technology*. https://doi.org/10.1007/s13197-023-05697-0

Hernández-López, A., Sánchez Félix, D. A., Zuñiga Sierra, Z., García Bravo, I., Dinkova, T. D., & Avila-Alejandre, A. X. (2020). Quantification of reducing sugars based on the qualitative technique of benedict. *ACS Omega*, *5*(50), 32403–32410. https://doi.org/10.1021/acsomega.0c04467

Hu, Y., Huang, P., Wang, Y., Sun, J., Wu, Y., & Kang, Z. (2023). Determination of Tibetan tea quality by hyperspectral imaging technology and multivariate analysis. *Journal of Food Composition and Analysis*, *117*, 105136–105136. https://doi.org/10.1016/j.jfca.2023.105136

IBEF. (2021, September 23). The tea industry in India. *India Brand Equity Foundation*. https://www.ibef.org/blogs/the-tea-industry-in-india

ISO 3103. (2019). *Tea — Preparation of liquor for use in sensory tests* (ISO 3103:2019(E)). International Organization for Standardization. htps:/standards.iteh.ai/catalog/standards/sist/a731214d-fbae-4850-9bc65c232bdd73c3/iso-3103-2019.

Jia, X., Ren, J.-N., Fan, G., Reineccius, G. A., Li, X., Zhang, N., An, Q., Wang, Q., & Pan, S. (2022). Citrus juice off-flavor during different processing and storage: Review of odorants, formation pathways, and analytical techniques. *Critical Reviews in Food Science and Nutritio*, *64*, 3018–3043. https://doi.org/10.1080/10408398.2022.2129581

Kamal, R. J., Wadood, A. W. M. R., & Silva, K. D. R. R. (2008). Total antioxidant capacity of selected grades of black tea grown in different geographical elevations in Sri Lanka. *Tropical Agricultural Research*, *20*, 155–161.

Liu, E-Hu., Qi, L.-W., Cao, J., Li, P., Li, C.-Y., & Peng, Y.-B. (2008). Advances of modern chromatographic and electrophoretic methods in separation and analysis of flavonoids. *Molecules*, *13*(10), 2521–2544. https://doi.org/10.3390/molecules13102521

Luqing, L., Lingdong, W., Jingming, N., & Zhengzhu, Z. (2015). Detection and quantification of sugar and glucose syrup in roasted green tea using near infrared spectroscopy. *Journal of near Infrared Spectroscopy*, *23*(5), 317–325. https://doi.org/10.1255/jnirs.1178

Markets and Trade Division - Economic and Social Development stream . (2022). *International tea market: market situation, prospects and emerging issues (FAO document CC0238EN/1/05.22)* (pp. 3–5). Food and Agriculture Organization of the United Nations.

Meng, L., Chen, X., Chen, X., Yuan, L., Shi, W., Cai, Q., & Huang, G. (2019). Linear and nonlinear classification models for tea grade identification based on the elemental profile. *Microchemical Journal*, *153*, 104512–104512. https://doi.org/10.1016/j.microc.2019.104512

Murat H., S. (2006). A brief review of feed-forward neural networks. *Communications Faculty of Science University of Ankara*, *50*(1), 11–17. https://doi.org/10.1501/commua1-2_0000000026

Nadeera, D. (2022, January 9). *Lankan tea exports earned $ 1.3 Bn in 2021*. The Island Online; Upali Newspapers (Pvt) Ltd. https://island.lk/lankan-tea-exports-earned-1-3-bn-in-2021/

National Science Foundation. (2025). National Instrument Database, National Science Foundation of Sri Lanka. https://nid.nsf.gov.lk/instrumentView/219

Nelum P. Piyasena, K. G., Hettiarachchi, L. S. K., D.P.S. Jayawardhane, S. A., U. Edirisinghe, E. N., & Jayasinghe, W. S. (2022). Evaluation of inherent fructose, glucose and sucrose concentrations in tea leaves (*Camellia sinensis* L.) and in black tea. *Applied Food Research*, *2*(1), 100100. https://doi.org/10.1016/j.afres.2022.100100

Nguyen, Q. H., Ly, H.-B., Ho, L. S., Al-Ansari, N., Le, H. V., Tran, V. Q., Prakash, I., & Pham, B. T. (2021). Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil. *Mathematical Problems in Engineering*, *2021*, 1–15. https://doi.org/10.1155/2021/4832864

Ostertagová, E. (2012). Modelling using Polynomial Regression. *Procedia Engineering*, *48*(1), 500–506. https://doi.org/10.1016/j.proeng.2012.09.545

Polat, A., Kalcıoğlu, Z., & Müezzinoğlu, N. (2022). Effect of infusion time on black tea quality, mineral content and sensory properties prepared using traditional Turkish infusion method. *International Journal of Gastronomy and Food Science*, *29*, 100559. https://doi.org/10.1016/

j.ijgfs.2022.100559

Przybylska, A., Gackowski, M., & Koba, M. (2021). Application of capillary electrophoresis to the analysis of bioactive compounds in herbal raw materials. *Molecules*, *26*(8), 2135. https://doi.org/10.3390/molecules26082135

Ren, G., Wang, S., Ning, J., Xu, R., Wang, Y., Xing, Z., Wan, X., & Zhang, Z. (2013). Quantitative analysis and geographical traceability of black tea using Fourier transform near-infrared spectroscopy (FT-NIRS). *Food Research International*, *53*(2), 822–826. https://doi.org/10.1016/j.foodres.2012.10.032

Ridder, M. (2023). *Leading tea exporters worldwide in 2022*. Statista. https://www.statista.com/statistics/264189/main-export-countries-for-tea-worldwide/#statisticContaine

Samanta, S. (2020). Potential bioactive components and health promotional benefits of tea (Camellia sinensis). *Journal of the American College of Nutrition*, *41*(1), 1–29. https://doi.org/10.1080/07315724.2020.1827082

Shashoa, N. A. A., Salem, N. A., Jleta, I. N., & Abusaeeda, O. (2017). Classification depend on linear discriminant analysis using desired outputs. *2016 17th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA)*, 328–332. DOI: 10.1109/STA.2016.7952041.

Shieh, W., & Djordjevic, I. (2010). OFDM for Optical Communications. In *Elsevier eBooks* (pp. 83–87). Academic Press. https://doi.org/10.1016/b978-0-12-374879-9.00003-4. https://doi.org/10.1016/B978-0-12-374879-9.00003-4.

Skauli, T. (2011). Sensor noise informed representation of hyperspectral data, with benefits for image storage and processing. *Optics Express*, *19*(14), 13031. https://doi.org/10.1364/oe.19.013031

Tea Research Institute. (2025). *Mid Country Regional Centre*. Tea Research Institute of Sri Lanka, Talawakelle. https://www.tri.lk/mid-country-regional-centre/

Tea Research Institute, Sri Lanka. (2003, September). *Manufacture of the tea by the CTC process*. (Serial No. 11/72 and T4, Serial No. 12/72).

Wang, H., Hu, L., Zhou, P., Ouyang, L., Chen, B., Li, Y., Chen, Y., Zhang, Y., & Zhou, J. (2021). Simultaneous determination of fructose, glucose and sucrose by solid phase extraction-liquid chromatography-tandem mass spectrometry and its application to source and adulteration analysis of sucrose in tea. *Journal of Food Composition and Analysis*, *96*, 103730. https://doi.org/10.1016/j.jfca.2020.103730

Wedagedara, H., Wanigasuriya, G., Nissanka , S. P., Mohotti, A., Mendis, E., & Botheju, W. S. (2019). Detection of sugar adulteration in black tea and its impact on quality parameters. *22nd Peradeniya University International Research Sessions (IPURSE)*, *22*, 151.

Wickramasinghe, W. A. N. D., Ekanayake, E. M. S. L. B., Wijedasa, M. A. C. S., Wijesinghe, A. D., Madhujith, T., Ekanayake, M. P. B., Godaliyadda, G. M. R. I., & Herath, H. M. V. R. (2021). Validation of multispectral imaging for the detection of sugar adulteration in black tea. *2021 10th International Conference on Information and automation for sustainability (ICIAfS)*, 494–499.

XU, Y., QIAO, F., & HUANG, J. (2022). Black tea markets worldwide: Are they integrated? *Journal of Integrative Agriculture*, *21*(2), 552–565. https://doi.org/10.1016/s2095-3119(21)63850-9

Xu, Y., Zhang, H., Zhang, C., Wu, P., Li, J., Xia, Y., & Fan, S. (2019). Rapid prediction and visualization of moisture content in single cucumber (Cucumis sativus L.) seed using hyperspectral imaging technology. *Infrared Physics & Technology*, *102*, 103034–103034. https://doi.org/10.1016/j.infrared.2019.103034

Yu, H., Qing, L.-W., Yan, D.-T., Xia, G., Zhang, C., Yun, Y., & Zhang, W. (2021). Hyperspectral imaging in combination with data fusion for rapid evaluation of tilapia fillet freshness. *Food Chemistry*, *348*, 129129–129129. https://doi.org/10.1016/j.foodchem.2021.129129

Zhang, R., Xu, P., Guo, L., Zhang, Y., Li, P., & Yao, D. (2013). Z-Score linear discriminant analysis for EEG Based brain-computer interfaces. *PLoS ONE*, *8*(9), e74433. https://doi.org/10.1371/journal.pone.0074433