Food Bioscience 59 (2024) 103946

Contents lists available at ScienceDirect

Food Bioscience

journal homepage: www.elsevier.com/locate/fbio

Geographical origin discrimination of lemon myrtle (*Backhousia citriodora*) leaf powder using near-infrared hyperspectral imaging

Maral Seididamyeh [a,*], Iman Tahmasbian [b,d], Anh Dao Thi Phan [a,c], Yasmina Sultanbawa [a,c]

[a] Centre for Nutrition and Food Sciences, Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St Lucia, QLD 4067, Australia
[b] Department of Agriculture and Fisheries, Queensland Government, Toowoomba, QLD 4350, Australia
[c] ARC Industrial Transformation Training Centre for Uniquely Australian Foods, Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Indooroopilly, QLD 4068, Australia
[d] Centre for Planetary Health and Food Security, Griffith Sciences, Griffith University, Nathan, Brisbane, QLD 4111, Australia

ARTICLE INFO

Keywords:
Backhousia citriodora
Geographical origin
Hyperspectral imaging
NIR
PLS-DA
SWIR

ABSTRACT

Lemon myrtle (LM), *Backhousia citriodora*, is a popular flavouring agent and herbal tea from Australia. To ensure traceability and consumer trust in global food supply chain, rapid and non-destructive tools are crucial. In this study, hyperspectral images were acquired from 91 L M samples sourced from three different origins (Malaysia, Queensland, and New South Wales (Australia)), within 950–2500 nm range. Classification models were developed using linear partial least squares-discriminant analysis (PLS-DA) with two approaches, pixel-based (trained by all spectral data points) and sample-based (trained by average spectrum). All models achieved classification accuracies above 96%. The sample-based PLS-DA model, trained by mean-centring transformed data, demonstrated the highest discriminatory performance. Both approaches show potential for LM origin classification, but the sample-based model is more suitable for automated and rapid industry applications due to its shorter calculation time. However, additional spectral data acquisition is necessary to improve the model and fully explore its capabilities and limitations.

1. Introduction

Backhousia citriodora F. Muell. (Lemon myrtle) belongs to the Myrtaceae family and is native to the rainforests of Queensland in Australia. The evergreen lemon myrtle trees usually reach 8 m in height and produce glossy green lanceolate leaves with creamy white flower clusters that appear in late spring to early summer (Mazzorana & Mazzorana, 2016). There are two essential-oil chemotypes of lemon myrtle namely citral and citronellal, the former is mainly used for flavouring and essential oil production and is commonly cultivated in Australia, while the latter is used for its insect-repellent properties (Archer, 2004).

The leaves have been traditionally used for culinary purposes as well as a herbal remedy for various infections such as respiratory, intestinal, and skin conditions (Mazzorana & Mazzorana, 2016). Several studies have demonstrated the potential health benefits of lemon myrtle such as antioxidant (Kang et al., 2020; Lim et al., 2022), antimicrobial (Lim et al., 2022), anti-inflammatory (Kang et al., 2020) and anti-diabetic (Jung et al., 2017) properties as well as hepatoprotective effects (Jung et al., 2017) and improving sarcopenia condition (Yamamoto et al., 2022). This can eventually pave its way into functional food, nutraceutical, pharmaceutical and cosmetic industries.

Recently, demand for lemon myrtle leaves and their products has been increasing in both domestic and international markets, where they are mainly used as a flavouring agent in different food products such as beverages, dairy, and bakery. Therefore, its commercial production has been extended by established plantations from north Queensland to northern New South Wales in Australia to supply this growing demand (Archer, 2004). However, the cultivation of lemon myrtle in other countries creates challenges for the industry to build consumer trust about the authenticity and provenance of the product in this expanding global market. Products originating from certain regions often have unique characteristics and qualities that are associated with that specific geographical area. Environmental factors such as soil composition and climate as well as the cultivation practices can lead to variations in chemical composition of the plant and therefore the aroma, flavour, and overall quality of the product. Hence, maintaining the integrity of regional products like lemon myrtle from Australia is essential for preserving consumer trust, ensuring product quality, and upholding the

* Corresponding author.
E-mail address: s.maral@uq.edu.au (M. Seididamyeh).

https://doi.org/10.1016/j.fbio.2024.103946
Received 12 January 2024; Received in revised form 18 March 2024; Accepted 20 March 2024
Available online 21 March 2024

reputation of the region's agricultural products.

The morphological differentiation between the products with different geographical origins may be difficult or even impossible by visual appearance, which is mainly due to alterations through different processing operations such as drying and grinding (Swetha et al., 2017). Several techniques have been studied to determine the authenticity and provenance of plant material including fingerprint studies of organic composition using high-performance liquid chromatography (HPLC) and gas chromatography coupled with mass spectrometry (GC-MS), DNA (deoxyribonucleic acid) sequencing and barcoding, and the inorganic elemental profile using inductively coupled plasma optical emission spectrometry (ICP-OES) (Fang et al., 2019; Su et al., 2019; Zhang et al., 2020). Although these techniques are accurate and reliable, they are time-consuming and still rely on well-equipped laboratories, expensive instruments and sampling techniques and are not suitable for rapid recognition of plant's origin and quality.

Nevertheless, on-site, rapid, non-destructive, and chemical-free techniques with minimal sample preparation are required to address the needs of traceability of products across the food value chain. While near infrared (NIR) spectroscopy has been used for identification purposes, it has certain limitations. For example, NIR spectroscopy fails to provide the information on spatial distribution of constituents within the sample since it measures one point of a sample and provides an averaged spectrum of the measured point. Measuring a single point necessitates using huge number of samples for calibration because the spectral variabilities existing within a sample will be overlooked, leading to reduced reliability of the models. Additionally, NIR spectroscopy predictions are unable to reveal physicochemical variabilities should they exist in the sample due to the absence of spatial measurements.

Hyperspectral imaging (HSI) is a computer vision system that integrates the regular imaging with the traditional NIR spectroscopy. HSI facilitates the simultaneous acquisition of both spatial (localisation) and spectral (identification) information of the entire load (Feng & Sun, 2012; Manley, 2014), and therefore allows for the identification and characterization of small-scale variations that might be overlooked by NIR spectroscopy, providing a more comprehensive understanding of the sample. Furthermore, the spatial information captured in the form of pixels via HSI enhances the statistical robustness of sample by measuring thousands to millions of detection points per sample, in contrast to the singular measurement obtained in NIR spectroscopy (Tahmasbian, Morgan, et al., 2021). This broader scope facilitates quicker and more reliable calibrations. The ability to measure spatial dimension rapidly enables scanning the entire load, which in turn minimizes/eliminates sampling error due to sampling techniques that otherwise is not practical when using NIR spectroscopy.

The potential of HSI for chemical detection and quantification as well as quality evaluation has been proven by several studies such as fruits (Kämper, et al., 2020), Darjeeling black tea (Firmani et al., 2019), Dianhong black tea (Ren et al., 2021), cumin powder (Florián-Huamán et al., 2022), and meat (Yao et al., 2019). HSI has been recently employed for provenance studies in different food products such as origin identification of loose-leaf tea (Mishra et al., 2018), green tea (Liu et al., 2022), wolfberries (Yin et al., 2017), saffron (Kiani et al., 2023), chia seeds (Choi et al., 2021), ginseng powder (Zhao, et al., 2021), white asparagus (Richter et al., 2019), pistachio (Vitale et al., 2013), cocoa beans (Mandrile et al., 2019), and mutton (Weng et al., 2021). To our knowledge, no prior HSI study has been conducted on lemon myrtle leaf powder as an herbal tea. Thus, this study represents the first attempt to provide NIR-HSI information on lemon myrtle leaf powders.

This study aimed to conduct a preliminary investigation into the potential use of NIR-HSI technology for distinguishing between commercial lemon myrtle tea products of various origins. Additionally, this study aimed to compare pixel-based and sample-based (average spectrum) calibration approaches for differentiation of the samples. We hypothesized that lemon myrtle samples sourced from different geographical areas exhibit distinct physicochemical differences, thereby influencing their spectral signatures and enabling differentiation. We also hypothesized that models trained using the spectra of individual pixels within a sample would outperform those trained using average spectra of samples, owing to the significantly larger statistical population utilized for model training.

2. Material and methods

2.1. Sample collection and preparation

The commercial lemon myrtle samples were provided by Australian Native Products Co. (NSW, Australia). A total of 91 samples were collected from main lemon myrtle-producing regions, of which 30, 31, and 30 samples were from Queensland (QLD) and New South Wales (NSW) in Australia, and from Kluang in Malaysia (MY), respectively. Samples were collected from different individual trees for each lemon myrtle origin during May–July 2022. The leaves were dried, ground, and packed in Aluminium airtight sealed pouches (GMD Packaging Pty Ltd., Queensland, Australia), followed by storing at room temperature (25 ± 2 °C). The drying condition for Australian samples was 55 °C for 3 h in a lab dehydrator and for Malaysian samples was by withering troughs, as is done for their commercial products. The moisture content of QLD, NSW, and MY samples were 10.28% ± 0.98%, 10.54% ± 0.99%, and 11.30% ± 0.26%, respectively and the differences were insignificant at $p > 0.05$. The samples were then delivered to the laboratory in October 2022 for analysis, where they were also stored at room temperature (25 ± 2 °C) until analysis.

2.2. Hyperspectral imaging system and image acquisition

Hyperspectral image acquisition was performed using a shortwave infrared (SWIR; 950–2500 nm) line-scan HSI system (HySpex, SWIR-384, Norsk Elektro Optikk, Oslo, Østlandet, Norway) mounted on a laboratory rack with a translation stage (Fig. 1). The images were captured in the full spectral range of 950–2500 nm at 5.45 nm spectral sampling intervals and with a spatial resolution of 384 pixels per line, and at room temperature (25 ± 2 °C). Samples were illuminated by two linear halogen light sources (100 W each) during imaging, which were positioned at two opposite sides of the camera and focused to illuminate a line overlapping with the camera's field of view (16°). These light sources were turned on 15 min prior to scanning to stabilise light-source temperature drift and spatial lighting uniformity. System operation and data acquisition were performed using the Breeze software (version 2022.1.0; Umeå, Prediktera, Sweden).

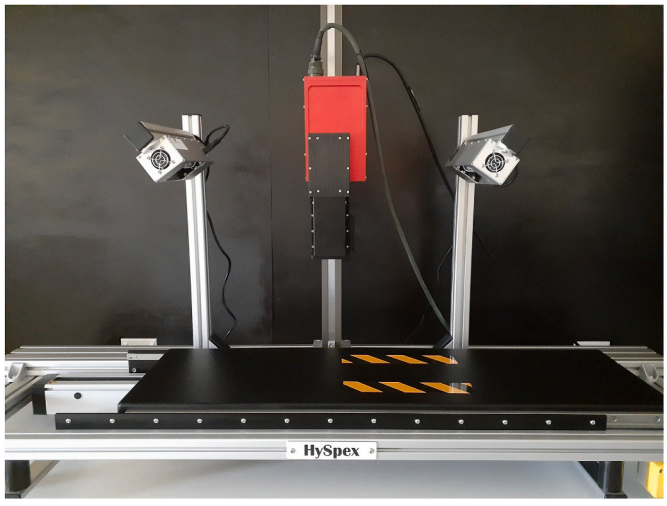

Fig. 1. The shortwave infrared hyperspectral imaging system used in this study.

Lemon myrtle samples in polyethylene sealable bags (65 × 75 mm; PPS, China; 2 cm depth) were placed on the black conveyor belt to acquire HSI data. Prior to image acquisition, the influential parameters on image quality were tested and set as follows: conveyor belt speed at 13.021 mm/s, integration time at 12 ms, and frame rate at 50 frames/s. The spectral analysis of the polyethylene bags did not show any measurable effect on the spectral data of the samples in the measured spectral range.

The background noises by dark current of camera, illumination source and environmental factors were corrected using the black and white references. The black reference image was captured before and after each sample image by automatically closing the camera shutter, and then averaging them. The white reference image was acquired by using a white diffuse reflectance standard (Zenith) of 50% reflectance to correct the variation in the illumination. The raw hyperspectral images were then corrected and transformed into reflectance hyperspectral images using the following equation (Park & Lu, 2015):

$$R = \frac{R_0 - D}{W - D} \quad (1)$$

where R is the corrected reflectance image, R_0 is the raw hyperspectral image, D is the dark reference image, and W is the white reference image.

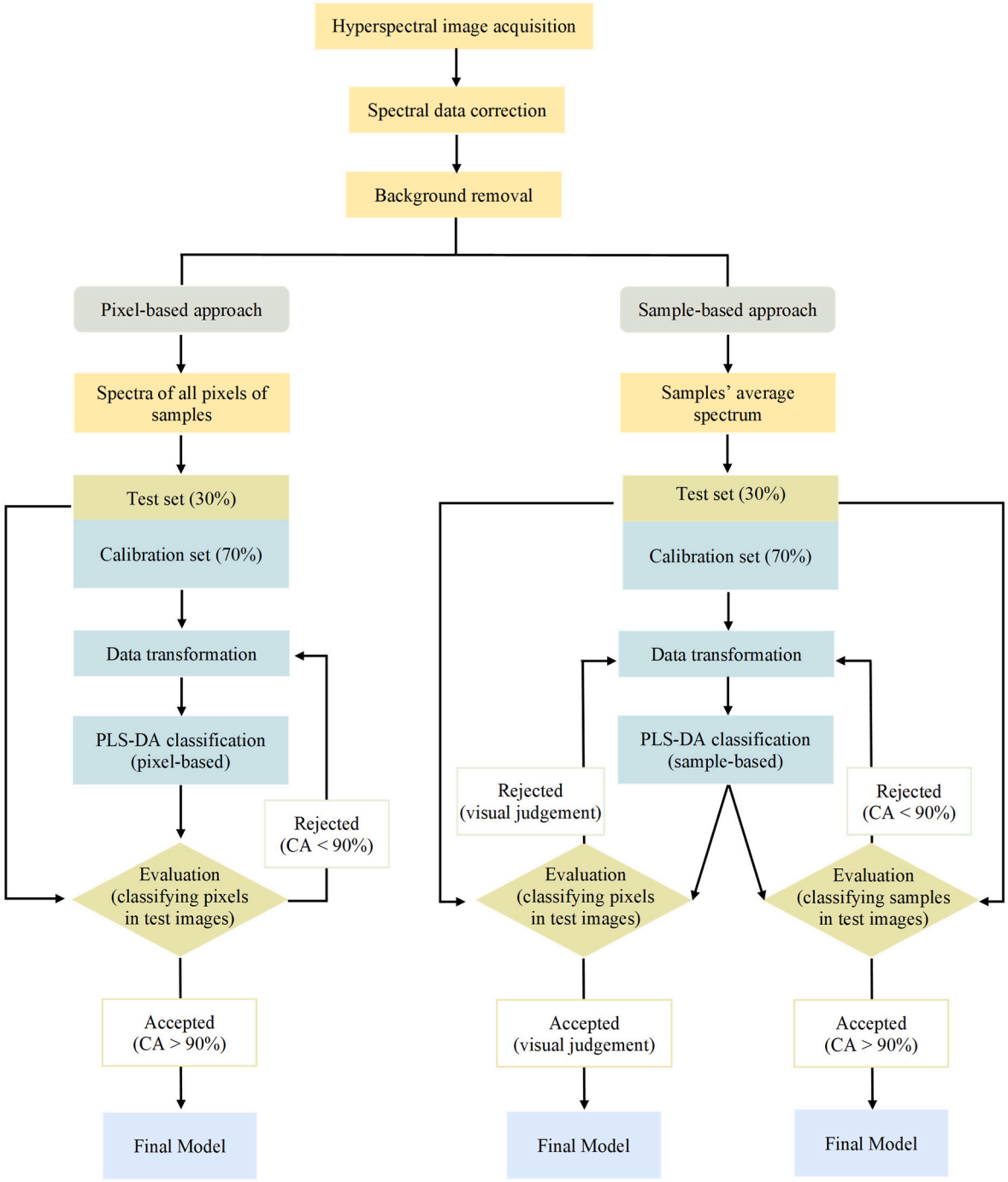

Fig. 2. Workflow of hyperspectral image processing and analysis using pixel- and sample-based approaches.

2.3. Spectral data extraction

The corrected hyperspectral images were loaded into Evince software (Version 2.7.17, Umeå, Prediktera, Sweden) to extract the spectral data. To facilitate the image processing, the image size was initially reduced by Evince resolution-reduction function, which selected data from every three rows and columns and every two wavelengths. The background and saturated pixel (if any) were removed using a principal component analysis (PCA) model with three components (Nturambirwe et al., 2021; Tahmasbian, Morgan, et al., 2021). The spectral data of each pixel within these images (approximately 5218 spectra) were then extracted and stored in the software for further analysis.

2.4. Data pre-processing and PLS-DA modelling

Partial least-square discrimination analysis (PLS-DA) algorithm was used as a supervised classification approach to identify and differentiate the lemon myrtle samples according to their geographical origins, in which full spectral data (950–2500 nm) extracted from samples were used as predictors. Initially, different pre-processing algorithms were applied and investigated individually or in combination to improve the accuracy of classification (Farrar et al., 2021; Khamsopha et al., 2021; Tahmasbian, Wallace, et al., 2021). These algorithms were applied to eliminate the potential artefacts, to correct non-linear behaviour, and to improve the data quality. The algorithms used included standard normal variate (SNV), multiple scatter correction (MSC), 1st and 2nd derivatives (DVT). All the data were subjected to mean centring (MC) prior to multivariate data analysis, regardless of the applied pre-treatment.

The spectral datasets were randomly divided into a calibration dataset containing 70% of total samples to develop the classification models and an external validation dataset containing the remaining 30% of samples to evaluate their performance. These datasets were used to train the classification models in two different approaches, namely pixel-based and sample-based (Fig. 2), to identify the most robust training approach for classification at pixel and sample levels. The former was carried out by applying the spectra extracted from all pixels of each image (i.e., approximately 5218 spectra per sample), and the latter was carried out by applying the average spectrum of all pixels of each image (i.e., one spectrum per sample). It is expected that more robust training models can be obtained using the pixel-based approach due to its larger datasets, although it is a more time-consuming training process and requires powerful computer processors. In contrast, using a smaller dataset in sample-based approach results in a simplified and faster training, however, it can cause difficulties in pixel classification for example to identify the impurities and integrities in the studied samples.

To avoid overfitting, the number of latent variables (LV) was adjusted by cross-validating the calibration models using a k-fold (k = 5) method (Kohavi, 1995). The optimal number of LVs was determined according to the highest coefficient of determination in cross-validation (R^2_{CV}) and the lowest number of LVs that resulted in the highest classification accuracy.

2.5. Model evaluation

The classification performance of the developed PLS-DA models was evaluated using the external validation dataset and the metrics described in the following equations (Sokolova et al., 2006, pp. 1015–1021; Tahmasbian et al., 2024), using the external validation dataset:

$$\text{Classification accuracy } (\%) = \frac{TP + TN}{TP + FP + TN + FN} \times 100 \quad (2)$$

$$\text{False positive error } (\%) = \frac{FP}{TP + FP + TN + FN} \times 100 \quad (3)$$

$$\text{False negative error } (\%) = \frac{FN}{TP + FP + TN + FN} \times 100 \quad (4)$$

$$\text{Sensitivity } (\%) = \frac{TP}{TP + FN} \times 100 \quad (5)$$

$$\text{Specificity } (\%) = \frac{TN}{TN + FP} \times 100 \quad (6)$$

$$\text{Precision } (\%) = \frac{TP}{TP + FP} \times 100 \quad (7)$$

where TP is true positives, TN is true negatives, FP is false positives, and FN is false negatives. The multivariate data analysis of the hyperspectral images was performed using the Evince software (Version 2.7.17, Prediktera).

3. Results and discussion

3.1. Spectral features

The mean spectra of lemon myrtle samples from three different geographical origins are illustrated in Fig. 3. The analysis of the mean spectra revealed a relatively similar spectral pattern in terms of peak shape and positions, which can be indicative of the unique phytochemical composition of *B. citriodora*. Notably, the spectra showed six characteristic absorption peaks (Fig. 3) that are integral to understanding the plant's molecular signatures. In general, the spectra in the near infrared range are relatively related to chemical compounds (Sun et al., 2021).

The peak centring at 1212 nm can be related to C–H stretching second overtones, reflecting the presence of aliphatic hydrocarbons within the samples (Mark & Workman, 2010; Qiu et al., 2018), which are common in plant oils and waxes. The absorption band at 2313 nm can be attributed to C–H bending and symmetric stretching (Lee et al., 2014), which could be characteristic of aromatic compounds and alkenes, suggesting the presence of aromatic rings that are fundamental in polyphenols and flavonoids. Moreover, the peak near 2357 nm suggests the presence of C–H vibrations specific to $–CH_3$ groups (Pu et al., 2008), which could be attribted to methylated phenolic compounds. The peak at 1736 nm is indicative of C–H first overtone vibrations (Tao et al., 2019), which can be associated with compounds that have $–CH_2–$ stretching vibrations, indicative of long-chain hydrocarbons or fatty acids that are common in plant lipids and waxes. The peak at 1768 nm can be related to O–H stretching first overtones (Posom et al., 2017), which is common in hydroxyl groups of alcohols, phenols and carboxylic acids suggesting the presence of polysaccharides and polyphenols. The peak at 1452 nm denotes the combination vibration of O–H stretching and H–O–H transformation, which is characteristic of water but can also indicate the presence of other hydroxyl containing compounds such as polysaccharides, polyphenols, and certain amino acids (Mark & Workman, 2010). The peak near 1932 nm can be associated with C═O band overtones, highlighting the presence of carbonyl groups (Qiu et al., 2018), which is typical for aromatic compounds, flavonoids, and certain amino acids. It may also indicate the presence of ketones, aldehydes, and acids within the phytochemical composition of a plant (Xiong et al., 2015). The spectra, therefore, suggest the presence of a wide range of phytochemical compounds in lemon myrtle samples, including polyphenols, flavonoids, aromatic compounds, polysaccharides, and amino acids.

There were however differences in reflectance intensity across the measured electromagnetic spectrum (950–2500 nm) among the samples from QLD, MY, and NSW (Fig. 3). The average spectrum of MY samples showed higher intensity levels compared to those from NSW and QLD, potentially due to variations in the drying process and therefore its effect on the chemical composition of lemon myrtle samples. Although there

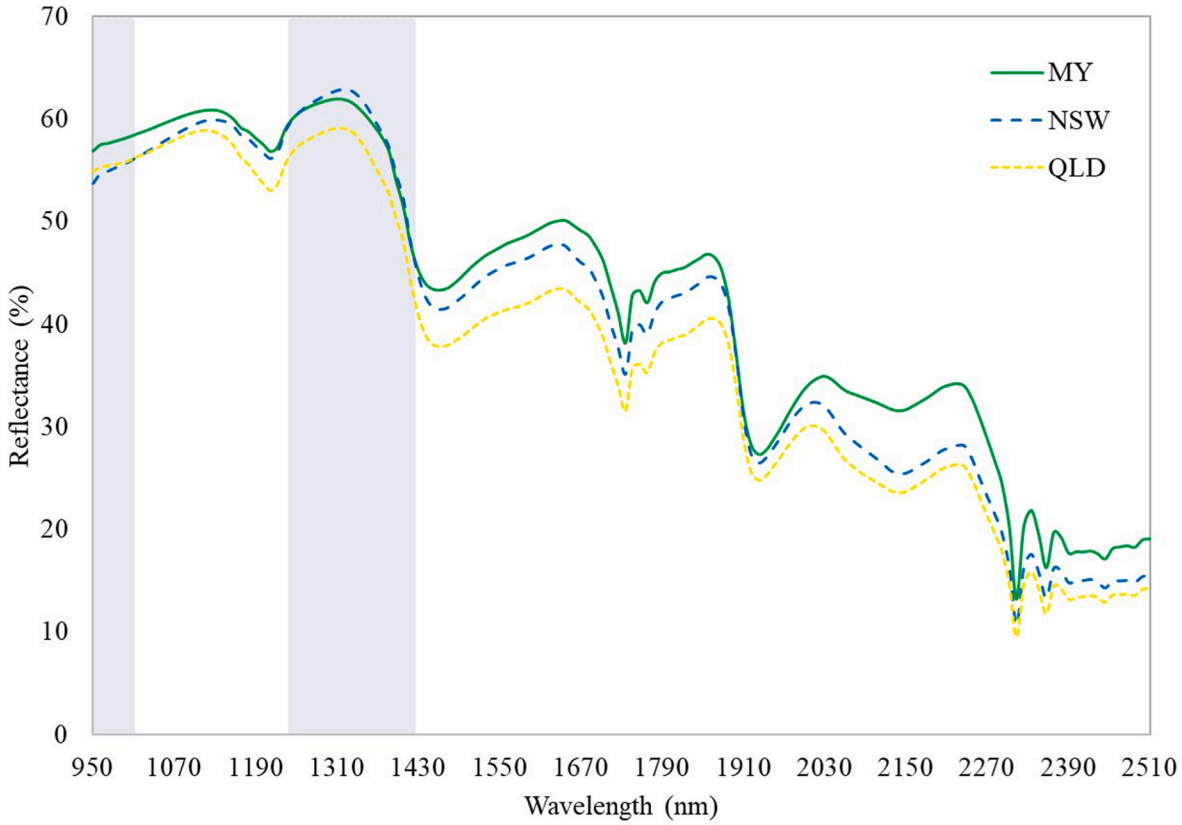

Fig. 3. Average raw spectral reflectance of the lemon myrtle samples from Malaysia (MY), New South Wales (NSW) and Queensland (QLD).

were shifts in the spectral pattern that are specified with vertical grey lines in Fig. 3. Specifically, the NSW samples showed lower intensity at the 950–1005 nm band compared to QLD samples, and higher intensity from 1245 to 1419 nm than MY samples. These intensity differences can be related to the geographical origins that influenced the chemical composition of plants and their relative concentrations through environmental factors (Granato et al., 2016; Long et al., 2023) as well as processing methods. The chemical variabilities measured using spectral signatures can therefore be employed to differentiate them in the global market. Notwithstanding, the spectral overlaps and crossovers as well as the complexity of lemon myrtle's phytochemical system did not allow to differentiate the geographical origins by spectra only. Thus, further analysis of spectral data was required to accurately classify the origin of lemon myrtle samples.

3.2. PLS-DA classification models

The PLS-DA classification models were established to discern the geographical origin of lemon myrtle samples. To establish the pixel-based PLS-DA model, different transformation algorithms were applied to the raw full spectral data to determine the optimal transformation method. As is shown inTable S1, the spectral overall classification accuracy was greater than 93% regardless of the transformation algorithm used, which demonstrates the ability of established models to easily differentiate samples of different geographical origins. However, a comparative analysis among the transformation algorithms revealed that the PLS-DA model achieved the highest overall classification accuracy of 94.6% with the MC and SNV transformations. Therefore, these transformations were selected to develop the pixel-based PLS-DA model, with the optimal number of LVs (=6) determined based on the highest R^2 (=0.78) in the cross-validation step.

In contrast, the sample-based models simplified the analysis and modelling process by averaging spectral data for each sample and reducing the size of hyperspectral datasets. Sample-based PLS-DA models successfully discriminated the three geographical origins in test images with 100% classification accuracy, regardless of applied transformations (Figs. 4 and 5). Therefore, the quantitative comparison (i.e., confusion matrix) was not reported for sample-based models. The optimal number of LVs was determined to be 4, aiming to simplify the model while maintaining high accuracy, thereby demonstrating its efficiency in classifying samples with reduced computational complexity. This was further validated by the clear separation of geographical origins in the PLS-DA score plots.

3.3. Model evaluation

A confusion matrix was used to assess the predictive ability of the developed pixel-based PLS-DA model in classifying lemon myrtle external validation datasets (Table 1). The model exhibited a classification accuracy higher than 96% for all three classes, demonstrating its substantial classification capability. Additionally, sensitivity, specificity, and precision were calculated to further assess model performance. While most predicted pixels were classified correctly, there were instances of misclassification and unclassified pixels. The model showed a slightly better classification efficiency for MY and NSW samples compared to QLD, as evidenced by higher accuracy and sensitivity values (Table 1). Despite higher accuracy for MY samples, lower specificity, and precision values along with a higher false positive error rate (98.6%, 97.3%, and 0.9%, respectively) suggested a bias towards false positive predictions for MY samples. Conversely, the lower classification accuracy observed for QLD samples (96.2%) was associated with a higher misclassification rate, lower sensitivity, and higher false negative error, where the latter could be attributed to the unclassified pixels. Precision values indicated a more reliable classification of true positives for QLD and NSW samples compared to MY samples (98.8%, 99.0%, and 97.3%, respectively), emphasizing the model's ability to accurately

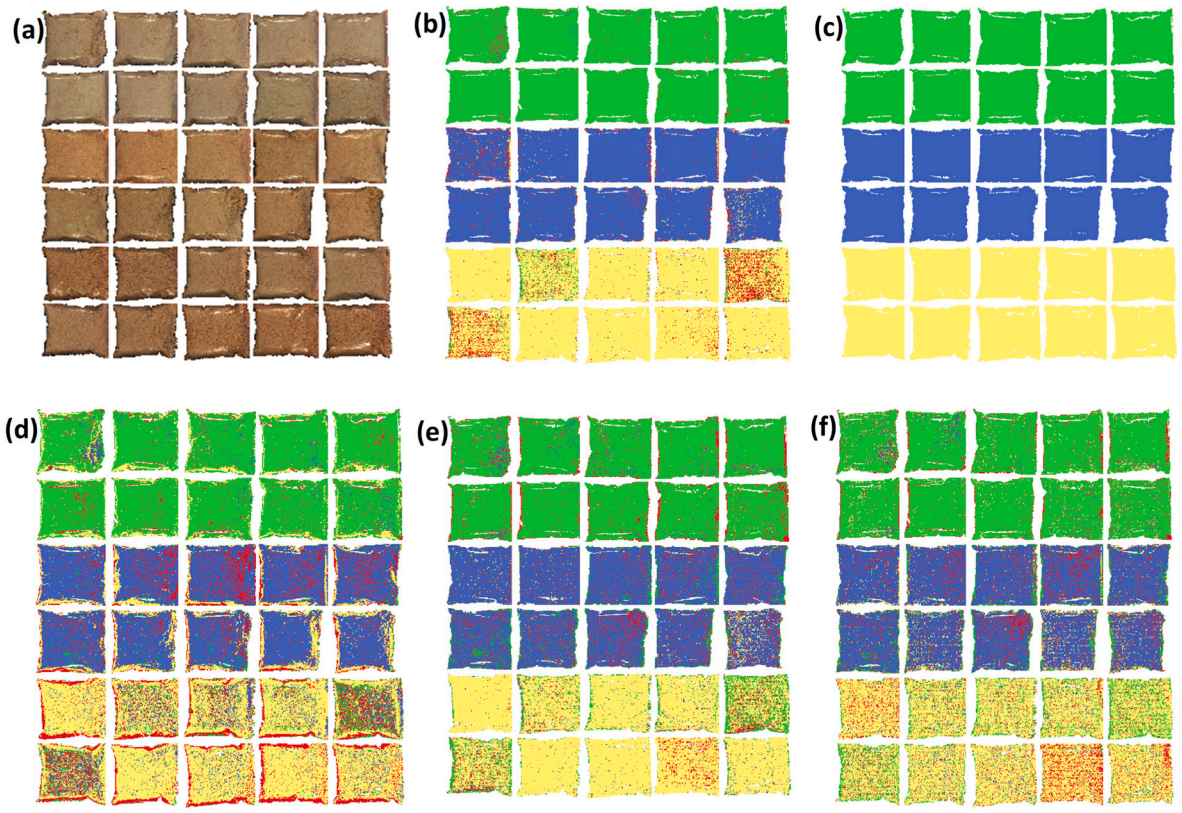

Fig. 4. Classification of lemon myrtle samples based on their geographical origins using PLS-DA analysis: (a) pseudo RGB images, (b) pixel classification by pixel-based PLS-DA model trained using MC + SNV transformed data, (c) sample classification by sample-based PLS-DA model trained using MC transformed data, and pixel classification by sample-based PLS-DA model trained using (d) MC, (e) MC + SNV, and (f) MC + SNV+2nd DVT transformed data. MC, SNV and 2nd DVT represent mean centring, standard normal variate and second derivative, respectively. The green, blue, and yellow colours represent pixels predicted as Malaysian (MY), New South Wales (NSW) and Queensland (QLD) samples, respectively, and the red colour shows unclassified pixels. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

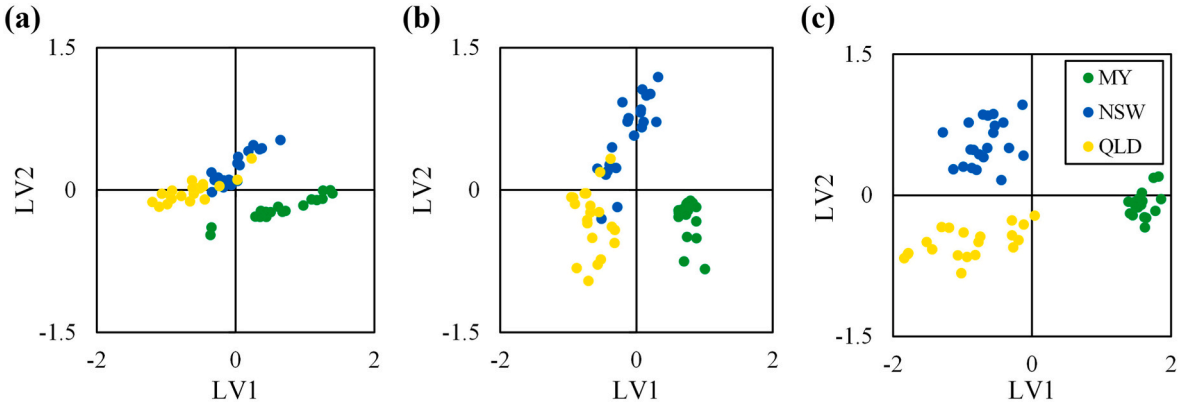

Fig. 5. PLS-DA scores plot (LV1 vs. LV2) representing the clustering of the lemon myrtle external validation spectra transformed using (a) mean-centring (MC), (b) MC + standard normal variate (SNV) and (c) MC + SNV + second derivative (2nd-DVT).

group tested samples (Cuadros-Rodríguez et al., 2020; Huang et al., 2022), which is thus crucial in food authentication applications.

The impracticality of full spectra application due to the need for massive data analysis (Kamruzzaman et al., 2015) suggests the utility of HSI in acquiring spectral information to generate classification maps. Therefore, the optimal model was applied to hyperspectral images of lemon myrtle samples from different geographical origins, generating a prediction map (Fig. 4). The original RGB images obtained by the hyperspectral camera did not provide any information about the geographical origin of samples (Fig. 4a). However, the developed pixel-based PLS-DA model using MC + SNV transformed hyperspectral data facilitated clear discrimination between images, assigning each geographical class to respective pixels in the prediction map (Fig. 4b). The visualization map effectively showed correctly classified, mis-classified, and unclassified pixels, where most of the pixels within one sample were accurately assigned to their geographical classes. Although a distribution of unclassified pixels was observed in images (mostly QLD images), over 89% of the pixels were correctly identified.

On the other hand, the sample-based models developed using different data transformations were evaluated through PLS-DA score

Table 1
Confusion matrix and classification performance of the MC + SNV-transformed pixel-based PLS-DA model evaluated using the external validation dataset.

		Actual Classes			CA (%)	FPE (%)	FNE (%)	Sensitivity (%)	Specificity (%)	Precision (%)	OCA (%)
		MY	NSW	QLD							
Predicted Classes (no. of pixels)	MY	**52103**	24	1441	98.6	0.9	0.5	98.6	98.6	97.3	94.6
	NSW	186	**50776**	326	98.1	0.3	1.6	95.4	99.5	99.0	
	QLD	32	517	**45369**	96.2	0.4	3.4	89.5	99.5	98.8	
	No Class	501	1889	3572							

Malaysian, New South Wales, and Queensland lemon myrtle samples are presented as MY, NSW and QLD, respectively. No class is for unclassified pixels.
MC, mean centring; SNV, standard normal variate; CA, classification accuracy; FPE, false positive error; FNE, false negative error; OCA, overall classification accuracy.
Bold values in confusion matrix represent samples classified correctly.

plots (Fig. 5). The score plot (LV1 vs. LV2) demonstrated clear separation of MY samples from other origins along the first predictive component, particularly when "MC + SNV" and "MC + SNV +2nd DVT" combinations were used for data transformations (Fig. 5b and c). The predictive ability of sample-based models was further evaluated by generating classification map on test images (Fig. 4c–f). The sample-based model developed by MC-transformed data was used to classify the test images in two different modes, which were by predicting the samples (i.e., average spectrum of each sample; Fig. 4c) and the pixels (Fig. 4d). Due to software limitation, we were unable to quantify the pixels classified using sample-based models. The visual judgement, however, demonstrates the ability of the developed model to clearly discriminate the geographical origin of the samples using their average spectrum. Although the sample-based model exhibited relatively poor performance in predicting individual pixels, which was particularly evident in the confusion between QLD and MY pixels (Fig. 4d–f). This may be explained by the comparable climatic conditions found in Malaysia (tropical) and Queensland (Australia; tropical/subtropical), which need to be further studied. Notably, the model developed with MC + SNV transformations showed fewer misclassified and unclassified pixels compared to other models (Fig. 4d–f). The visualization map revealed the higher discrimination ability and predictability of the sample-based PLS-DA model developed by MC transformed average spectra (Fig. 4c), which was followed by the pixel-based model (Fig. 4b).

4. Discussion

The application of PLS-DA models, through different transformation algorithms, showed significant potential in distinguishing lemon myrtle samples based on their geographical origins with high accuracy. Both pixel-based and sample-based models demonstrated robust classification capabilities, albeit with strengths and constraints. Despite being data-intensive, the pixel-based model's strength lies in its detailed classification accuracy, which enables precise geographical origin determination at agranular level as was observed in the predication map as well. However, potential biases – evidenced by a higher false positive rate for MY samples and a lower sensitivity for QLD samples – indicated the model's limitations and areas for improvement.

The sample-based models, on the other hand, streamlined the data analysis process by reducing data complexity and analysis time, and excelled in overall classification accuracy. This simplification did not compromise the model's ability to distinguish between geographical origins, although less granular than pixel-based approaches, which was evidenced by the PLS-DA score plots and classification maps. The models' performance in separating the MY samples from other Australian origins with clear distinctions between the three classes highlighted their effectiveness. Our results showed better classification accuracy compared to those reported on using PLS-DA models for geographical origin discrimination of green tea with 92.5% accuracy (Liu et al., 2022), Longjing tea with 91.98% accuracy (Hong & He, 2020). Nevertheless, sample-based approach suggests a highly effective model for broader classification tasks and is potentially more suitable for rapid screening.

A challenge in both models was the confusion between QLD and MY samples, likely due to the spectral similarities caused by relatively similar climatic conditions. The observation of misclassified pixels in the test images, particularly between QLD and MY samples, indicates challenges in dealing with overlapping spectral features. This hurdle has been also observed in previous studies such as using ECOC-SVM (error correcting output codes-support vector machines) models to discriminate six different commercial tea products (Mishra et al., 2018). The results from using different data transformation combinations suggest that specific transformations may enhance the model's discriminative power, further emphasizing the need for careful selection of pre-processing techniques in developing robust classification models. This points to the need for further research into additional discriminative features such as spectral and extrinsic factors or the integration of advanced pre-processing methods to reduce misclassification and improve model performance. These advancements will reinforce the use of PLS-DA models combined with hyperspectral imaging as a non-destructive and rapid method for food provenance determination. The ability of these models to visually represent the geographical origins of samples demonstrated their practical utility and the effectiveness of the transformation algorithms applied, despite the observed limitations.

These findings affirm the practical applicability of PLS-DA models in food provenance studies. However, the variation in model performance and the challenges in classification highlight the importance of ongoing model evaluation in identifying and addressing limitations, and adaptation to specific analytical needs to ensure reliable and accurate classification outcomes. Continued research and model refinement are essential for harnessing the full potential of these analytical methods, offering a promising alternative to traditional methods through enhanced efficiency and accuracy in food authentication.

5. Conclusion

This study demonstrates the potential of combining HSI with multivariate data analysis to determine the origin of lemon myrtle leaf powders rapidly and non-destructively. The classification models were developed by two approaches including pixel-based (using all spectra) and sample-based (using average spectrum), which exhibited over 96% accuracy in distinguishing the origin of lemon myrtle samples. The sample-based model was found to be more robust in classifying results when built with MC-transformed datasets, compared to other combination transformations and the pixel-based model. Since average spectra is used to develop the sample-based model, data handling complexity and calculation time is therefore reduced. Overall, this technique is promising for traceability applications due to its automation, non-destructive nature, rapidity, and objectivity. However, to build a robust and proper calibrated classification model, additional spectral data collection is necessary. This would result in rapid sample identification with acquisition of HSI data, which is beneficial for industrial applications.

CRediT authorship contribution statement

Maral Seididamyeh: Writing – original draft, Methodology, Investigation, Conceptualization. **Iman Tahmasbian:** Writing – review & editing, Visualization, Methodology, Formal analysis. **Anh Dao Thi Phan:** Writing – review & editing. **Yasmina Sultanbawa:** Writing – review & editing, Resources, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The authors acknowledge the Traditional Owners of the lands on which the *Backhousia citriodora* was harvested and respect the knowledge and experience the Traditional Owners hold regarding the care, harvest, and use of this plant. Authors would also like to acknowledge Kirsty Langdon (Australian Native Products Co., NSW, Australia) for her assistance with sample collection. This research was funded by the University of Queensland, Vice Chancellor Strategic Initiative Funding for the project titles "reimagining intellectual property to build the bioeconomy".

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fbio.2024.103946.

References

Archer, D. (2004). *Backhousia citriodora F. Muell. Lemon scented myrtle: Biology, cultivation and exploitation*. Toona: Sutton, Australia.

Choi, J.-Y., Kim, H.-C., & Moon, K.-D. (2021). Geographical origin discriminant analysis of Chia seeds (*Salvia hispanica* L.) using hyperspectral imaging. *Journal of Food Composition and Analysis, 101*, Article 103916.

Cuadros-Rodríguez, L., Valverde-Som, L., Jiménez-Carvelo, A. M., & Delgado-Aguilar, M. (2020). Validation requirements of screening analytical methods based on scenario-specified applicability indicators. *TrAC, Trends in Analytical Chemistry, 122*, Article 115705.

Fang, S., Ning, J., Huang, W.-J., Zhang, G., Deng, W.-W., & Zhang, Z. (2019). Identification of geographical origin of Keemun black tea based on its volatile composition coupled with multivariate statistical analyses. *Journal of the Science of Food and Agriculture, 99*(9), 4344–4352.

Farrar, M. B., Wallace, H. M., Brooks, P., Yule, C. M., Tahmasbian, I., Dunn, P. K., & Hosseini Bai, S. (2021). A performance evaluation of Vis/NIR hyperspectral imaging to predict curcumin concentration in fresh turmeric rhizomes. *Remote Sensing, 13*(9), 1807.

Feng, Y.-Z., & Sun, D.-W. (2012). Application of hyperspectral imaging in food safety inspection and control: A review. *Critical Reviews in Food Science and Nutrition, 52*(11), 1039–1058.

Firmani, P., De Luca, S., Bucci, R., Marini, F., & Biancolillo, A. (2019). Near infrared (NIR) spectroscopy-based classification for the authentication of Darjeeling black tea. *Food Control, 100*, 292–299.

Florián-Huamán, J., Cruz-Tirado, J., Barbin, D. F., & Siche, R. (2022). Detection of nutshells in cumin powder using NIR hyperspectral imaging and chemometrics tools. *Journal of Food Composition and Analysis, 108*, Article 104407.

Granato, D., de Magalhães Carrapeiro, M., Fogliano, V., & van Ruth, S. M. (2016). Effects of geographical origin, varietal and farming system on the chemical composition and functional properties of purple grape juices: A review. *Trends in Food Science & Technology, 52*, 31–48.

Hong, Z., & He, Y. (2020). Rapid and nondestructive discrimination of geographical origins of Longjing tea using hyperspectral imaging at two spectral ranges coupled with machine learning methods. *Applied Sciences, 10*(3), 1173.

Huang, J., He, H., Lv, R., Zhang, G., Zhou, Z., & Wang, X. (2022). Non-destructive detection and classification of textile fibres based on hyperspectral imaging and 1D-CNN. *Analytica Chimica Acta, 1224*, Article 340238.

Jung, K. I., Kim, P. K., Gal, S. W., & Choi, Y. J. (2017). The hepatoprotective effects of Hep G2 cells and the alcohol-metabolizing enzyme activities of lemon-myrtle (*Backhousia citriodora*) leaf extracts. *Journal of Life Science, 27*(11), 1262–1268.

Kämper, W., Trueman, S. J., Tahmasbian, I., & Bai, S. H. (2020). Rapid determination of nutrient concentrations in Hass avocado fruit by Vis/NIR hyperspectral imaging of flesh or skin. *Remote Sensing, 12*(20), 3409.

Kamruzzaman, M., Makino, Y., Oshita, S., & Liu, S. (2015). Assessment of visible near-infrared hyperspectral imaging as a tool for detection of horsemeat adulteration in minced beef. *Food and Bioprocess Technology, 8*, 1054–1062.

Kang, E.-J., Lee, J.-K., Park, H.-R., Kim, H., Kim, H.-S., & Park, J. (2020). Antioxidant and anti-inflammatory activities of phenolic compounds extracted from lemon myrtle (*Backhousia citriodora*) leaves at various extraction conditions. *Food Science and Biotechnology, 29*(10), 1425–1432.

Khamsopha, D., Woranitta, S., & Teerachaichayut, S. (2021). Utilizing near infrared hyperspectral imaging for quantitatively predicting adulteration in tapioca starch. *Food Control, 123*, Article 107781.

Kiani, S., Yazdanpanah, H., & Feizy, J. (2023). Geographical origin differentiation and quality determination of saffron using a portable hyperspectral imaging system. *Infrared Physics & Technology, 131*, Article 104634.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *International joint conference on artificial intelligence (IJCAI)* (pp. 1137–1145). Canada.

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 Chemistry, 158*, 351–357.

Lim, A. C., Tang, S. G. H., Zin, N. M., Maisarah, A. M., Ariffin, I. A., Ker, P. J., & Mahlia, T. M. I. (2022). Chemical composition, antioxidant, antibacterial, and antibiofilm activities of *Backhousia citriodora* essential oil. *Molecules, 27*(15), 4895.

Liu, Y., Huang, J., Li, M., Chen, Y., Cui, Q., Lu, C., Wang, Y., Li, L., Xu, Z., & Zhong, Y. (2022). Rapid identification of the green tea geographical origin and processing month based on near-infrared hyperspectral imaging combined with chemometrics. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 267*, Article 120537.

Long, W., Wang, S.-R., Suo, Y., Chen, H., Bai, X., Yang, X., Zhou, Y.-P., Yang, J., & Fu, H. (2023). Fast and non-destructive discriminating the geographical origin of Hangbaiju by hyperspectral imaging combined with chemometrics. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 284*, Article 121786.

Mandrile, L., Barbosa-Pereira, L., Sorensen, K. M., Giovannozzi, A. M., Zeppa, G., Engelsen, S. B., & Rossi, A. M. (2019). Authentication of cocoa bean shells by near- and mid-infrared spectroscopy and inductively coupled plasma-optical emission spectroscopy. *Food Chemistry, 292*, 47–57.

Manley, M. (2014). Near-infrared spectroscopy and hyperspectral imaging: Non-destructive analysis of biological materials. *Chemical Society Reviews, 43*, 8200.

Mark, H., & Workman, J. J. (2010). *Chemometrics in spectroscopy*. Elsevier.

Mazzorana, G., & Mazzorana, M. (2016). Cultivation of lemon myrtle (*Backhousia citriodora*). In Y. Sultanbawa, & F. Sultanbawa (Eds.), *Traditional herbal medicines for modern times Australian native plants cultivation and uses in the health and food industries* (1st ed., pp. 113–126). CRC Press Taylor and Francis.

Mishra, P., Nordon, A., Tschannerl, J., Lian, G., Redfern, S., & Marshall, S. (2018). Near-infrared hyperspectral imaging for con-destructive classification of commercial tea products. *Journal of Food Engineering, 238*, 70–77.

Nturambirwe, J. F. I., Perold, W. J., & Opara, U. L. (2021). Classification learning of latent bruise damage to apples using shortwave infrared hyperspectral imaging. *Sensors, 21*(15), 4990.

Park, B., & Lu, R. (2015). *Hyperspectral imaging technology in food and agriculture*. Springer.

Posom, J., Saechua, W., & Sirisomboon, P. (2017). Evaluation of pyrolysis characteristics of milled bamboo using near-infrared spectroscopy. *Renewable Energy, 103*, 653–665.

Pu, Y., Ragauskas, A. J., Lucia, L. A., Naithani, V., & Jameel, H. (2008). Near-infrared spectroscopy and chemometric analysis for determining oxygen delignification yield. *Journal of Wood Chemistry and Technology, 28*(2), 122–136.

Qiu, G., Lü, E., Lu, H., Xu, S., Zeng, F., & Shui, Q. (2018). Single-kernel FT-NIR spectroscopy for detecting supersweet corn (*Zea mays* L. saccharata sturt) seed viability with multivariate data analysis. *Sensors, 18*(4), 1010.

Ren, G., Wang, Y., Ning, J., & Zhang, Z. (2021). Evaluation of Dianhong black tea quality using near-infrared hyperspectral imaging technology. *Journal of the Science of Food and Agriculture, 101*, 2135–2142.

Richter, B., Rurik, M., Gurk, S., Kohlbacher, O., & Fischer, M. (2019). Food monitoring: Screening of the geographical origin of white asparagus using FT-NIR and machine learning. *Food Control, 104*, 318–325.

Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006). *Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation*. Springer Berlin Heidelberg.

Su, H., Wu, W., Wan, X., & Ning, J. (2019). Discriminating geographical origins of green tea based on amino acid, polyphenol, and caffeine content through high-performance liquid chromatography: Taking Lu'an guapian tea as an example. *Food Science and Nutrition, 7*(6), 2167–2175.

Sun, Y., Li, Y., Pan, L., Abbas, A., Jiang, Y., & Wang, X. (2021). Authentication of the geographic origin of Yangshan region peaches based on hyperspectral imaging. *Postharvest Biology and Technology, 171*, Article 111320.

Swetha, V., Parvathy, V., Sheeja, T., & Sasikumar, B. (2017). Authentication of *Myristica fragrans* Houtt. using DNA barcoding. *Food Control, 73*, 1010–1015.

Tahmasbian, I., McMillan, M. N., Kok, J., & Courtney, A. J. (2024). Underwater hyperspectral imaging technology has potential to differentiate and monitor scallop populations. *Reviews in Fish Biology and Fisheries, 34*, 371–383.

Tahmasbian, I., Morgan, N. K., Hosseini Bai, S., Dunlop, M. W., & Moss, A. F. (2021). Comparison of hyperspectral imaging and near-infrared spectroscopy to determine nitrogen and carbon concentrations in wheat. *Remote Sensing, 13*(6), 1128.

Tahmasbian, I., Wallace, H. M., Gama, T., & Bai, S. H. (2021). An automated non-destructive prediction of peroxide value and free fatty acid level in mixed nut samples. *LWT-Food Science and Technology, 143*, Article 110893.

Tao, F., Yao, H., Hruska, Z., Kincaid, R., Rajasekaran, K., & Bhatnagar, D. (2019). Potential of near-infrared hyperspectral imaging in discriminating corn kernels infected with aflatoxigenic and non-aflatoxigenic *Aspergillus flavus*. *Sensing for Agriculture and Food Quality and Safety, XI*, Article 1101603.

Vitale, R., Bevilacqua, M., Bucci, R., Magrì, A. D., Magrì, A. L., & Marini, F. (2013). A rapid and non-invasive method for authenticating the origin of pistachio samples by NIR spectroscopy and chemometrics. *Chemometrics and Intelligent Laboratory Systems, 121*, 90–99.

Weng, S., Guo, B., Du, Y., Wang, M., Tang, P., & Zhao, J. (2021). Feasibility of authenticating mutton geographical origin and breed via hyperspectral imaging with effective variables of multiple features. *Food Analytical Methods, 14*, 834–844.

Xiong, C., Liu, C., Pan, W., Ma, F., Xiong, C., Qi, L., Chen, F., Lu, X., Yang, J., & Zheng, L. (2015). Non-destructive determination of total polyphenols content and classification of storage periods of Iron Buddha tea using multispectral imaging system. *Food Chemistry, 176*, 130–136.

Yamamoto, A., Honda, S., Ogura, M., Kato, M., Tanigawa, R., Fujino, H., & Kawamoto, S. (2022). Lemon myrtle (*Backhousia citriodora*) extract and its active compound, Casuarinin, activate skeletal muscle satellite cells *in vitro* and *in vivo*. *Nutrients, 14*(5), 1078.

Yao, X., Cai, F., Zhu, P., Fang, H., Li, J., & He, S. (2019). Non-invasive and rapid pH monitoring for meat quality assessment using a low-cost portable hyperspectral scanner. *Meat Science, 152*, 73–80.

Yin, W., Zhang, C., Zhu, H., Zhao, Y., & He, Y. (2017). Application of near-infrared hyperspectral imaging to discriminate different geographical origins of Chinese wolfberries. *PLoS One, 12*(7), Article e0180534.

Zhang, J., Yang, R., Li, Y. C., Wen, X., Peng, Y., & Ni, X. (2020). Use of mineral multi-elemntal analysis to authenticate geographical origin of different cultivars of tea in Guizhou, China. *Journal of the Science of Food and Agriculture, 100*(7), 3046–3055.

Zhao, S., Song, W., Hou, Z., & Wang, Z. (2021). Classification of ginseng accorsing to plant species, geographical origin, and age using laser-indused breakdown spectroscopy and hyperspectral imaging. *Journal of Analytical Atomic Spectrometry, 36*, 1704.