Advancing Mango Ripeness Assessment: A Comprehensive Study Integrating VNIR Spectroscopy and SIMCA Modelling for 'Dashehari' Cultivar

Abstract

This study addresses the challenges associated with assessing mango ripeness, particularly in the Dashehari cultivar, a popular mid-season mango in northern India. Farmers faced many problems during harvesting season. Traditional ripeness assessment methods are deemed inaccurate and time-consuming, necessitating the development of non-destructive techniques. The research focuses on the application of Visible and Near-Infrared (VNIR) spectroscopy, coupled with chemical models, to create a versatile tool for predicting Soluble Solids Content (SSC) in thin-skinned fruits with similar physicochemical characteristics. The investigation extends to the effectiveness of VNIR spectroscopy in combination with classification models for mango identification and ripening stage prediction. The chosen wavelength regions, guided by preprocessing techniques and Principal Component Analysis (PCA), demonstrate distinct clustering among unripe, half ripe, and fully ripe mangoes, particularly in the 670-850 nm range. The Soft Independent Modelling by Class Analogy (SIMCA) model, incorporating PCA, achieves remarkable classification accuracy rates of 100%, 96.66%, and 93.33% for unripe, halfripe, and fully ripe fruits, respectively, within the 670-850 nm wavelength region. In the context of the Dashehari mango, known for its green skin even when fully ripe, the study provides valuable insights into precise ripeness assessment. The proposed approach holds significance for the mango industry, aiding in quality assurance and post-harvest strategies for marketing, transportation, and storage. The combination of VNIR spectroscopy and SIMCA modelling emerges as a promising solution, offering advantages in terms of accuracy, efficiency, and reduced post-harvest losses.

Keywords- VNIR spectroscopy; SIMCA model; Ripening stage; PCA

1. Introduction

The mango (*Mangifera indica* L.) is a climacteric fruit that is grown in tropical and subtropical regions worldwide. Its cultivation is on the rise annually due to growing consumer demand (Marques et al., 2016).The leading mango-producing countries, including India, China, Thailand, Indonesia, Mexico, Pakistan, and Brazil, collectively contribute to approximately 72% of global mango production (FAOSTAT, 2022). Mangoes are typically harvested slightly before reaching full maturity to prevent the initiation of climacteric respiration during their transportation to distant markets (Jha et al., 2014)

While mango production holds global significance, challenges associated with fruit quality hinder widespread consumption of this fruit (Prasad et al., 2018). One major issue involves marketing fruits with varying maturity stages and consumer quality within the same batch. Dashehari, a mid-season mango, stands out as one of the most favored varieties in northern India because of its substantial yield potential and excellent fruit quality (Prasad et al., 2019). The ripening of fruits on the tree leads to the production of heterogeneous quality fruits, both in terms of size and sensory attributes (Gill et al., 2017). Therefore, mango fruits are harvested when they are mature, green, and unripe. A significant drawback of the Dashehari cultivar is that its skin remains green even after reaching full ripeness. To address this issue, a non-destructive technique is needed for detecting the ripeness of mangoes (Saroj et al., 2023).

The digitalized forecasting of maturity holds significant importance for both export purposes and in minimizing substantial post-harvest losses. Proficiency in computing and predicting maturity levels can aid in developing effective strategies for marketing, transportation, and storage (Palafox-Carlos et al., 2012). Traditional Dashehari mango ripeness assessment methods, like smell, cutting samples, hand pressing, or days post-harvest, are time-consuming, inaccurate, potentially damaging, and lack batch representation (Dhaneshwari et al., 2023). Consumer reliance on surface features can be misleading. Harvest maturity is vital for end-use, and choosing appropriately mature fruits is crucial for quality assurance. Mango processing industries require precise ripeness assessment for products like juice, pulp, jam, and jelly (Lawson et al., 2019).

A blend of NIRS (Near-Infrared Spectroscopy) and chemical models was employed to create a versatile tool. This tool was then applied to establish a comprehensive model capable of predicting Soluble Solids Content (SSC) in thin-skinned fruits sharing similar physicochemical characteristics. Near-infrared spectroscopy (NIRS) detection methods offer the benefits of being non-invasive, user-friendly, environmentally conscious, and safe (Shi & Yu 2017).Fruits and vegetables comprise organic compounds with various chemical bonds (C-H, N-H, S-H, O-H). NIR spectroscopy's spectral profile reflects information on hydrogen-containing groups, providing insights into the organic matter and composition of diverse biochemical structures (Shao et al. 2012). NIRS is increasingly accepted for its cost-effectiveness, ease of use, and online applicability. Visual and NIR spectroscopy accurately and rapidly predict maturity and sweetness in certain mango cultivars based on single properties like firmness and TSS. (Schmilovitch et al., 2000; Jha et al., 2014).

This study investigated the effectiveness of VNIR spectroscopy combined with classification models for identification purposes. The objectives included obtaining spectral data from mango fruits during ripening, preprocessing the data, selecting suitable wavelengths for model development, and ultimately constructing SIMCA models to predict mango ripening stages.

2. Material and methods

2.1 Mango fruit sample collection and preparation

200 Dashehari mangoes were harvested at 75-80% maturity based on color and specific gravity from the Indian Agricultural Research Institute, New Delhi, in season 2022 and 2023. Immediate pre-cooling during postharvest maintained quality. Selected mangoes, with flawless skins, were uniquely identified for non-destructive assessments every 5 days over 25 days. Concurrent destructive analysis confirmed ripening stage of mango.

2.2 Capturing Mango Spectral Signatures with a Spectroradiometer

Utilizing the Spec3 Analytical Spectral Device (ASD) spectroradiometer, whole mangoes' reflectance-mode spectral signatures were acquired. Covering a wavelength range of 350-2500 nm at 1nm intervals, the device underwent careful calibration with a standard white reference before capturing the spectral data. Precautions were taken to ensure no gaps between the mango fruit and the handheld probe, preventing signal losses. Spectral signatures were collected at 0, 5, 10, 15, 20, and 25 days post-harvest during Dashehari mango ripening. Each mango underwent scanning at 8 locations, and the data were averaged for analysis. For internal disorder confirmation, 10% of fruits per storage interval underwent destructive analysis by being cut open.

2.3 Confirmation of Ripening Stage through Destructive Sample Analysis

Following the acquisition of spectral signatures, mango samples were halved near the mango stone using a sharp knife, minimizing damage to the fruit pulp. Digital photographs were taken using a DSLR camera (Nikon, Japan) for subsequent analysis.

2.4 Selection of critical wavelength for VNIR spectroscopy-based defect detection

Selecting specific wavelengths involves narrowing down the range where non-destructively acquired spectral reflectance can effectively categorize mango fruits. Various preprocessing, data visualization, and data reduction techniques were applied to process the spectral reflectance. This facilitated the identification of a sensitive wavelength region for subsequent modeling.

2.5 Principal Component Analysis (PCA) of spectral data

Principal Component Analysis (PCA) can be applied to spectral data for various purposes, such as feature extraction, dimensionality reduction, and visualizing spectral data. Furthermore, PCA has the capacity to streamline intricate datasets, emphasize trends, and simplify model building by reducing data dimensionality. In this research, PCA was utilized to categorize data into three classes [unripe (Specific gravity- 0.96 to 0.98), half ripe (Sp. Gravity- 1.0 to 1.01), and fully ripe (Sp. Gravity- 1.03 to 1.04)] and identify essential wavelength regions for constructing classification models, utilizing PC score, loadings, and influence plots.

2.6 Development of SIMCA model

SIMCA creates distinct models for each ripening stage, accurately representing the unique variations and relationships among characteristics specific to each type of fruit. Before modeling, the spectral data was divided into training and testing sets in an 80:20 ratio. Utilizing the statistical properties of class-specific models, SIMCA establishes control limits, outlining the range within which data points for each class should lie.

3.Results and Discussion

3.1 Effectiveness of VNIR spectroscopy for detection of ripening stage

Reflectance spectra were obtained in ripening mango fruits using a Vis-NIR spectroradiometer in reflectance mode, as illustrated in Fig. 1. It was noted that the reflectance of mango fruits declined during storage or ripening. Fresh mango fruits exhibited a reflectance value exceeding 0.85, while for half ripe mangoes, it ranged between 0.85 and 0.75. Fully ripe mangoes displayed reflectance below 0.75 within the wavelength range of 700 to 1200 nm. The current research demonstrates the efficacy of Vis-NIR (Visible and Near-Infrared) spectroscopy in identifying spongy tissue occurrence in Alphonso mangoes. However, further investigations are required to comprehend and delineate the fruit quality parameters influencing Vis-NIR spectra within the 670 nm to 970 nm wavelength range and their correlation with internal disorders in the fruit. In a study by Buccheri et al. (2019) on Annurca apples, NIR spectroscopy was employed for quality monitoring, achieving a correct classification rate of 93.3% for fresh apples, which decreased to 70% after 7 days of storage. The experimental findings of the present study reveal that the optimal wavelength range for detecting ripening stages in mangoes is between 670 nm and 850 nm, surpassing other tested ranges, specifically 480 nm to 650 nm and 1100 nm to 1430 nm, in accurately identifying internal ripening in mango fruit.

Following the visualization and profiling of mango fruit spectral reflectance, various preprocessing techniques were employed to select reflectance peaks and wavelength regions conducive to improved modeling. Gap segment first derivative was applied to the spectral reflectance to identify significant reflectance peaks and choose regions of interest. It was observed that wavelength ranges of 670-850, 900-970, 1100-1170, and 1270-1430 nm exhibited noteworthy reflectance peaks, along with substantial differences in reflectance among ripening stages (Fig. 2). Other wavelength regions demonstrated minimal or negligible differences.

3.2Effectiveness of SIMCA Model in Binary Classification of Mango Based on Spectral Reflectance

3.2.1 Performance of principal component analysis (PCA) for classification of mangoes

Following the choice of wavelength regions guided by the first derivative, Principal Component Analysis (PCA) was employed for feature extraction, data reduction, classification, and visualization of spectral data. Notably, in the 670-850 nm wavelength range, distinct clustering was observed among unripe, half ripe, and fully ripe mangoes. Although some samples exhibited overlap at the boundary between two clusters, the first two components (PC1 and PC2) still accounted for up to 90% of the variance, as depicted in Fig. 3. In this study, principal component analysis was employed for selecting wavelength regions. Post-PCA, the wavelength regions of 670-850, 900-970, and 1100-1170 nm distinctly categorized mango spectral signatures into two clusters, whereas regions 480-670 and 1270-1430 nm provided partial classification of healthy and damaged mango. Raghavendra et al. (2021) investigated wavelength regions for detecting internal defects in mango using NIR spectroscopy, revealing that the 702 to 752 nm range yielded optimal results with an 84.5% classification accuracy. The study observed that reflectance within the 670 to 850 nm wavelength range effectively facilitated clustering mango fruits into healthy and damaged categories.

3.2.2 Performance of Soft Independent Modelling by Class Analogy (SIMCA) model on selected wavelength regions

The range of wavelengths from 670 to 850 nm provided credible identification results with a significance level of α= 5%, as illustrated in Fig. 4. In the Si vs Hi plot for unripe, halfripe, and fully ripe fruits, all ripening stages were precisely identified, positioned within the quadrilateral area with minimal leverage and distance from both axes.To check the classification performance of SIMCA model 96 samples were used for testing purposes.It was found that wavelength region of 670 to 850 nm showed highest classification accuracy rate of 100, 96.66 and 93.33% for unripe, half ripe and fully ripe fruits, respectively.

In this study, the main goal was to develop a model capable of differentiating between mangoes affected by spongy tissue and those that are healthy. To achieve this objective, researchers applied feature extraction and selection techniques to efficiently choose relevant wavelengths. In the current research, a classification model, specifically SIMCA combined with PCA, was utilized. Only selected regions within the entire wavelength range, such as 670-850 nm, were used for SIMCA modeling. The model's performance was assessed based on classification accuracy, yielding an impressive accuracy ranging from 93% to 100% for wavelengths between 670 nm to 850 nm when utilizing the original features with the SIMCA model incorporating PCA.

Sanaeifar et al. (2014) employed computational techniques and an electronic nose to monitor the ripeness of bananas, utilizing PCA, SVM, and SIMCA methods. PC1 and PC2 were found to contribute to 82% of the variance, with the SIMCA model achieving 92% accuracy and the SVM model recording 98.66% accuracy. In a study by Pu et al. (2019), bananito fruits were classified based on ripeness using visible spectroscopy and hyperspectral imaging, achieving a total correct classification rate (TCC) of 86.7% with the SIMCA model. Shi et al. (2022) utilized Neat Infrared Spectroscopy (NIRS) and a SIMCA model to classify rice varieties, obtaining accuracies of 100% for Kenjing No-5, 100% for No-6, and 97.5% for No-9.

Conclusion

This study presents a comprehensive approach for non-destructive ripeness assessment of Dashehari mangoes, addressing the challenges associated with traditional methods. The combination of Visible and Near-Infrared (VNIR) spectroscopy with chemical models, specifically the Soft Independent Modelling by Class Analogy (SIMCA) model, proves effective in predicting ripening stages with remarkable accuracy rates within the 670-850 nm wavelength range.

The research emphasizes the global significance of mango production and the need for precise ripeness assessment, particularly in the context of the Dashehari cultivar's unique challenges, where the green skin persists even after full ripeness. By employing VNIR spectroscopy and SIMCA modeling, the study achieves a reliable identification tool for mango ripeness, offering advantages such as non-invasiveness, user-friendliness, and environmental consciousness.

Looking ahead, the future thrust of this research lies in expanding the applicability of the developed model to a broader range of mango varieties and exploring its potential integration into mango processing industries. Additionally, ongoing efforts should focus on refining the model's robustness to varying environmental conditions and storage durations. Continuous research and development in this direction hold promise for advancing the efficiency of mango ripeness assessment, contributing to improved post-harvest management, reduced losses, and enhanced product quality in the mango industry.

References

Buccheri M, Grassi M, Lovati F, Petriccione M, Rega P, Scalzo RL, Cattaneo TM. Near infrared spectroscopy in the supply chain monitoring of Annurca apple. *J. Near Infrared Spectrosc. 2019;*27(1), 86-92.

Dhaneshwari DSPS, Deshmukh UB. Effect of ethrel on the ripening behaviour of mango (Mangifera indica L.) variety Dashehari. 2023.

Gill PPS, Jawandha SK, Kaur N, Singh N. Physico-chemical changes during progressive ripening of mango (Mangifera indica L.) cv. Dashehari under different temperature regimes. *J.Food Sci. Technol. 2017;*54, 1964-1970.

Jha SN, Narsaiah K, Jaiswal P, Bhardwaj R, Gupta M, Kumar R, Sharma R.Nondestructive prediction of maturity of mango using near infrared spectroscopy*. J. Food Eng. 2014;*124, 152-157.

Lawson T, Lycett GW, Ali A, Chin CF. Characterization of Southeast Asia mangoes (Mangifera indica L) according to their physicochemical attributes. *Sci.Hortic. 2019;*243, 189-196.

Marques EJN, de Freitas ST, Pimentel MF, Pasquini C. Rapid and non-destructive determination of quality parameters in the ‘Tommy Atkins’ mango using a novel handheld near infrared spectrometer. *Food Chem. 2016;*197, 1207-1214.

Palafox-Carlos H, Yahia E, Islas-Osuna MA, Gutierrez-Martinez P, Robles-Sánchez M, González-Aguilar GA. Effect of ripeness stage of mango fruit (Mangifera indica L., cv. Ataulfo) on physiological parameters and antioxidant activity. *Sci.Hortic. 2012;*135, 7-13.

Pu YY, Sun DW, Buccheri M, Grassi M, Cattaneo TM, Gowen A. Ripeness classification of Bananito Fruit (Musa acuminata, AA): A comparison study of visible spectroscopy and hyperspectral imaging. *Food Anal. Methods. 2019;*12, 1693-1704.

Raghavendra A, Guru DS, Rao MK. Mango internal defect detection based on optimal wavelength selection method using NIR spectroscopy. *Artif.Intell. Agric. 2021;*5, 43-51.

Sanaeifar A, Mohtasebi S, Ghasemi-Varnamkhasti M, Ahmadi H, Lozano Rogado JS. Development and application of a new low-cost electronic nose for the ripeness monitoring of banana using computational techniques (PCA, LDA, SIMCA, and SVM). 2014.

Schmilovitch ZE, Mizrach A, Hoffman A, Egozi H, Fuchs Y. Determination of mango physiological indices by near-infrared spectrometry. *Postharvest Biol. Technol. 2000;*19(3), 245-252.

Shao X, Ning Y, Liu F, Li J, Cai W. Application of near-infrared spectroscopy in micro inorganic analysis. *Acta ChimicaSinica. 2012;*70(20), 2109.

Shi G, Zhang X, Qu G, ChenZ. Classification of Rice Varieties Using SIMCA Applied to NIR Spectroscopic Data. *ACS omega. 2022;*7(50), 46623-46628.

Prasad K, Sharma R. R, Sethi S, Srivastav M. Influence of harvesting method on postharvest loss, shelf-life and quality of mango (*Mangifera indica*) fruits. Indian J Agric Sci, 2019;(3), 445-449.

Prasad K, Jacob S, Siddiqui M. W. Fruit maturity, harvesting, and quality standards. In Preharvest Modulation of Postharvest Fruit and Vegetable Quality (pp. 41-69). Academic Press, USA 2018.

Saroj N, Prasad K, Singh S. K, Kumar V, Maurya S, Maurya P, Kumar R. Characterization of bioactive and fruit quality compounds of promising mango genotypes grown in Himalayan plain region. PeerJ, 2023;11: e15867.

Shi H, Yu P, Comparison of grating-based near-infrared (NIR) and Fourier transform mid-infrared (ATR-FT/MIR) spectroscopy based on spectral preprocessing and wavelength selection for the determination of crude protein and moisture content in wheat. *Food Control. 2017;*82, 57-65.

Figures

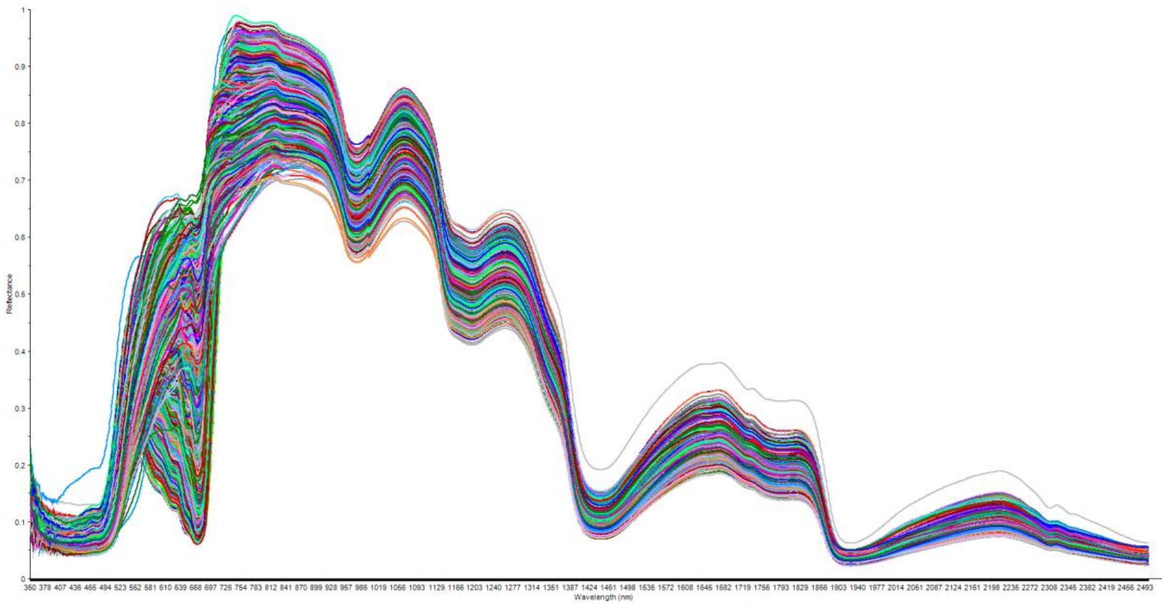


Fig. 1 Spectral reflectance of mango fruits in VNIR region

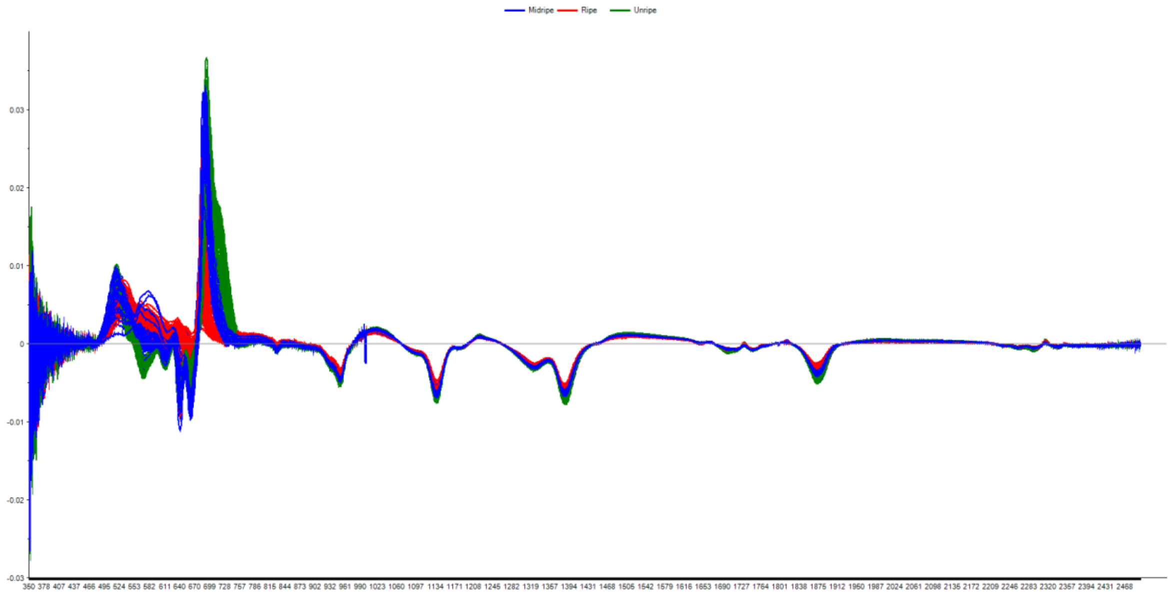


Fig. 2 First derivative of mango spectral signatures

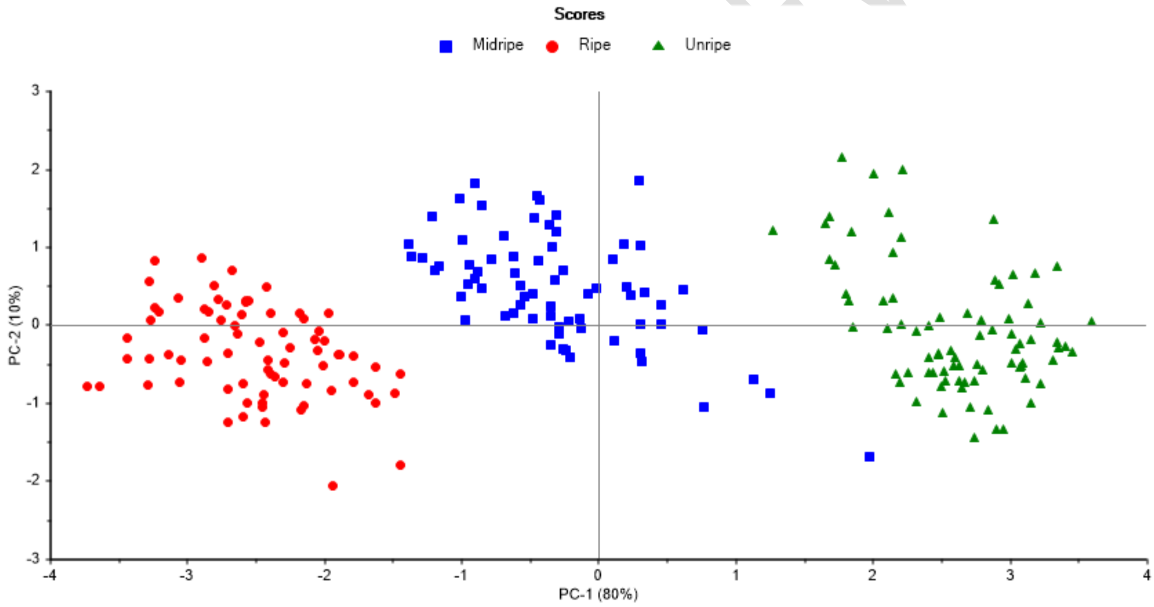
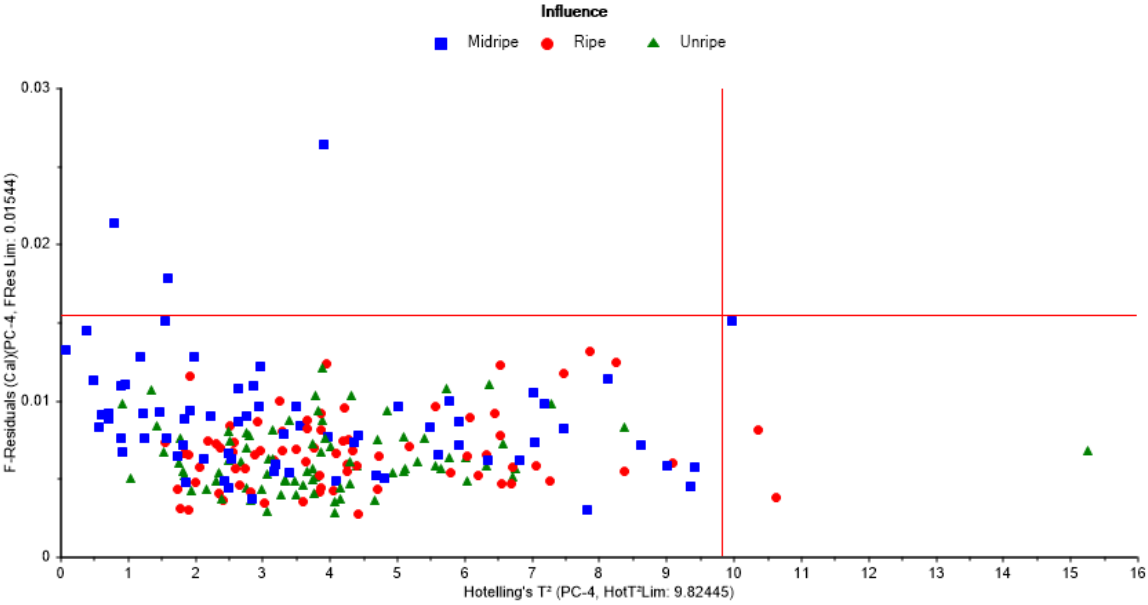
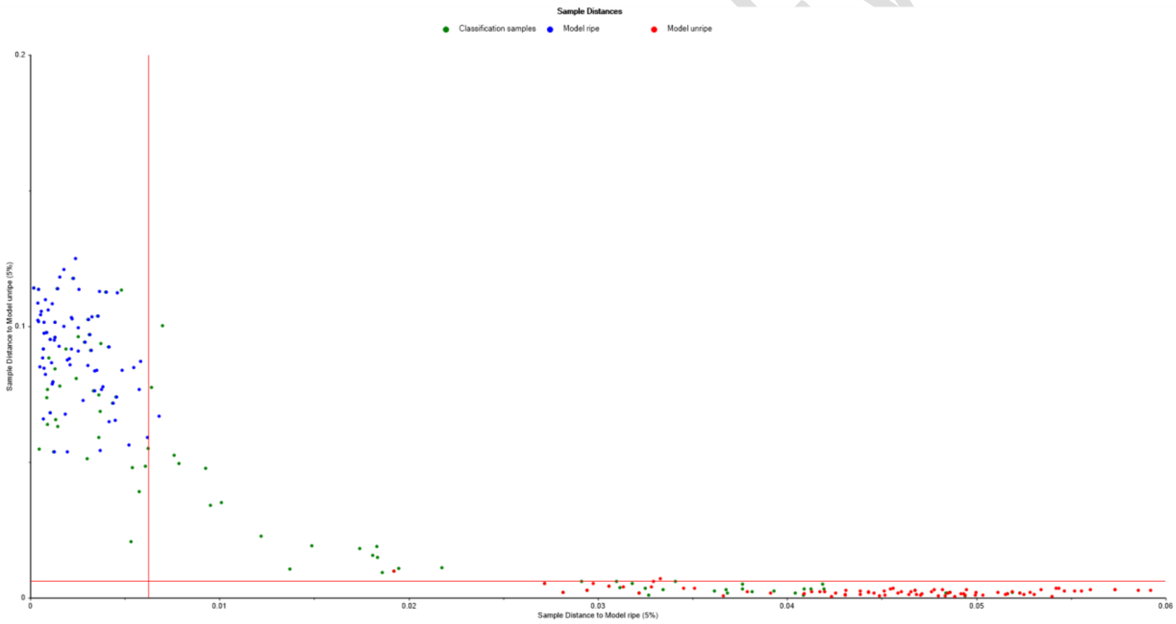


Fig. 3 PCA score of spectral data in the wavelength region of 670-850 nm (Green- unripe, Blue- half ripe, Red- full ripe)



(A)



(B)

Fig. 4 SIMCA classification of mango in wavelength region of 670-850 nm with A) Coomans plot and B) Si vs Hi plot for healthy fruit