DEVELOPMENT OF DECISION SUPPORT SYSTEM FOR NON-DESTRUCTIVE FRUIT QUALITY ESTIMATION

By

Ayesha Zeb

(Registration number: 00000115218)

Department of Electrical Engineering

Military College of Signals

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

(2022)

DEVELOPMENT OF DECISION SUPPORT SYSTEM FOR NON-DESTRUCTIVE FRUIT QUALITY ESTIMATION

By

Ayesha Zeb

(Registration number: 00000115218)

A thesis submitted to the National University of Sciences and Technology, Islamabad,

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in

Electrical Engineering

Thesis Supervisor: Brig. Dr. Abdul Ghafoor

Co-Supervisor: Dr Muhammad Imran

Department of Electrical Engineering

Military College of Signals

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

(2022)

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of PhD Thesis written by **Miss. Ayesha Zeb,** Registration No. **00000115218**, of **Military College of Signals** has been vetted by undersigned, found complete in all respects as per NUST Statutes / Regulations / PhD Policy, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of PhD degree. It is further certified that necessary amendments as pointed out by GEC members and foreign/local evaluators of the scholar have also been suitably incorporated in the said thesis.

Signature: _____

Prof Dr Abdul Ghafoor

Date: _____

Signature : _____

HoD

Date: _____

Signature: _____

Dean/Principal

Date: _____

Form PhD-7

DOCTORAL PROGRAM
OF STUDY

National University of Sciences & Technology
REPORT OF DOCTORAL THESIS DEFENCE

Name: **Ayesha Zeb** NUST Regn No **00000115218**

School/College/Centre: **Military College of Signals**

Title: **Development of Decision Support System for Non-Destructive Fruit Quality Estimation**

DOCTORAL DEFENCE COMMITTEE

Doctoral Defence held on **21th Dec 2022**

	QUALIFIED	NOT QUALIFIED	SIGNATURE
GEC Member 1: Assoc Prof Dr Alina Mirza	☐	☐	_____
GEC Member 2: Assoc Prof Dr Naima Iltaf	☐	☐	_____
GEC Member 3: Assoc Prof Dr Attiq Ahmed	☐	☐	_____
GEC Member (External): Dr Waqar Shahid	☐	☐	_____
Supervisor: Prof Dr Abdul Ghafoor, PhD	☐	☐	_____
Co-Supervisor: Assoc Prof Dr Muhammad Imran	☐	☐	_____
External Evaluator 1: Dr Muhammad Naveed Tahir (Local Expert)	☐	☐	_____
External Evaluator 2: Dr Fazli Subhan (Local Expert)	☐	☐	_____
External Evaluator 3: Dr Raheel Nawaz (Foreign Expert)	☐	☐	
External Evaluator 4: Dr Siti Zaiton Binti Mohd Hashim (Foreign Expert)	☐	☐	

FINAL RESULT OF THE DOCTORAL DEFENCE
(Appropriate box to be signed by HOD)

☐ PASS ☐ FAIL

The student **Ayesha Zeb** Regn No **00000115218** is accepted for Doctor of Philosophy Degree.

Dated: _____

_____
Dean/Commandant/Principal/DG

CERTIFICATE OF APPROVAL

This is to certify that the research work presented in thesis titled **"Development of Decision Support System for Non-Destructive Fruit Quality Estimation"** was conducted by **Ayesha Zeb**, under the supervision of **Prof Dr Abdul Ghafoor.** No part of this thesis has been submitted anywhere else for any other degree. This thesis is submitted to the Department of Electrical Engineering, Military College of Signals in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the field of Electrical (Telecommunication) Engineering, Department of **Electrical (Telecommunication) Engineering,** of **Military College of Signals**, **National University of Sciences and Technology, Islamabad**.

Student Name: **Ayesha Zeb** Signature: _____

Examination Committee:

a) External Examiner 1: **Dr Muhammad Naveed Tahir** Signature: _____
(Assistant Professor COMSATS University Islamabad)

b) External Examiner 2: **Dr Fazli Subhan** Signature: _____
(Assistant Professor Air University Islamabad)

c) Internal Examiner 1: **Assoc Prof Dr Ata ur Rehman** Signature: _____
(Electrical Engineering Department, MCS NUST)

Supervisor: **Prof Dr Abdul Ghafoor** Signature: _____

Dean: **Prof Dr Asif Masood** Signature: _____

AUTHOR'S DECLARATION

I, **Ayesha Zeb**, hereby state that my PhD thesis titled "**Development of Decision Support System for Non-Destructive Fruit Quality Estimation**" is my own work and has not been submitted previously any by me for taking any degree from National University of Sciences and Technology (NUST), Islamabad, or anywhere else in the country / world.

At any time if my statement is found to be incorrect even after my Graduation, the university has the right to withdraw my PhD degree.

Dated:

_____

Ayesha Zeb

PLAGIARISM UNDERTAKING

I solemnly declare that research work presented in the thesis titled "**Development of Decision Support System for Non-Destructive Fruit Quality Estimation**" is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and that of National University of Sciences and Technology (NUST), Islamabad, towards plagiarism. Therefore I, as an Author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred / cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of PhD degree, the University reserves the rights to withdraw / revoke my PhD degree and that HEC and the University has the right to publish my name on the HEC / University Website on which names of students are placed who submitted plagiarized thesis.

Date:

_____

Ayesha Zeb

DEDICATION

Dedicated to

My beloved mother, my brother Yasir Zeb and my husband

ACKNOWLEDGEMENTS

All praise to the Almighty Allah for his countless blessings. I thank my Allah for blessing me with the strength and intellect to complete this degree and providing ease for me throughout this journey.

I would like to express my deepest gratitude to my supervisors, Brig. Dr. Abdul Ghafoor, Dr Muhammad Imran and Dr Waqar Shahid Qureshi for their invaluable guidance, time and encouragement throughout my degree. Without their guidance and support, it would not have been possible for me to complete the research work. This research was supported by Pakistan Agriculture Research Council (PARC), Agriculture Linkage Program. Grant numbers: AE-007 and National Center of Robotics and Automation (NCRA), Robot Design and Development Lab (RDDL) Grant numbers: DF-1009-31. Special thanks to my external supervisor, Dr Waqar Shahid Qureshi, the Co-PI of RDDL, NCRA, and Dr Moshin Tiwana for providing me access to the equipment of their funded project by PARC for my research work. I am also grateful to Prof Dr Aman Ullah Malik from Institute of Horticulture Sciences, University of Agriculture Faisalabad (UAF), Pakistan, for his guidance and providing access to the lab of UAF for mango destructive testing experiments. And to Dr Muhammad Amin from Islamia University of Bahawalpur for facilitating us and providing access to the mango orchards in Multan and Bahawalpur for harvesting mango samples. I am also highly indebted to Prof Kerry Walsh from Central Queensland University Australia, for his precious guidance and time-to-time review of my research work. I am thankful to the management of GR Farms in Chakwal for permitting us to work in their orange orchard for dataset collection. Bundle of thanks to my GEC members, Dr Alina Mirza, Dr Attiq Ahmed, Dr Naima Iltaf; my teachers, administrative staff, colleagues and friends for their kind assistance throughout my degree.

Last but not the least, I am indebted to my family for their prayers and support throughout my academics. Specially my mother who motivated me to pursue studies till PhD degree. During the times of difficulty, she is my source of comfort and only one smile of her gives me enough strength to keep struggling. She provided so much ease to me during my life that without her I would not have been at the place where I am today. And my father who believed in me and raised me up like his sons. My brothers specially, Yasir bhai for everything he has done for me. Whatever I am today, I am because of him. He is my strength and my inspiration. He always guided me in every phase of my life including my studies and my career. My beloved sister-in-law and my nephews for their cute suggestions and questions regarding my research. My father-in-law, mother-in-law, Sohaib bhai and my husband, without their support it would not have been possible for me to complete my degree in time.

ABSTRACT

To reduce fruit loss and ensure quality, harvest timing and load information is critical to farm management (of labour and packing consumables). Early harvest brings poor eating quality fruit to the market, while late harvest decreases the available shelf life of fruit. These factors drive the need for quantitative tools for fruit maturity and quality testing. The assessment of harvest time is generally based on time (number of days from flowering) and physical features (size, shape and surface characteristics, firmness and pulp color). The assessment of these physical features is subjective and requires experience labour. The current quality inspection methods in Pakistan include weight-based segregation and packaging, therefore the quality of each fruit is not traceable. A few value chains have now set standards for fruit dry matter (DM) content at the time of harvest assessed non-destructively via near infrared spectroscopy (NIRS) (e.g. Australian Mango Industry Association). Pakistani supply chain also needs to adopt such a system that provides traceability and visibility to each sample within fruit packs.

The focus of this thesis is to investigate the short-wave NIRS (SWNIRS) for fruit quality inspection and present a decision support system for the Pakistani horticulture. First in this thesis, I have developed a decision support system for prediction of Pakistani fruit's quality index values using SWNIRS. The investigated fruits, i.e. export varieties of mango ('Sindhri', 'Samar Bahisht Chaunsa' and 'Sufaid Chaunsa'), export variety of mandarin ('Kinnow') and loquat, hold high commercial significance to Pakistan, but also provide examples of fruit with relatively thin and thick skin and with relatively thick and thin edible flesh. These differences in morphology can be expected to impact the non-invasive assessment of flesh characters using SWNIRS. Locally developed partial least squares regression (PLSR) models returned an coefficient of determination (R^2) of 0.90 and root mean square error (RMSE) of 0.95 ^{o}Brix for solids soluble content (SSC) and R^2 of 0.80 and RMSE of 1.17% for DM in the prediction of a mango test set, and an R^2 of 0.71 and RMSE 0.65 ^{o}Brix in the prediction of SSC in a Kinnow mandarin test set. For cultivar 'Tanaka' loquats, the locally developed PLSR model achieved an R^2 of 0.90 and RMSE of 0.95^{o}Brix in the prediction of a test set. The results confirm the suitability of NIRS for the non-invasive evaluation of thin-skinned fruit and highlight the need for a data model trained on spectra obtained from multiple varieties when indicated by quality control on prediction performance.

Most of the reported literature on non-destructive fruit quality estimation uses an indirect approach to classify fruit sample i.e. predict the quality index value using some machine learning regression algorithm and based on the predicted value judge the sample quality (which requires prior knowledge about standards). Second, in this thesis, I have proposed a direct sweetness classifier for fruit sweetness classification as opposed to thresholding based indirect measure of quality index value. I have defined acceptance criteria for melons and oranges based on direct classification method to predict the sweetness level using NIR spectroscopy. I have compared performance of our classification-based approach with that of regression-based thresholding methods reported in literature. The proposed classifier has been tested for sweetness classification of Pakistani melon (variety: 'Honey melons') and orange (variety: 'Blood red', 'Mosambi' and 'Succari') fruits. The best SSC model was obtained using multiple linear regression on second derivative of spectral data (for wavelength range 729–975 nm) with correlation coefficient (R) = 0.93, and RMSE = 1.63 on test samples. Sweetness of test samples were obtained using ◦Brix thresholding with an accuracy of 55.45% for three classes. The best

direct sweetness classifier was obtained using K nearest neighbor (KNN) on second derivative of spectral data (for wavelength range 729–975 nm) with an accuracy of 70.3% for three classes on test samples. For oranges, PLSR models were developed for Brix, titratable acidity (TA), Brix: TA, and BrimA (Brix minus acids) estimation with a correlation coefficient of 0.57, 0.73, 0.66, and 0.55, respectively, on independent test data. The ensemble classifier achieved 81.03% accuracy for three classes (sweet, mixed, and acidic) classification on independent test data for direct fruit classification. Extensive evaluation validates our argument that modelling a direct sweetness classifier is a better approach as compared to estimation of quality indices for sweetness classification using NIR spectroscopy.

Automated fruit classification is a significant task in many industrial applications. For instance, it may help a supermarket cashier in identifying the fruit, its cultivar and subsequently its price. Computer vision based automatic fruit classification is relatively a mature field now, but it requires complex computer vision algorithms and systems to accurately classify different fruits. Third, I present SWNIR spectral data-based classifier for fruit classification problems. The research focuses on O-H and C-H overtone features of fruit and its correlation with SWNIRS and therefore opens a new dimension of fruit classification problems using SWNIRS. Eleven fruits, which include apple, cherry, hass, kiwi, grapes, mango, melon, orange, loquat, plum, and apricot, were used in this study to cover physical characteristics such as peel thinness, pulp, seed thickness, and size. Different shallow machine learning architectures were trained to classify fruits using spectral feature vectors. At first, using 83 features vectors within the range of 725-975nm (3nm-resolution) and then using only four features of wavelength 770nm, 840nm, 910nm, and 960nm (corresponding to O-H and C-H overtone features). For the 83 spectral features range as an input, the QDA classifier achieved a cross-validation accuracy of 100% and a test data accuracy of 93.02%. For the four features vector as an input, the QDA classifier achieved a cross-validation accuracy of 97.1% and test data accuracy of 90.38%. The results demonstrate that fruit classification is mainly a function of absorptivity of SWNIR radiation primarily with respect to O-H and C-H overtones features. An LED-based device mainly having 770nm, 840nm, 910nm, and 960nm range LEDs can be used in applications where automation in fruit classification is required.

The decision support system presented in this dissertation will not only aid the Pakistani supply chain for automated, efficient and non-destructive quality assessment of fruits in particular, but in general also opens a new dimension for fruit segregation in two applications: 1) using direct classification based on acceptance criteria of sweetness instead of quantitative assessment of quality attributes and 2) fruit type classification using SWNIR spectral features.

Table of Contents

List of Figures

List of Tables

ACRONYMS

NIRS	Near Infrared Spectroscopy
SWNIRS	Short Wave Near Infrared Spectroscopy
SSC/TSS	Solids Soluble Content / Total Soluble Solids (^{o}Brix)
DM	Dry Matter
TA	Titratable Acidity
BrimA	Brix minus acids
R_P	Correlation Coefficient of Prediction
R_{CV}	Correlation Coefficient of Cross Validation
R^2_P	Coefficient of Determination of Prediction
R^2_{CV}	Coefficient of Determination of Cross Validation
$RMSE_P$	Root Mean Square Error Prediction
$RMSE_{CV}$	Root Mean Square Error Cross Validation
MSC	Multiplicative Scatter Correction
SNV	Standard Normal Variate
SG 1-D	Savitzky Golay 1st derivative filter
SG 2-D	Savitzky Golay 2nd derivative filter
PCA	Principle Component Analysis
MLR	Multiple Linear Regression
PCR	Principle Component Regression
PLSR	Partial Least Squares Regression
ANN	Artificial Neural Networks
LDA	Linear Discriminant Analysis
SVM	Support Vector Machine
KNN	K Nearest Neigbor
MPLSR	Modified Partial Least Squares Regression
LSSVM	Least Square Support Vector Machine
CARS	Competitive Adaptive Reweighted Sampling
UVE	Uninformative Variable Elimination
CARS-SPA	Competitive Adaptive Reweighted Sampling-Successive Projections Algorithm
UVE-SPA	Uninformative Variable Elimination- Successive Projections Algorithm
BPNN	Back Propagation Neural Network
O-H bond	Oxygen-Hydrogen bond
C-H bond	Carbon-Hydrogen bond

Chapter 1 : INTRODUCTION

This chapter summarizes the motivation, research problems, scope, objectives and limitations of the presented research.

1.1. Motivation:

Each year 1.3 billion tonnes of the world's food supply worth around $1 trillion is squandered due to inefficient harvesting and transportation procedure. Concurring to UNESCO sustainable consumption and production goal, accomplishing more with less is the key to a sustainable future. Reducing environmental degradation, enhancing resource efficiency, and supporting more sustainable lives also accounts for this objective.

Despite being an Agricultural country, Pakistan's agricultural contribution to GDP declined to 24%, from 51% in 1957 [1]. According to Agriculture marketing information service, around twenty-nine types of fruits are produced throughout the year in Pakistan. However, mostly the produced fruits are expended in domestic markets. Pakistan exported 768 thousand tons of fruits FY 2019-20. Production of citrus, mango, apple, melons, dates, guava, and apricot reflects strong domestic market demand for horticulture crops. The highest production has reached i.e. Citrus fruit & Mango 2.4 and 1.7 million tons, the largest fruit crop group by volume and are major export revenue earner. Pakistan is the 4th largest producer of mango in the world. In the year 2019-20 125000 tons of mango were exported to different regions of the world, earning $72 million for Pakistan. **Figure *1.1*** shows Pakistan's mango export statistics since 2012[1]. From 2014-17, a decrease in mango export can be seen due to deteriorated quality of exported fruit (the EU imposed a ban on Pakistani mango import). Hence, new regulations were made to ensure quality.

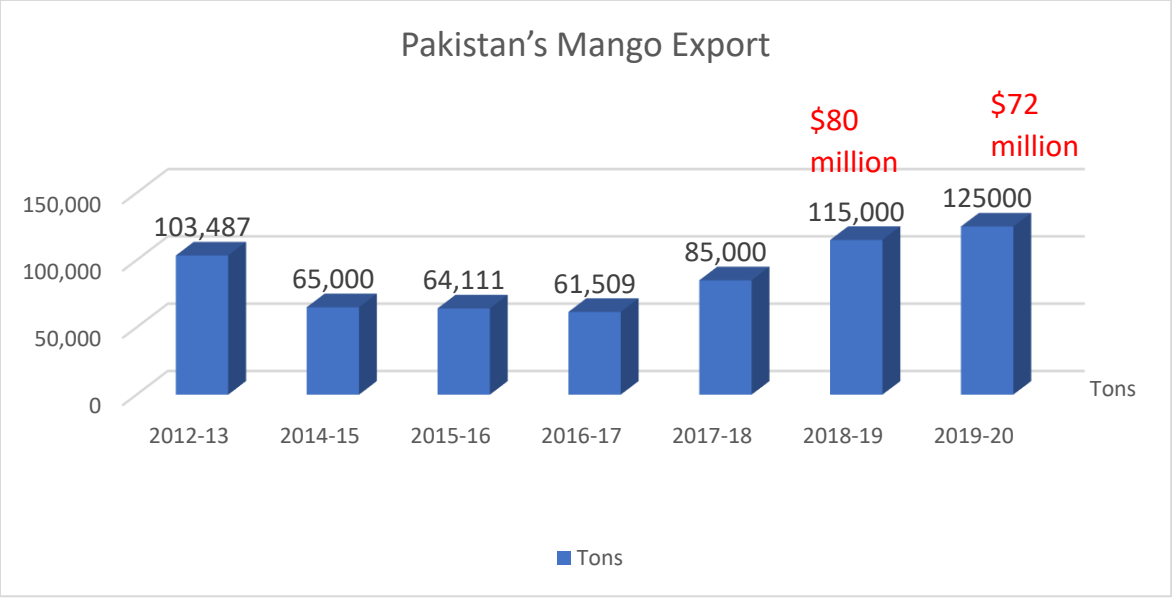


Figure 1.1: Pakistan's mango export statistics since 2012

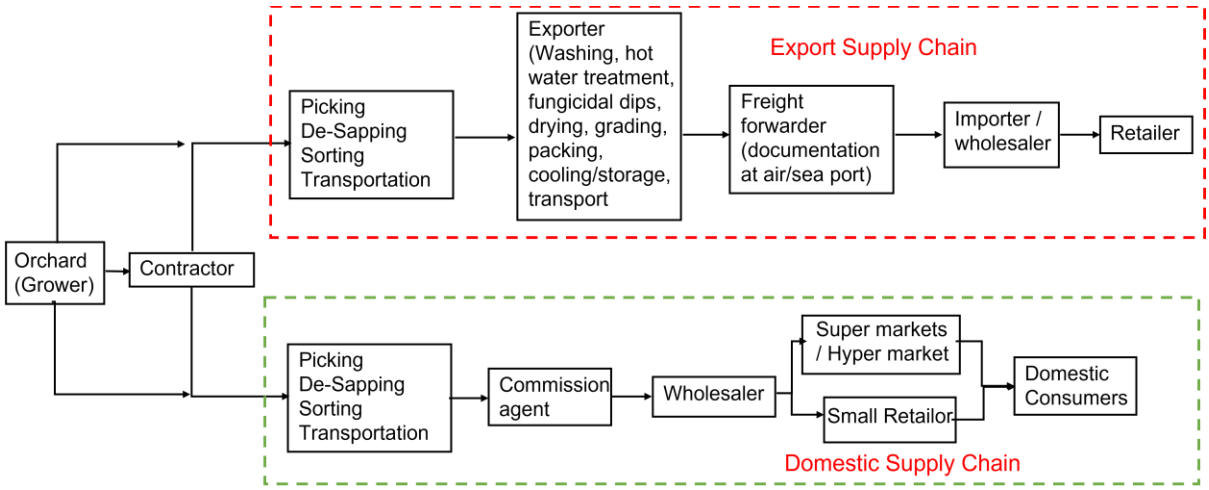


Figure 1.2: Pakistan's mango supply chains

Like other fruits and vegetables, Pakistan's mango supply chain involves multiple agents and procedures (see **Figure *1.2***). Fruit harvest maturity/quality is of great concern for growers and dealers. Early harvest of fruit results in delivery of poor eating quality fruit to market with consequent loss of purchases, while late harvest of fruit results in post-harvest loss by delivery of over ripe fruit to markets. The assessment of harvest time is generally based on time (number of days from flowering) and physical features (size, shape and surface characteristics, firmness and pulp color). Chemical and physical methods determine quality indices such as solids soluble

contents (SSC), dry matter (DM) and titratable acidity (TA). However, these methods are destructive and time-consuming procedures [2]. Currently, in Pakistan, the quality inspection methods are either manual, destructive or are weight-based. Thus, a need arises for non-destructive, fast, accurate instrumental techniques capable of screening the samples on-line/at-line.

Non-destructive testing using portable near infrared spectroscopy (NIRS) has been used in quantitative measure of fruit quality indices, which are correlated with fruit maturity [2]. In recent years, researchers have used NIRS over different wavelength regions with machine learning regression algorithms to develop maturity indexes prediction models of various fruits including apple [3],[4], pear [5],[6], mango [7], banana [8], [9], melon [10], [11], mandarin [12], [13], strawberry [14],[15], apricot [16], [17], kiwifruit [18],[19], persimmon [20], grape [21],[22], loquat [23] and pineapple [24]. However, due to the diversity in varieties and growing conditions, it is essential to develop the maturity index prediction model for a particular variety, growing region and for local or export varieties [2]. The Vis and NIR region of the light spectrum has the range 400-750 nm and 750-2500 nm, respectively. The short wave NIR (SWNIR) or Herschel region lies between 750-1100 nm and the extended NIR region lies between 1100-2500 nm. The SWNIR region is used commercially for the assessment of internal quality attributes of intact fruit, in preference to the extended NIR region [25]. Longer wavelength ranges offer narrower and stronger absorption features as compared to SWNIR and thus better evaluation of internal parameters however, the SWNIR wavelengths have greater effective penetration depth into the fruit, hence, offer robustness across independent populations and given the variation in outer layer attributes. The short-wave Vis-NIR option is preferred for commercial purposes due to (currently) lower hardware costs [25], [26].

The focus of this thesis is to investigate the SWNIR spectroscopy for fruit quality inspection and present a decision support system for the Pakistani horticulture. Recent reviews [25], [26] of the use of NIRS in fruit quality evaluation have noted the need for further work in the development of instrumentation, chemometric procedures including a need to demonstrate model robustness in prediction of fruit populations from a range of varieties and growing conditions, and for further application development in terms of application of NIRS to assessment of new commodities and attributes.

First, in this thesis, I developed a decision support system for the prediction of Pakistani fruit's quality index values using SWNIRS. The current quality inspection methods in Pakistan

include weight-based segregation and packaging, therefore the quality of each fruit is not traceable. A few value chains have now set standards for fruit DM content at the time of harvest assessed non-destructively via NIRS (e.g. Australian Mango Industry Association, [27]). The Pakistani supply chain also needs to adopt such a system that provides traceability and visibility to each sample within fruit packs.

Second, in this thesis, I have proposed a direct sweetness classifier for fruit sweetness classification as opposed to thresholding based indirect measure of quality index value. The proposed classifier has been tested for sweetness classification of Pakistani melons and orange cultivars. Most of the reported literature has reported chemometric models based on machine learning regression algorithms to predict fruit quality index value. Decision is taken based on the predicted value of quality index. For fruit sorting, a qualitative assessment (e.g. classification technique) can be used rather than a quantitative measurement of the amount of an attribute. Many claims exist in literature about fruit population classification, such as based on cultivars or species, geographic locations, or production systems [25]. Without explicit a priori information of the NIR evaluated component, NIRS can be utilised to determine an attribute. Jacobs et al. [28], for example, used NIR spectra to determine the storage life of lettuce.

Third, I present SWNIR spectral data-based classifier for fruit classification problems. The research focuses on O-H (Oxygen - Hydrogen) and C-H (Carbon - Hydrogen) overtone features of fruit and its correlation with SWNIRS and therefore opens a new dimension of fruit classification problems using SWNIRS. Fruit classification is a complex and challenging problem due to interclass similarities and uneven intraclass attributes. Choice of data acquisition sensors and relevant features is critical. Knowledge in the domain of computer vision has been widely used to solve the fruit classification problem using many sensors from black and white cameras to non-visual sensors like acoustic and tactile sensors [29]. For non-destructive classification, the acoustic and tactile sensors have limitations such as physical contact or fruit excitation requirements [30]–[36]. Vision based sensors also have limitations like high sensitivity to light conditions and background environment which introduce problems like reflection, refraction, translation, rotation and scale dependence.

To summarize, I have chosen to investigate the potential of SWNIR spectroscopy for non-destructive fruit sorting and quality inspection using direct and indirect approach to classify fruits.

1.2. Research Problems:

Section 1.1 summarised the research's objective and narrowed down its focus to a specific set of applications. In this section, I describe the research questions I set out to answer.

1. To provide traceability and visibility to each sample within a fruit pack, is it possible to measure non-destructively fruit quality estimates (such as, DM and/or SSC) during packaging of export fruit varieties in Pakistan?
2. Is it possible to have a decision support system that segregates fruit samples based on quality attributes using SWNIR spectral absorbance of fruits through direct classification instead of estimating fruit quality attributes from SWNIR spectral absorbance (such as, DM, SSC or etc. through regression) using machine learning models?
3. Is it possible to classify different types of fruit using a SWNIRS or a multiwavelength LED based simple and cost-effective device?

1.3. Scope, Objectives and limitations:

In this section, the scope, objectives and limitations of my research are described.

1. To introduce quality traceable / quality-based segregation system in Pakistani supply chain, I developed a decision support system for prediction of Pakistani fruit's quality index values non-destructively. The investigated fruits hold high commercial significance to Pakistan, but also provide examples of fruit with relatively thin and thick skin and with relatively thick and thin edible flesh. These differences in morphology can be expected to impact the non-invasive assessment of flesh characters using NIRS. The objectives are:
   (a) To develop NIRS based models for on-tree prediction of SSC and DM for three important export varieties of mango i.e. 'Sindhri', 'Sufaid Chaunsa' and 'Samar Bahisht Chaunsa'.
   (b) To develop SSC prediction model for our export variety of mandarin i.e. 'Kinnow'.
   (c) To assess the applicability of field portable NIRS to a novel commodity, Loquat (investigated variety, 'Tanaka').

The developed mango, mandarin and loquat models were trained with dataset collected at a single temperature (25^{o}C). To obtain a robust model against sample temperature variations (for on-tree predictions), the model needs to be trained with samples scanned at multiple temperatures.

2. To present a decision support system that can directly classify fruit samples as acceptable class sample or not, I choose melon and orange fruit as a case study. Melon and orange are high-value fruit crop grown on a large scale in Pakistan. They are not only cherished in Pakistan but the citrus fruit brings handsome revenue to the country by its export to many countries. Another reason for choosing these fruits is their thick inedible peel (average rind thickness of melons is 6.68mm and oranges is 5mm) which makes the penetration of NIR radiation challenging. Investigated local cultivars include one cultivar of melon i.e. 'Honey' melons and three cultivars of oranges i.e. 'Mosambi', 'Succari' and 'Blood red'. The objectives are as follows:
   (a) To define acceptance criteria for melons based on direct classification method to predict sweetness level using NIR spectroscopy.
   (b) To define acceptance criteria for oranges based on direct classification method to predict taste level (sweet/mix/acidic) using NIR spectroscopy.
   (c) Suitable design of the decision support system elements and selection of suitable pre-processing and classification techniques.
   (d) Compare proposed direct classification method with literature reported indirect method of quality index value prediction.

   Considering orange fruit, along with sweetness, the consumer acceptance is usually also linked with richness and tartness. As in the case of succari cultivar which is always sweet but not as tasty as mosambi and blood red cultivars. Our proposed sweetness classifier for orange is based on acceptance criteria linked only with sweetness of sample and not the tartness and richness parameters.

3. To investigate the potential of SWNIRS for fruit classification problem, I collected datasets using the F-750 for eleven fruits including apple, cherry, hass, kiwi, grapes,

mango, melon, orange, loquat, plum, and apricot to cover physical characteristics such as peel thinness, pulp, seed thickness, and size. The objectives are:

(a) To test different shallow machine learning architectures for fruit type classification
(b) Compare two sets of spectral feature vectors i.e. 83 features vectors within the range of 725-975nm (3nm-resolution) and only four features of wavelength 770nm, 840nm, 910nm, and 960nm (corresponding to O-H and C-H overtone features).

1.4. Contributions:

The main contributions of this dissertation are as follows:

1. I present a decision support system for non-destructive quality-based fruit sorting and grading for Pakistani horticulture.
2. I present a novel decision support system to predict sweetness of melons and oranges which is based on direct classification approach.
3. I present SWNIR spectroscopy-based solution to fruit classification problem.

1.5. Publications:

This section lists the papers I have published and submitted for review as part of this dissertation.

Published papers:

1. Zeb, A., Qureshi, W. S., Ghafoor, A., Malik, A., Imran, M., Iqbal, J., & Alanazi, E. (2021). Is this melon sweet? A quantitative classification for near-infrared spectroscopy. Infrared Physics & Technology, 114, 103645. (IF: 2.638) [37]
2. Zeb, A., Qureshi, W. S., Ghafoor, A., & O'Sullivan, D. (2022, February). Learning fruit class from short wave near infrared spectral features, an AI approach towards determining fruit type. In 2022 8th International Conference on Mechatronics and Robotics Engineering (ICMRE) (pp. 193-196). IEEE. [38]

3. Zeb, A., Walsh, K.B., Qureshi, W.S., Malik, A., Shah, S.S.A., Ghafoor, A., Alanzai, E. Vis-NIR spectroscopy for fruit quality evaluation: case study for Pakistani apple, mango, mandarin and loquat production. Accepted in International Food research journal. (IF: 1.1014)

Submitted papers:

1. Zeb, A., Qureshi, W. S., Ghafoor, A., Malik, A., Imran, M., Alina, M., Tiwana, M., & Alanazi, E. Sensory assessment classification using short-wave NIR spectroscopy for Pakistani orange cultivars.
   Revised version submitted to Scientific Reports (IF: 4.99)

1.6. Dissertation outline:

This thesis is organized as follows:

Chapter 2: This chapter presents the basic model of the proposed decision support system and introduction/literature review of all the basic blocks, i.e. NIRS and chemometrics.

Chapter 3: This chapter presents decision support system for prediction of Pakistani fruit's quality index values. The investigated fruits include apple, mango, mandarin and loquat. It includes relevant literature, materials and methods, results and conclusion sections.

Chapter 4: This chapter presents the proposed direct classification based decision support system for prediction of acceptance (sweetness) level of melons. It includes introduction, materials and methods, results, and conclusion sections. Includes material from [37]

Chapter 5: This chapter presents the proposed direct classification based decision support system for prediction of sweetness level of oranges. It includes introduction, materials and methods, results, and conclusion sections.

Chapter 6: This chapter presents the SWNIRS based proposed classification method for fruits type classification. It includes introduction, materials and methods, results, and conclusion sections. Includes material from [38].

Chapter 7: This chapter briefly summarizes the dissertation and concludes the research findings along with future recommendations.

Chapter 2 : DECISION SUPPORT SYSTEM

In this thesis, a decision support system for fruit quality estimation using NIR spectroscopy has been presented. Figure 1 shows the basic block diagram of the decision support system. To judge fruit quality, the sample is exposed to NIR radiation and as a result the molecular bonds of macro constituents absorb the radiation and vibrate. A spectrometer is used to measure this absorption of light with respect to individual wavelengths. Chemometric analysis is then performed on the raw absorbance spectra to tease out useful information about fruit internal quality for instance DM and Brix level etc. The decision support system is based on the following hypothesis:

"Fruit quality is a function of absorptivity of short wave NIR radiation primarily with respect to O-H and C-H overtones features"

These blocks in **Figure *2.1*** are explained in detail in sections below.

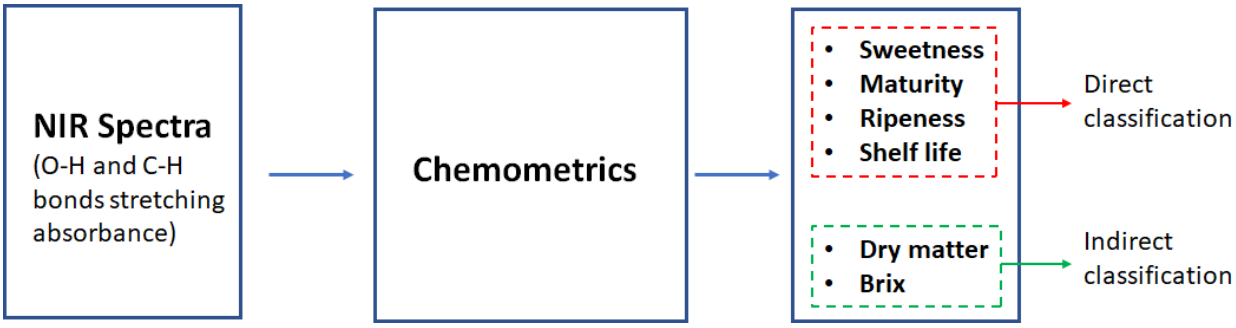


Figure 2.1: Decision support system for fruit quality estimation

1.1. Vis- NIR Spectroscopy:

Spectroscopy is the study of light and other radiation absorption and emission by materials, as well as the relationship between these processes and the wavelength of radiation. Spectroscopic techniques are used in nearly every modern discipline of science and technology. The visible and near-infrared (Vis and NIR) regions of the light spectrum have wavelength ranges of 400-750 nm and 750-2500 nm, respectively. Whereas, the short-wave NIR (SWNIR), or Herschel area, ranges from 750 to 1100 nm, while the extended NIR ranges from 1100 to 2500 nm.

Visible spectroscopy depends on the absorption of light of a matching energy level, i.e. at specified wavelengths, as a result of electronic transitions in molecules. As a result, with an average full width half maximum (FWHM) of about 20 nm, strong absorption peaks occur [39].

Light absorption associated to the vibration and stretching of molecular bonds is measured in NIR spectroscopy. This mainly comprises absorption of light associated to the C-H and O-H bonds stretching in intact fruits [40], [41], which is mostly related to water and storage reserves (the major macro constituents). The basic absorption ranges linked with these characteristics are found in the infrared region of light (>2500 nm), depicting far finer and greater absorption peaks than those found in the NIR region.

In the 1960s, Karl Norris of the United States Department of Agriculture (USDA) came up with the idea of employing NIR spectroscopy to quantify analyte levels in solid materials. The early research was based on agricultural goods with a low moisture content [42]. Later, the study was expanded to include intact fruits and vegetables with significant moisture content. Birth et al. [43] measured the amount of DM in onions, while Dull et al. [44] measured the amount of total soluble solids (TSS) in melons. Then, in Japan, a research and development programme began with the publication of the first study on non-destructive TSS assessment in peaches [45].

NIR region gives lower absorptivity of overtones features than those of infrared (IR) region. Due to this characteristic, the effective pathlengths of NIR radiation within fruit are in the millimetres to centimetres range. Whereas, in case of IR radiation the effective pathlengths are in the order of micrometres. Similarly, in the SWNIR region higher overtone characteristics with longer effective pathlengths are observed as compared to the overtones in NIR region. This is the reason that SWNIR radiation can be used to estimate fruit quality of intact fruits.

Absorption at 740 nm and 960 nm is associated with third and second overtone of O-H stretching and 840nm is associated with O-H combination feature. While, absorption at 910 nm is associated with third overtone of C-H stretching. However, the overall peak positions shift with temperature and solute concentration because the amount of H-bonding can change which influences the vibration of O-H bonds. In practice, its hard to interpret short-wave NIR (being second and third overtones, so weak and broad) compared to extended NIR and IR regions. However, the features related to water can be interpreted, which is the main NIR active molecule in fruit. Around 80-90% of fleshy fruit is composed of water. This is why every other variable is

examined in relation to water's enormous absorption characteristics. Increase in any other macroconstituent, e.g. solids soluble content (SSC) and dry matter (DM), causes decrease in water content resulting in a negative correlation with water. The penetration depth of NIR in fruits is comparatively greater in the 700–900 nm range [46] hence more information on internal quality attributes can be fetched by using wavelength absorption data of this region. Understanding the raw absorption spectra is challenging because all absorption characteristics are wide and overlapping. The use of NIRS over IR spectroscopy is successful due to chemometrics, which enabled useful information to be fetched out of the spectra.

Because H bonding affects the vibration of O-H bonds, water absorption characteristics are sensitive to temperature. Water exists in various states linked with the extent of H-bonding, from complete H bonding ('crystalline') in ice to none in water vapor. Increase in temperature or soluble sugars concentration causes decreased H-bonding and the absorption peaks linked with O-H bonds shift to shorter wavelengths [41]. The effect of temperature on DM prediction is mainly on accuracy (i.e., bias) instead of precision (i.e., bias corrected $RMSE_P$) (e.g., [47]).

2.1.1. Vis- NIR spectroscopy assessable attributes:

2.1.1.1. Dry matter and solids soluble content:

Vis-NIR spectroscopy is being used in commercial practice to estimate DM and SSC of intact thin-skinned fruits. According to a review [25], Vis-NIRS can be employed to measure DM and SSC of intact thin-skinned fruits to root mean square error prediction ($RMSE_P$) of less than 1% for both.

DM content is utilized as an indicator of:

- Storage reserves i.e. sum of starch and soluble sugars for instance in un ripped mangoes and apples, or oil content for instance in avocado or olive fruit
- Future soluble sugars content in climacteric fruits that store starch which is converted to sugars with ripening

Consequently, DM content at the time of harvest is an index of SSC of fruit after ripening.

A few value chains have now set standards for fruit DM content at the time of harvest assessed non destructively via NIRS (e.g. Australian Mango Industry Association, [27]). DM is also an important indicator of fruit harvest maturity, with the desired DM level altered by growing condition for example, in mango [48], [49] and in durian [50].

DM prediction models using NIRS have been successfully built that are robust to ripening phase i.e. to quantity of starch and sugars. But, the accuracy of SSC prediction models is affected with variations in starch levels e.g. in apple and mango ripening phase [51], [52]. The reason is probably that NIR spectra for sugar and starch in water are similar. Hence, for climacteric fruits, the accuracy of SSC prediction models with NIRS is highest for fully ripped fruits i.e. when starch is completely transformed to sugar (e.g. [52]). This conclusion does not apply to fruits that contain soluble sugars rather than starch as their primary reserve, such as stone fruit or grapes.

2.1.1.2. Internal defects:

After DM and SSC levels, the highest commercial adoption of Vis-NIRS is on packaging lines for fruit sorting based on internal defects The apparent absorption spectra of a fruit is influenced by both dispersion and absorption. Internal fruit features that cause variations in dispersion of light inside the fruit can be detected with Vis-NIRS just as easily as absorbance-related features. A decrease in the intensity of light scattering in a fruit is linked to an internal water core defect (e.g., [53]), and identification of this disease using Vis-NIRS has been published by many authors such as apple mouldy core [54] and apple internal browning [55].

2.1.1.3. Acidity:

In fruits like lemons and limes, which contain roughly 70 g L^{-1} (citric acid equivalents), acidity is a major constituent. However it is a micro-constituent in most other fruit, for example Subedi et al. [56] found that the average acidity of peach fruit was 8.8 g L^{-1} (citric acid equivalents). SSC measured of lime juice via refractometer is led by the influence of organic acids instead of soluble sugars, whereas the opposite is valid for peach. Subedi et al. [56], [57] used SWNIRS to

determine the $RMSE_P$ of an unbiased group of uncut and cut lime fruit to be 3.0 and 1.6 g L^{-1} (citric acid equivalents), respectively. Provided such estimate of $RMSE_P$, the consistent direct evaluation of acidity in fruits having low acidity amounts is doubtful (reliable means that test set is independent of training set). An indirect assessment includes correlation of one characteristic (i.e. acidity level) to another characteristic quantifiable by Vis-NIR spectroscopy (such as chlorophyll level). And if the indirect link holds true in practice, indirect acidity evaluation may be useful.

2.1.1.4. Firmness:

In postharvest research, a non-destructive measurement of fruit firmness remains a "holy grail." The usual technique of evaluation involves inserting a penetrometer into the fruit, which is damaging. Many claims have been made for employing Vis-NIRS to determine fruit firmness. Indeed, firmness assessment was mentioned in 10% of all studies published between 2015 and 2020 [25].

Slight variations in chemical (such as pectin levels) are linked to changes in firmness. The use of NIRS to estimate these chemical variations in uncut fruit seems unlikely. Changes in cell wall adherence are linked to changes in firmness. Changes in cell shape may cause changes in light dispersion inside the fruit, which can be observed in the perceived absorbance spectrum, although these variations in dispersion have not been associated significantly with firmness amount. Water content, pigment level, and starch-sugar conversion during ripening are all correlated to firmness levels, thus indirect Vis-NIRS assessments may be established (e.g., [58]).

As a result, there is no agreement that Vis-NIRS can reliably (and directly) detect firmness.

2.1.1.5. Other macro attributes:

Literature shows that with proper hardware, i.e. optical geometry, physiological states like maturity or ripening stage, as well as the extent of internal flaws like internal browning, can be evaluated [25].

2.1.1.6. Minor constituents:

It's also improbable that minor constituents could be directly and non-destructively assessed using Vis-NIRS in high moisture content fruit, except via indirect association with the amount of some other macro-constituent or pigment. For instance, [59] performed experiments on bell peppers using Vis-NIRS to estimate ascorbic acid non-destructively and obtained an $RMSE_{CV}$ of 15.1–18.9 mg per 100 g (fresh weight). The fruit was tested at various maturity levels. Because numerous characteristics change throughout fruit development, the ascorbic acid evaluation could be of a different characteristic correlating to ascorbic acid level.

2.1.1.7. Discrimination:

For fruit sorting, a qualitative assessment (e.g. classification technique) can be used rather than a quantitative measurement of the amount of an attribute. Many claims exist in literature about fruit population classification, such as based on cultivars or species, geographic locations, or production systems [60]. Without explicit a priori information of the NIR evaluated component, NIRS can be utilised to determine an attribute. Jacobs et al. [28], for example, used NIR spectra to determine the storage life of lettuce. However, robustness of such models needs to be proved with independent test data. For example, Vis-NIRS can determine whether the apple sample is acceptable or not, based on internal browning level [55].

2.1.2. NIR wavelength region:

In commercial practise, the SWNIR range (750-1100 nm) is employed to estimate internal characteristics of intact fruit, such as SSC and DM value, rather than wavelengths >1100 nm. However, many published papers on the evaluation of interior fruit features (57 percent,[25]) use equipment with longer wavelength ranges (>1000 nm). Use of extended NIR region instead of SWNIR, provides narrower and stronger absorption features which could lead to better results in the evaluation of internal quality attributes than the SWNIR. In the SWNIR band, however, the

water absorption coefficient is comparatively low, allowing for increased penetration depth inside fruit. Given the diversity in outer layer characteristics, the reduced effective penetration depth because of longer wavelengths could restrict model's robustness across independent populations. Even if the extended NIR region produced equivalent results, the Vis-NIR option will be chosen for commercial applications due to lesser hardware cost [61].

For a bell-pepper application, Vis-SWNIR produced superior findings than short wave IR reflectance spectroscopy [59]. Applications like carrot SSC [62], grape colour, phenolics, SSC, and ripeness stage [63], and persimmon astringency [64], Vis-SWNIR outperformed NIR. For assessing SSC of intact beets, the Vis-NIR range produced better results than the NIR range, while sugar beets slices produced nearly identical results [65], and for these wavelength ranges this is in agreement with the actual penetration depth. In contrast, [61] found that a sensor (using a reflectance geometry) working in the 850-1888 nm range performed better in the evaluation of apple characteristics than a sensor working in the 340-1014 nm region.

2.1.3. NIRS applications:

Vis-NIRS has been widely employed to estimate fruit maturity/quality parameters of many fruits including (but not limited to) apple, mango, guava, avocado, pear, persimmon, grapes, melon, citrus etc. Few of the reported studies have been summarized below:

2.1.3.1. Apple:

SSC is a critical quality metric in apples as it is related to maturity, harvest time, and taste [66]. It is critical to estimate SSC using efficient tools such as NIR spectroscopy to meet increasing consumer demand for high-quality apples. Brown et al. [67], demonstrate in spectroscopic studies that non-bruised tissues have a higher reflectance at Vis-NIR wavelengths than freshly bruised tissue. Bilanski et al. [68] and Pen et al. [69] expanded these fundamental studies by categorising wavelengths to

distinguish between non-defected and defected peeled apple tissue. Moons et al. [70] recognized a correlation among pH, SSC, and titratable acidity (TA) and NIR spectra. Lammertyn et al. [71] demonstrated through promising results that non-destructive estimation of apple internal quality attributes is possible. PLSR and PCR have been used to create a correlation between apple quality parameters and reflectance spectra. Standard error of prediction (SEP) of 0.068, correlation coefficient (R) of 0.93, SEP of 0.61, R of 0.82, and SEP of 2.49, respectively, has been reported for the pH, SSC, and firmness models. Whereas the prediction performance for mechanical texture properties was poor.

Zude et al. [72] used non-destructive spectral analysis to determine the chlorophyll content of two apple varieties, 'Elstar' and 'Jonagold'. The experimentation demonstrated good results for 'Elstar' using 650 to 730 nm range. Moreover, it has been recommended that wavelengths higher than 720 nm should not be used for the 'Jonagold' apple, this has been suggested because certain fruits have varying characteristics that have an inconsistent impact on light reflectance outside this wavelength range. Liu et al. [73] evaluated FT-NIR spectroscopy for predicting fruit quality of intact apples in 814 – 1100 nm range. For TA and SSC, chemometric models based on PLSR were developed. Liu et al. [74] evaluated the same FT-NIRS in interactance geometry (812-2537 nm) for Fuji apples.

During cold storage of 'Cripps Pink' apples, Mogollon et al. [75] discovered that the most effective wavelength range for detecting internal browning is 600–830 nm. As per existing literature, SSC and firmness are critical parameters in determining the maturity of Apple fruit. PLSR, is a widely used prediction method, alongside multiplicative scatter correction (MSC), Savitzky-Golay (SG) Smoothing, and principal component analysis (PCA) as pre-processing techniques. PLSR models provide encouraging results for estimating maturity of apple. SSC in apples has been predicted with good accuracy using NIRS, which is not the case with firmness. The applications of NIRS for the prediction of apple quality attributes are summarised in **Table *2.1***.

2.1.3.2. Mango:

Walsh et al. [76] discovered that the amount of DM in mango corresponded to amount of carbohydrate in the fruit. In terms of fruit maturity, it is considered to be an important indicator. The applications of NIRS for mango quality evaluation are summarised in **Table *2.2***. Lately, scientists have established spectroscopic approaches and handheld tools that can accurately quantify on-tree fruit [48], [76].

Intact fruit DM can be assessed on the tree using NIRS in accurate and non-invasive manner. Guthrie et al. [77] stated the use of a Multiple Linear Regression (MLR) model for single cultivar of mango at harvest maturity stage, while Saranwong et al. (2004) reported the use of a PLSR model for the same cultivar, with results of (R^2 = 0.96 and $RMSE_P$ = 0.79) and (R^2 = 0.89, SEP = 0.41, bias = 0.07), respectively. PLS multi-cultivar model for mango DM estimation created by Subedi et al. (2011) was used to accurately estimate samples of independent populations (R^2_P = 0.79, $RMSE_P$ = 0.97).

According to Spreer et al. [78], measurements of mango fruit's size can be taken non-destructively and linked to the fruit's weight (R^2 = 0.96, with an accuracy of 0.1 g). moreover, it is stated that non-destructive procedures make it possible to regularly monitor fruit on the tree to avoid sampling problems that are linked with destructive sampling.

Nagle et al., [79] have argued that change in intercellular distance could alter light scattering pattern, which could affect the calculation of NIR based DM. The calibration of water deficient fruits (R^2_C = 0.35, $RMSE_C$ = 4.6% DM) has yielded poor results using SWNIRS. Walsh et al., [76] employed handheld spectrometers with PLSR to measure DM of mango on-tree and proved their usefulness in making harvesting decisions. Internal browning based sorting of ‘Keit’ mangoes is done by Gabriels et al., [80] by comparing PLSR and ANN models, where ANN outperformed the PLSR model.

Table 2.1: Applications of NIRS for apple quality estimation

Optical Geometry	Spectrometer	Cultivar	Parameter estimated	Wavelength range (nm)	Data set size	Pre-processing algorithm	Estimation model	R^2	Ref
Reflection	Vis-NIR	Red delicious, Gala	SSC	800- 1100	960 + 800	PCA	PCR, Mahalanobis Distance (MD)	0.93, 0.96	[81]
			Firmness	400-1800				0.22, 0.79	
		Multiple (20)	DM	408-2498	229	SNV, SG 1-D	PLSR	0.94	[82]
		Multiple	SSC TA	400-1022	100	MSC, SG Smoothing, PCA	BPNN, GRNN, PSO	-	[83]
		Multiple (7)	SSC, DM	729-975	640	PCA	PLSR	0.89, 0.95	[84]
		Braeburn	SSC, Firmness	650-980	30	-	PLSR	0.81, 0.83	[85]
	FT-NIR	Golden Smoothee, Golden Delicious, Granny Smith	SSC, DM, TA	900-2500	960	-	PLSR	0.80, 0.83, 0.80	[86]
Interaction	Vis-NIR	Pink Lady	Flesh browning, SSC,	500-975	296	SG smoothing, PCA	PLS DA	0.86	[55]

			DM						
		Fuji	SSC	350-1030	150	MSC, SNV, SG Smoothing	SMLR	0.63	[87]
	FT-NIR	Fuji	SSC, TA, pH	812-2357	333	SG Smoothing	PLSR	0.91, 0.51, 0.58	[73]
		multiple	Glucose, Fructose, Sucrose	814-1100	130	MSC, SG Smoothing	PLSR	0.95, 0.97, 0.97 (R)	[74]
Transmission	Vis-NIR	Fuji	SSC, Water core degree	600-1000	663	MSC, SG Smoothing, SNV	SI-PLSR, SI-SPA-PLSR, GA-PLSR, CARS-PLSR	0.95, 0.94, 0.92, 0.98 (R)	[53]

Table 2.2: Applications of NIRS for mango quality estimation

Optical Geometry	Spectrometer	Cultivar	Parameter estimated	Wavelength range (nm)	Data set size	Pre-processing algorithm	Estimation model	R^2	Ref
Reflection	Vis-NIR	Tommy Atkins	SSC, Firmness, TA	1200-2400	80	SG 2-D	MLR, PCA, PLSR	MLR: 0.93, 0.82, 0.60	[88]
			SSC, DM, TA, Firmness	950-1650	400	SG smoothing, SNV, EMSC	PLSR	0.92, 0.67, 0.50, 0.72	[89]
		Caraboa	SSC, DM	700-1100	200	MSC, SG 2-D	MLR, PLSR	0.96, 0.97	[90]
				700-990	1200	derivative	PLSR	0.84, 0.77	[91]
		Osteen	Ripening index	600-1750	140	SG smoothing, EMSC		0.83	[92]
		Cogshall	SSC, DM, TA	800-2300	250	SG 1-D, 2-D		-	[93]
		Kenington Pride, Calypso	DM	300-1050	350	-		0.82	[49]
		Chokonan, Rainbow,	SSC	900-1700	80	EMSC	SVM, PLSR	0.95, 0.86	[94]

		Kai Te							
		Keitt	Internal browning	400-1000	576	Thresholding	PLSR, ANN	0.53, 0.57	[80]
	FT-NIR	Kent	SSC, TA, Ascorbic acid	1000-2500	58	SNV, MSC, SG 2-D	PLSR, PCR	0.66, 0.95, 0.61	[95]
Interactance	Vis-NIR	Palmer	SSC, DM	699-999 SSC, 699-981 DM	149	SNV, SG 1-D	PLSR	0.87, 0.84	[96]
			DM	699-981	200	-		0.82	[97]
Transmission	Vis-NIR	Sunshine	pH	300-1000	120	Baseline correction, MSC, SG Smoothing, Normalization	PLSR	0.93	[98]

2.1.3.3. Grapes:

The criteria for table grapes maturity is based on three critical attributes: SSC, TA, and pH. Although SSC and TA are important quality markers, there exist additional elements that significantly influence sensory judgement [99]. As seen in **Table *2.3***, the majority of researches have considered TA, SSC, and pH as grape quality indexes. Sensory evaluation is the most appropriate method for estimating the market acceptability of table grapes. NIRS is amongst the most effective non-destructive techniques for estimating the maturity of table grapes.

Fernandez-Novales et al., [100] employed transmission optical geometry with NIRS to predict SSC in 700–1060 nm range and obtained good results. Two spectral bands i.e. 650–1100 and 750–1100 nm, and two optical geometries were compared i.e. diffuse transmission and interaction, were utilised [101] to create PLSR models for SSC estimation in the three grape varieties 'Carmenere', 'Chardonnay' and 'Cabernet Sauvignon'. Models using the first spectral band (650-1100nm) had a R value more than 0.90 and an $RMSE_P$ value less than 1.2 for all cultivars. In comparison to interactance, the transmittance spectral analysis performed slightly better. As demonstrated by Larrain et al., [102], pre-processing procedures with reflectance spectra had no significant effect on estimating the SSC in the 640–1100 nm range. For SSC, PLSR models depicted good performance with R^2_P values from 0.93 to 0.96 and $RMSE_P$ values ranging from 1.01 to 1.27 for various varieties. Moreover, for pH estimation of several grape varieties the wavelength range of 640–1100 nm with reflectance spectra was used and PLSR models returned R = 0.75–0.89 and $RMSE_P$ = 0.088–0.16.

Cao et al., [103] estimated SSC, pH and variety of three table grape cultivars in the 400–1000 nm range. The genetic algorithm (GA) was used for wavelength selection and the least squares support vector machine (LSSVM) with an R^2_P of 0.98 and an $RMSE_P$ of 0.13. Omar et al., [104] assessed SSC and pH of table grapes and revealed that from 922–923 and 990–995 nm wavebands are most significant for SSC and pH assessment. The optimum result was an MLR model based on the three wavelengths: 605, 923, and 990 nm (R = 0,87; RMSE = 0,11). Additionally, the

anthocyanin concentration in several grape varieties was also determined using a PLS-based model in the 640–1100 nm band, with good results (R = 0.79–0.83). **Table *2.3*** summarises the uses of NIRS for grape quality evaluation. From **Table *2.3***, it can be seen that to eliminate noise and fluctuations in the data, pre-processing algorithms such as smoothing, MSC, SNV, and Savitzky-Golay (SG) derivative filters have been used. PLS has demonstrated superior outcomes in terms of R^2 values when compared to other regression algorithms.

Table 2.3: Applications of NIRS for grape quality estimation

Optical Geometry	Spectrometer	Cultivar	Parameter estimated	Wavelength range (nm)	Data set size	Pre-processing algorithm	Estimation model	R/R^2	Ref
Reflection	Vis-NIR	Graciano	Phenolic compounds	1100-2000	84	MSC, SNV, detrend	MPLSR	0.98 (R)	[105]
		Cabernet Sauvignon, Merlot, Syrah, Carmene`re	OH	450-1660	650, 450, 450, 225	SG 1-D	PLSR	0.93, 0.92, 0.98, 0.81 (R)	[106]
		Nebbiolo	SSC, TA, pH	450-980	71 fresh, 156 homogenized	MSC, SG 1-D, 2-D, Smoothing	PLSR	0.82, 0.81, 0.81 (R)	[21]
		Manicure finger, Ugni Blanc	Surface color, SSC, Total phenolic compounds	400-1000	270 berries at 5 stages	MSC, SNV	LS-SVM	0.870 (SSC), 0.872 (TP) R^2	[63]
		Grenache	Amino acids	570-1000 1100-2100	4480 berries	MSC, SNV, detrend	MPLSR	0.60 (R^2)	[107]
		Syrah, Cabernet Sauvignon	TSS, Total anthocyanins,	450-1800	432, 576 berries	Median filter, Moving average filter,	PCR, MLR, PLSR	R^2> 0.90, R^2> 0.90, R^2> 0.70,	[108]

			yellow flavonoids			SG 1-D, 2-D, MSC			
		Tempranillo, Syrah	Total phenolic, Anthocyanin, Flavanols	908-1676	400 berries	-	MPLSR	-	[109]
		Cabernet Sauvignon, Sangioves, Merlot	Reductant sugars, TA, Anthocyanin, pH, maturity index (reductant sugars/TA)	400-1000	30 berries/ bunch 45 berries/day Total 4 days	-	PLSR	0.75, 0.78, 0.61, 0.70, 0.80 (R)	[110]
		Shiraz, Cabernet, Chardonnay, Merlot e Pinot	SSC	400-1000	-	Averaging, Normalization, transformation	MLR	R^2 0.95 for red, 0.83 for white varieties	[111]
	FT-NIR	Prime Seedless, Thompson Seedless,	TSS, TA, TSS/TA, pH, BrimA	800-2500	338 bunches	Moving Smoothing windows + MSC, SNV, SG 1-D, 2-D, MSC	PLSR	0.71, 0.33, 0.57, 0.28, 0.77	[112]

		Regal Seedless,						(R^2)	
Interactance	Vis-NIR	Cabernet Sauvignon, Carménère, Merlot, Pinot Noir, and Chardonnay	SSC, pH, Anthocyanin concentration	640-1100	60 berries or each variety / week for 4 months	-	PLSR	0.930, 0.923, 0.921, 0.875, 0.874, (R^2 values of SSC for each variety)	[102]
		White Malanga	TA, pH, SSC, Firmness, No of Seeds	800-1100	-	SG 2-D	PLSR	0.897, 0.632, 0.980, 0.661, 0.903 (R^2)	[113]
Transmission	Vis-NIR	Caignan, Mouerdre, Ugni blanc	TA	680-1100	371	-	PLSR, MLR, LS-SVM	0.76, 0.69, 0.83 (R^2)	[114]
		Jufeng	SSC	200-1100	115 bunches	MSC	PLSR	0.83 (R^2)	[115]
		Albarino	Volatile compounds	190-2500	52 berries	SNV, SG-1D, 2-D	PLSR	0.85 (R^2)	[116]
	FT-NIR	Grape must, White wine	Calcium	1100-2300	60, 98	Baseline correction, MSC, SNV, SG 1-D	PLSR	0.91, 0.93 (R^2)	[117]

2.1.3.4. Melon:

SSC assessment using NIRS may be more difficult for fruits having thick peel and large size, for example melons and watermelons. Depending upon the type of fruit and intensity of light, the NIR radiation can penetrate to a depth of approximately 4–20 mm within the fruit, which may be appropriate to measure the SSC for fruits with thin peels and small sizes, as significant spectral information can be collected. However, while melons have a thick peel, the peel thickness varies from stem end to calyx in melons. As a result, if the location having the thickest peel is picked to capture spectral data, sufficient information about fruits inner characteristics is not acquired. A suitable representative location for spectra collection is necessary in this scenario.

Guthrie et al. [118] built first SSC-prediction model using NIR spectroscopy in reflectance mode for intact melons. A total of 100 melons of three different varieties (El Dorado, Eastern star and Hammersly) were harvested on three different dates. A combined calibration model built using MLR performed good with a data set taken from all three harvests (R^2 = 0.72, MSE_P = 1.84 °Brix). Sugiyama et al., [119] investigated the relationship between SSC and NIR spectra (collected in interactance mode) in melons. Six Andes melons at three ripeness stages (unripe, mature and fully mature) were investigated. Sugar distribution maps for melons were generated at 676 nm wavelength, which depicted that sugar levels rise gradually from melons rind towards inner pulp. Greensill et al. [120], accomplished transfer of prediction models for melons SSC between Fourier transform NIR (FT-NIR) spectrometer (630–1030 nm) and scanning-grating based NIR instruments using various chemometric techniques. Two populations of Rock melons i.e. Dubloon (201 samples) and Navajo (198 samples) were used. Guthrie et al. [121] tested robustness of multivariate calibration models for SSC of 100 rock melons using NIR spectroscopy (695–1045 nm). The SSC of mesocarp was reported to be highest around the equator and increasing towards seeds cavity and, decreasing towards stem and calyx end of the melon but more so towards the stem. Equatorial region was selected as the representative region for collection of spectra and destructive testing for SSC. Modified PLSR (MPLSR) was used with second derivative pre-processing of

absorbance data reported to perform better than standard partial least squares regression (PLSR). Long et al. [122] also investigated the variation of SSC within same sample in Eastern star and Malibu melons (total 149 samples), and in this regard supported the previous discussion of Guthrie et al. [121]. Long et al., [122] added that cold storage of fruit (0–14 days) does not affect NIR model performance.

Flores et al. [123] also built NIR prediction models using SSC as quality index in intact and cut melons and watermelons. Examined variety included Cantaloupe and Galia melons. Suh et al. [124], [125] concluded that interactance mode of NIRS provides more information of fruits internal attributes as compared to reflectance and transmittance modes, to predict SSC and firmness in thick skinned fruits such as muskmelons. Tian et al. [126] conducted experiments to predict SSC for two varieties of melons i.e. Baipi and Hetao, using transmittance mode of NIRS (350–1000 nm). Three spectral pre-processing methods i.e. first derivative, second derivative and Norris derivative smoothing were used for creating the PLSR model and principal component regression (PCR) model. Sanchez ´ et al. [127] used NIRS to estimate melon pulp color in Cantaloupe and Galia melons. MPLSR was used along with several pre-processing techniques and principle component analysis (PCA).

Khurnpoon et al. [128] concluded that only two of the textural properties i.e. rupture force and penetrating force for melons pulp can be estimated by NIRS using FT NIR spectrometer in reflectance mode. Three varieties of melons including Manao, Jinhongbao and Xizhoumi (120 samples each) were tested [10] using NIR spectrometer in reflectance mode. SSC calibration model was built using spectral data collected at stylar and equator ends. Hu et al., [129] and Zhang et al. [130] have done a comparison between different positions for acquiring spectra and different models for predicting SSC in Hami melons. Three local models (at calyx, equator, and stem position) and one global model was built using PLSR with different pre-processing techniques. The results concluded that the models generated from data acquired from equator-region and the global data (from all three positions) depicted similar performance and better than the calyx and stem regions-based models. Hence, it was recommended that equator position is suitable for spectral data collection in intact Hami melons in order to predict SSC.

Sensory test was conducted by Lu et al. [131] to establish internal quality standards for Elizabeth, M-1, M-3 and M-5 variety of melons. A total of 16 melons (4 melons of each variety) were tested by a panel of 10 judges. Satisfactory class samples included melons with SSC over 12 °Brix and firmness 4–5.5 kgf cm^{-2} whereas melons over 10 °Brix and firmness 3.5–6.5 kgf cm^{-2} were considered average class samples. NIRS predictive models for SSC and firmness were built using wavelength selected genetic algorithm for mix cultivars using spectral data collected from stylar end of samples. Firmness model was reported to be inferior than SSC predictive model. **Table *2.4*** lists the applications of melon quality estimation reported in literature using NIRS.

2.1.3.5. Other fruits:

Attempts have been made to non-destructively estimate fruit internal characteristics of avocado, pear, peach, persimmon, mandarin, guava, banana and strawberries etc using Vis-NIRS (see **Table *2.5***).

Table 2.4: Applications of NIRS for melon quality estimation

Optical Geometry	Spectrometer	Cultivar	Parameter estimated	Wavelength range (nm)	Data set size	Pre-processing algorithm	Estimation model	R/R^2	Ref
Reflection	Vis-NIR	Hami	SSC	550-950	120	1) None 2) Smoothing 3) Smoothing + MSC 4) Smoothing + normalization 5) Smoothing + mean centering	PLSR, MLR, LSSVM	0.89, 0.89, 0.88, 0.93, 0.89 (R)	[129]
						Smoothing-MSC Smoothing-normalization Smoothing-SNV	PLSR, CARS, UVE, CARS-SPA, UVE-SPA	0.91, 0.914, 0.92 (R)	[130]
		El Dorado, Eastern star, Hammersly		700-1100	100	SG 2-D	MLR	0.82, 0.44, 0.70 (R^2)	[118]
		Cantaloupe, Galia		515-1650	1000	--	MPLS	0.74 (R^2)	[123]
		Manao, Jinhongbao, Xizhoumi		750-950	360	MSC, SG 1-D	PLSR	0.83 (R^2)	[10]

		Cantaloupe, Galia	Pulp color	535-1650	432	SNV, Detrending, Window filtering (1,5,5,1)	MPLS	a*- 0.97 b*- 0.88 c*- 0.85 h*-0.97 (R2)	[127]
	FT-NIR	Reticulata CV. Green net	Rupture force, Penetrating force	800-2500	249	Constant offset elimination, straight line subtraction, vector normalization, min-max normalization, multiplicative scatter correction (MSC), first derivatives, second derivates, first derivative + straight line subtraction, first derivative + vector normalization, first derivative + MSC	PLSR	0.85, 0.84 (R^2)	[128]

Interactance	Vis-NIR	Muskmelons	SSC, Firmness	475-1100	256	1) normalization with mean 2) normalization with max 3) normalization with range 4) MSC 5) median filter (size 3,5,7) 6) SG 1-D (size 3,5,7) 7) SG 2-D (size 3,5,7) 8) Norris Gap 1-D (size 3,5,7) Norris Gap 2-D (size 3,5,7)	PLSR	0.710 0.732 0.755 0.702 0.700,0.699, 9.701 0.386, 0.594, 0.627 0.014, 0.003, 0.012 0.619, 0.499, 0.667 0.125, 0.243, 0.412 (R^2)	[124]
		Elizabeth, M-1, M-3, M-5		500-1010	16	SG (25,3)	PLSR	0.832, 0.573 (R^2)	[131]
Transmission	Vis-NIR	Baipi, Hetao	SSC	350-1000	-	SG 1-D	PLSR, PCR	0.907, 0.921 (R)	[126]

Table 2.5: Applications of NIRS for fruit quality estimation

Optical Geometry	Spectrometer	Fruit	Cultivar	Parameter estimated	Wavelength range (nm)	Data set size	Pre-processing algorithm	Estimation model	R^2	Ref
Reflection	Vis-NIR	Kiwi	Hayward	SSC, Firmness	400-2450	3840	SG smoothing	PLSR, SVM	0.83, 0.60	[132]
		Peach	Akatsuki	Firmness	500-1000	40	SG 2-D	PLSR	0.80	[133]
		Kiwi	Xixuan, Huayou	SSC	860-1700	200	SNV	PLSR, LSSVM	0.76, 0.92	[134]
		Pear	Cuiguan, Huanghua, Qingxiang	Firmness	350-1800	330	MC-UVE-SPA	PLSR, LSSVM	0.93, 0.93	[135]
		Peach	Calrico	Firmness	400-1060	260	SNV, MSC, SG 2-D	PLSR, LSSVM	0.76, 0.78	[136]
		Mandarin	Clemevilla	Firmness, SSC, pH, TA	1600-2400	256	SNV, Detrending	MPLSR	0.28, 0.64, 0.79	[137]
		Avocado	Hass	DM, Oil Content, Moisture content	700-2500	155	SNV, MSC, SG 2-D	PLSR	0.84, 0.84, 0.58	[138]
		Pear	Abate, Cascade,	SSC, Firmness	500-1010	240	MSC, Standardization,	PLSR, MLR	0.84, 0.67,	[139]

			Conference, Wujiuxiang, Red Comice				SG 2-D		0.84, 0.61	
	FT-NIR	Kiwi	Hayward	SSC, DM, Firmness	850-2500	100	SNV, MSC, SG 2-D	PLSR	0.91, 0.83, 0.89	[140]
		Peach	Aurora-1	SSC, Firmness	1000-2500	530	SNV + Detrend, MSC, SG 2-D	PLSR	0.45, 0.40	[141]
		Peach	Baby Gold	SSC, pH, TA	800-2500	50	SNV, MSC, SG 2-D	PLSR	0.84, 0.52, 0.43	[142]
		Orange	Penggan	SSC, pH	1000-2500	600	Multiple	PLSR	0.835, 0.735 (R)	[143]
Interaction	Vis-NIR	Pear	Yunhe	SSC	590-1091	134	SNV, MSC, SG 1-D	PLSR	0.90	[144]
		Kiwi	Saehan	DM, SSC	729-975	100	SG 2-D	PLSR	0.73, 0.73	[145]
		Banana	Robusta	DM, SSC, Firmness	500-1050	204	-	PLSR	0.88, 0.97, 0.98	[52]
		Mandarin	Nanfeng	SSC, TA, Vitamin C,	400-1040	153	SG 1-D, 2-D, MSC	SVM, BPNN, PLSR	0.93, 0.66, 0.81,	[146]

				Surface color					0.57 (R)	
		Avocado	Shepard and Hass	DM	720-975	690	SG 2-D	PLSR	0.88	[147]
		Orange	Shatangju, Huangyanbendizao	SSC	400-1000	300	SG Smoothing, MSC	PLSR	0.837, 0.866 (R)	[148]
Transmittance	Vis-NIRS	Mandarins	Satsuma	Citric acid content	-	-	SG 2-D + mean centring	PLSR	0.83 (R)	[149]
		Orange	Navel	SSC	550-980	195	multiple	PPSO-PLSR	0.788	[150]

2.1.4. NIRS instruments:

2.1.4.1. Optical Geometry:

There are three types of optical geometries i.e. reflectance, partial transmittance or interactance and full transmission. Recognizing the fruit's specular as well as diffuse reflection, a reflection geometry allows the detector to see an illuminated portion of the fruit. Internal characteristics of the sample are not revealed by specular reflections. Detection of specular reflections can exceed detection of diffusely reflected light in an incorrectly designed system. Incident rays illuminating an object at 45 degrees to the plane of the object will reduce amount of specular rays received by the detector. To assess fruit quality, the mesocarp, rather than the superficial layers should be the source of the diffusely scattered light detected by a reflectance geometry, to avoid biassing the results.

Detector, sample and lamp are organised in a way that light must travel from the sample to reach the detector, with no direct passage of light between lamp and detector. Arrangements like this can have a 180° angle between lamp, sample and detector ('full transmission' geometry) or smaller angles ('interactance' or 'partial transmission'). With moving fruit on a packline, the 'shadow probe' configuration [120] is used for contactless sample evaluation. It is a specific partial-transmission geometry in which, probes are installed in front of the collinear light source, thus a shadow is formed on the fruit as viewed by the probe. As a result, the detected signal does not depends on detector to sample distance. Optical density and tissue inhomogeneity in many fruits limit the application of a full transmission geometry. As a result, powerful light sources and/or precise detectors are required for the former issue, while the latter results in an optically sampled volume that does not match the desired attribute (e.g., flesh TSS).

The optical geometry chosen for a particular postharvest activity should be guided by an examination of the fruit structure in relation to the spectrophotometric system's optical geometry. Although a vast number (75%) of research published since 2015 [25] utilise Vis-NIRS to analyse internal characteristics of fruit use a reflection geometry, still no commercial spectrophotometer designed for fruit applications uses this geometry. This could

be due to the apparatus available in research facilities, with reflection geometry predominating in food applications [151]. The majority of food applications utilise a standardized sample for which a surface (reflection) examination can be used to represent the entire sample. Future papers on the development of applications utilising Vis-NIRS for fruit characteristic evaluation should include a justification for the optical geometry used.

Vis-NIR spectroscopy of undamaged fruit demands light to pass through the 'skin' of the fruit (usually the exocarp and non-edible mesocarp), the edible mesocarp, and back out of the skin. Spectra recorded with a reflection geometry will largely contain data about the surface of the sample in terms of diffusely reflected and specular light. Variation in the morphology and surface characteristics of the fruit will affect the amount of specular reflection, hence altering the perceived absorption levels. Due to the fact that specular light is similar to incoming light, its influence to the perceived absorption spectrum can be reduced 'physically' by polarising filters or numerically using pre-processing techniques such as derivatives. However, the latter procedure does not increase the signal-to-noise ratio (SNR) of diffuse reflection portion.

The bulk of diffuse reflected light will originate within 5 mm of the surface of apple fruit [46], [152], and from a shallower depth for fruits with a thicker (greater scattering factor) exocarp, like avocado. Internal attribute evaluation using reflection geometry demands a connection between the internal characteristic and the spectra of the fruit's external layers.

As compared to reflection geometry, transmission geometry is more suitable for gathering useful information from more depth within fruit. Schaare and Fraser [153] compared the performance of interaction, reflection and full transmission geometries for evaluation of SSC in intact Kiwi. They found interaction geometry to be most useful in their assessment as reflection geometry spectra contained superficial layer information due to specular and diffuse reflection and full transmission geometry spectra had low SNR and also contained information about central seeds and tissues. As reflection geometry highly depends upon surface reflectance instead of internal reflection, any change in surface properties will affect the results. Thus, chemometric models based on reflection geometry are expected to depict poor robustness when predicting independent populations [25].

Internal faults are typically evaluated using a acceptable/non-acceptable (discriminant) judgement. However, the imperfection may not be distributed evenly across the fruit. A reflection geometry entails evaluating a single section of the fruit. Using a full transmission geometry, light is distributed throughout the fruit, including some reflection from the inner surface of skin, and hence this geometry enables evaluation of a bigger fraction of the fruit content [154]. As a result, this geometry may produce a more beneficial outcome than that of other geometries, as demonstrated in the evaluation of internal browning in apple and pear [55], [155], as well as internal decay in citrus [156].

It can be concluded that NIRS technique can be successfully applied to thin skin intact fruits having a homogenous pericarp where the optical geometry applied matches to fruit structure.

2.1.4.2. SWNIR in-line equipment:

Constraints regarding in-line applications include the sample assessment speed suitable to conveyer belt speed, operating conditions, and the geometry of conveyer belt. As compared to in-field use, light conditions can be reasonably controlled in the in-line applications. Main issues include:

- Effect of sample movement during the interval of measurement
- Appropriate optical geometry to reduce the effect of change in fruit size and shape
- Spectra collection from a representative section of the fruit, where attributes' natural distribution within fruits may vary
- Robustness of the models

Japan first introduced in-line commercial use of NIRS technology in late 1980s / early 1990s, for sorting application in fruits on attribute such as SSC. In-line commercial equipment was initially introduced by the Japanese companies like Mitsui Metals and Mining, Sumitomo, Fantac and Emitec. Initially the systems were based on reflectance geometry which was soon replaced by partial or full transmittance geometry, as the reflectance geometry had model robustness issues. Except the Sumitomo system which used diode lasers as a light source, all other systems employed

halogen lamps. The use of diode lasers was later discontinued due to its cost and laser output intensity stability issues. In the early 2000s, many new in-line systems were manufactured by the global grading equipment manufacturers i.e. Color Vision Systems (Australia, now part of MAF RODA, France) and Compac (now Tomra, Norway), Greefa (Holland), Multiscan technologies (Spain), Aweta (Holland), Sacmi and Unitec (Italy).

Current state of the art commercial systems sort upto 10 fruits per second (depending on fruit size and simultaneous operation at multiple lanes) at an assessment speed of 1 $m.s^{-1}$. Such systems employ SWNIR spectrometers with a partial or full transmittance geometry. There is lack of performance benchmarking of commercial in-line systems in literature because of commercial concerns except [26]. Although researchers have tried to mimic in-line system issues for example, Ignat et al. [61] investigated the performance of PLSR models developed with the use of two different spectrometers for static and moving apples. However, such attempts have used either wavelength region of >1050nm or a reflectance geometry (e.g. [61], [157]), both of which are not used in standard commercial systems.

In comparison to machine vision and weight-based segregation (i.e. load cells), SWNIRS adoption in in-line commercial systems lacks behind. The probable reasons include accuracy, complexity and value addition to the current systems. As far as accuracy is concerned, NIRS based systems indirectly measure fruit attributes unlike the direct assessment in weight-based systems, hence, NIRS application depends upon the strength of the indirect correlation of NIRS with the attribute of interest. Moreover, to increase reliability of the NIRS system, the calibration models need to be updated regularly [26] which increases operational complexity. And lastly, value creation can be observed in Japanese fruit markets which offered rewards for top quality fruit. This incentive served as a motivational factor for manufacturing of commercial in-line NIRS systems by multiple industries in Japan, however, at extremely high prices. In most of the markets (e.g. in western countries), fruit rejection (with a strong penalty) is done on the basis of internal browning instead of SSC values. Hence, NIRS based defected fruit sorting is being promoted in Western markets regardless of the degree of uncertainty involved [55].

2.1.4.3. SWNIR devices for handheld use:

Many handheld NIR spectrometer-based models exist that are designed to be used with the fruit and predict its attribute's value. The basic difference between in field use and indoor use is the ability of NIRS device to tolerate changes in ambient light conditions and temperature levels. Many generic small handheld spectrometers (not specific to fruit) are available with guiding material available on how to do fruit evaluation with them [25]. In early 2000s, handheld devices purposely built for in-field use became commercially available for example, NIRGun by Fantec from Japan, pigment analyzer by CP from Germany, Sacmi from Italy, Nirvana by integrated spectronics from Australia and Felix instruments from USA. Few of them have left the field like NIRGun, pigment analyzer and Nirvana, many new handheld devices have been released like SunForest and GWon from South Korea, Atago Hikari and FHK from Japan [25].

Many other detector techniques also exist [25] within generic spectrometric devices. There are two types of detector technologies in handheld NIR spectrometers i.e. single detector and array detector [158]. VIAVI MicroNIR 1700 by Santa Rosa from USA is probably the first commercial handheld NIR spectrometer used in many applications of qualitative and quantitative assessment of seafood, pharmaceutical drugs, soil hydrocarbon contaminants and food nutrients. It employs an array detector within wavelength range 908-1676 nm along with a linear variable filter (LVF) as monochromator. However, single detectors have low cost as compared to array-based detectors. Hence, many new advancements were seen in NIRS systems with single detector. The DLP NIRscan Nano EVM by Dallas TX from USA uses a single element detector with digital micro mirror device covering wavelength range of 900-1701nm. Another MEMS based FT-NIR system (wavelength range 1298 to 2606 nm) containing single chip Michelson interferometer was introduced by the Si-ware systems from Egypt. The FT-NIR technology offers multiple advantages over other technologies like higher optical resolution and ease of model transfer between instruments. These technologies have not been implemented in fruit-specific spectrometers with the proper optical geometry, wavelength range, and user interface. It's not out of the question that this kind of change will occur within the next decade.

Majority of fruit spectrometers employ tungsten halogen lamp as a light source, a CMOS linear array detector or grating and Silicon photodiode operating in the 350-1100 nm with a

interactance optical geometry. Variations in ambient light can be accommodated by referencing each sample. Amongst all commercial handheld devices, F-750 produce quality meter by Felix instruments [159] is a commercial handheld non-destructive fruit quality meter intended for use with fruit in the field. It employs a halogen lamp as light source, equipped with a Carl Zeiss MMS1 spectrometer with wavelength range 285-1200 nm (+/- 10nm) in interactance mode and spectral sampling of 3nm and spectral resolution of 8-13nm. It is equipped with a GPS to record spatial coordinates for possible generation of orchard maturity maps and provides referencing on every sample to cater ambient light variations. The interface is user-friendly with the ability to develop chemometric models for a particular fruit cultivar. It is a research-oriented device from Felix Instruments. It gives the data scientist the ability to capture raw spectra, apply different pre-processing filters, develop statistical models for fruit quality estimates, and test those models in the field. It takes extensive research and training before F-750 can be used in the field as a portable fruit quality estimator. There are at least a total 40 research articles published recently that use F-750 as a benchmark to capture spectral data for research and developing models for estimation of different fruit quality estimates [96], [160], [161], [162], [163], [164], [165], [166], [167], [168], [55],[169],[170],[171],[97],[172],[173]. Felix Instruments have recently also started ready to use the F-751 series for only a few fruits i.e., Mangoes, avocados, and Melon. The F-750 is used as the industry standard (benchmark instrument) e.g. Australian mango industry association [27].

Low-cost LED based instrumentation is also available for assessment of specific pigment in fruit. For instance, the Atago Hikari uses LEDs as a light source within the SWNIR region. Advantage of using LEDs as a light source is less power consumption and negligible heat dissipation, thus there isn't any need for heat dissipation systems used in the case of halogen lamp. The DA meter, Kiwi meter and Cherry meter [174] are handheld devices that measure chlorophyll or red pigment detection at two or three wavelengths (DA meter: 670 nm and 720 nm; Kiwi meter and Cherry meter: 560 nm, 640 nm and 750 nm). The Multiplex330 [175] is also a handheld device employs three LEDs i.e. 435 nm, 685 nm and 735 nm, to measure chlorophyll, flavonols and anthocyanins content of leaves and fruits. The FIORAMA [176] handheld device uses 5 LEDs of wavelengths 535 nm, 570 nm, 685 nm, 720 nm and 950 nm to measure chlorophyll and red pigments.

2.2. Chemometrics

NIRS measures the absorption of light associated to the vibration of molecular bonds when exposed to light radiations thus providing chemical data. Chemometrics is a data driven science that uses mathematics and statistics to extract useful information out of this chemical data. All the overtone (absorption) features acquired via NIRS are broad and overlapping which makes the interpretation of the acquired spectra in raw form, difficult. When NIRS was introduced as an alternative to infrared spectroscopy, it was made possible by the invention of chemometrics, which made it possible to fetch out relevant information from the acquired spectra. It typically involves feature extraction (i.e. data pre-processing) and machine learning based model development (**Figure 2.2**). Sections below explain each of these in detail:

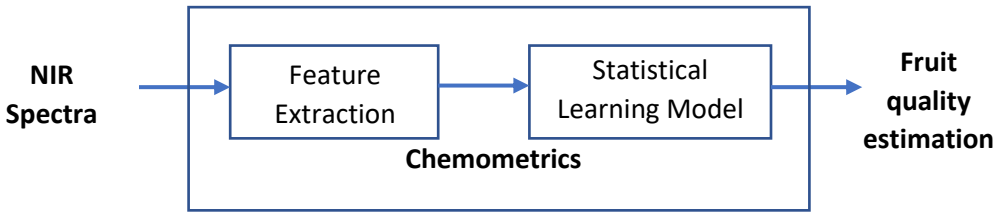


Figure 2.2: Main elements of chemometrics block of the decision support system

2.2.1. Feature Extraction:

The NIR spectra represents the pattern of NIR radiation absorption by the sample. Factors affecting the shape of NIR spectra are:

1) Due to chemical characteristics of the sample, each wavelength of light will get absorbed differently by the sample. Mostly, this is the desired signal which can be further interpreted as it relates to the feature of interest.
2) Depending on the wavelength, light scatters at different angles due to different particle sizes.
3) Differences in path lengths among samples because of positioning variations and sample surface irregularities. The scattering effects due to particle size and due to different path length are the main cause of variations in the NIR spectra.

Additive effects such as differences in path lengths introduce a shift in spectrum baseline along the vertical axis. While multiplicative effects such as due to particle size alter the slope of the spectrum. The purpose of data pre-processing is to eliminate effects that are unrelated to the characteristics of the target sample, which would result in better features and thus better data modelling. Many pre-processing techniques are available in literature that can be applied on spectral data before the model development. Researchers in the past have used these pre-processing algorithms (individual and in combination, both) to optimize their modelling results (Anderson and Walsh, 2022). Standard pre-processing algorithms include two main categories i.e. scatter correction algorithms (mean centring, standard normal variate (SNV), and multiplicative scatter correction (MSC)) and derivative algorithms. Wavelength selection (WS) methods have also been used in literature with chemometric data to reduce redundancy in data and thus number of features resulting in reduced complexity of modelling algorithms [177], [178].

Initially, the research reports in this field presented extensive comparisons of pre-processing algorithms to optimize the results. However, now it has been accepted that the optimum pre-processing technique (and its parameter settings) depend upon instrumentation used and the application. Hence, the trend in current reports is to just mention the used (optimum in their case) pre-processing technique along with parameter settings employed [177]. Many reviews are available in literature that summarize the pre-pre-processing techniques used in horticultural applications e.g. [179]–[181]. Some of the most widely used pre-processing techniques are explained in section below:

2.2.1.1. Standard normal variate (SNV):

When NIR radiations interact with molecules of the target sample, the resultant scattering of light often generates a shift in absorbance levels. As a result, it may become more difficult to interpret and model the resulting spectra. Light scattering causes variations in path lengths that result in varying background signal levels (variation with the wavelengths creating baseline shift and curvature) within the sample and among different samples [182].

SNV pre-processing transformation reduces the scattering effects and differences in global signal intensities by subtracting the mean of the spectrum from each spectrum and then dividing each spectrum by the standard deviation of the spectra.

$$x_{i.j(SNV)} = \frac{(x_{i,j} - \mu_i)}{\sigma_i} \quad (2.1)$$

Where $x_{i,j(SNV)}$ is the transformed element, $x_{i,j}$ is the raw spectra element of the ith spectrum at variable j, μ_i is the mean of the ith spectrum and σ_i is the standard deviation of spectrum i. SNV benefits include:

1) Correction of baseline shift and global variations in intensity
2) Improvement in PLSR results when used with NIR spectra with scattering effects

SNV drawbacks include:

1) SNV resulting spectra has 0 centred positive and negative values, which may make interpretation more difficult
2) SNV works on the assumption that multiplicative effects are uniformly distributed over the whole spectra, which may not be the case in all applications, hence it may introduce more artifacts.

2.2.1.2. Multiplicative scatter correction (MSC):

MSC is a demanding pre-processing technique that requires a reference spectrum which is ideally free of all scattering effects. Getting such reference spectrum is a challenge itself. Usually that reference is obtained by taking average of all the spectral observations. MSC is done in two steps [183]:

1) An ordinary least squares regression equation of the form given below is used to regress each spectrum x_i against the reference spectrum x_m:

$$x_i = a_i + b_i \,.\, x_m \quad (2.2)$$

2) The MSC corrected spectrum is then calculated as:

$$x_{i(MSC)} = \frac{(x_i - a_i)}{b_i} \quad (2.3)$$

2.2.1.3. Spectral derivatives:

Spectral derivatives remove both the additive and multiplicative effects in the spectra. The first derivative only eliminates the additive effect (baseline shift) while the second derivative eliminates both the additive and multiplicative effects (baseline and slope). There are two different spectral derivative methods i.e. Norris William (NW) and Savitzky-Golay (SG). Both these methods first perform smoothing to avoid reduction of signal to noise ratio (SNR) in corrected spectra. Finite difference method is the main method behind finding derivatives. For first derivative the difference of two consecutive spectral measurements is calculated as [183]:

$$x_i' = x_i - x_{i-1} \tag{2.4}$$

Second derivative is calculated by taking difference of two consecutive points of first derivative i.e.

$$x_i'' = x_i' - x_{i-1}' = x_{i-1} - 2\,.\,x_i + x_{i+1} \tag{2.5}$$

In literature, the most widely used spectral derivative pre-processing technique is the SG method. Savitzky and Golay [184] introduced a method of numerical derivation including the smoothing. A symmetric window size is selected and then a polynomial is fitted on the raw data within that window to find derivative at the central point of the window. The parameters of the polynomial are then calculated, which can be used to calculate derivative of any order of this polynomial analytically. The derivative equation is used to calculate the derivative of the centre point of the window. This method is repeated for all points in the spectra. The window size and polynomial order are to be decided based on the application. The polynomial order defines the highest derivative that can be calculated for example, a fourth order polynomial can be used to find fourth order derivative.

2.2.1.4. Principal component analysis (PCA):

The spectroscopic sensors capture information about the sample in the form of light signals at multiple wavelengths. In multivariate data analysis, first step is to identify clear differences in the captured signal among different samples. However, in case of NIRS and SWNIRS, the overtone features are so broad and overlapping that there exists strong correlation among wavelengths which makes it difficult to interpret the raw spectra. For good modelling of data, the input features to a machine learning model should be uncorrelated with each other. Correlation among features provides no extra information to the model about the data to be modelled. Thus, it is desirable to reduce the redundancy in the features, which can be achieved by reducing the number of features while maintaining maximum variation in the data using some technique like principal component analysis (PCA)[183].

For a matrix **X** with dimensions (m x n), m number of rows corresponding to observations and n corresponding to absorbance values (features) at n number of wavelengths. PCA transforms this m x n dimensional space having 'n' number of correlated variables into a m x p dimensional space having 'p' number of uncorrelated variables (called as the principal components 'PCs'), which are linear combination of the original variables. The matrix **X** is decomposed as:

$$\boldsymbol{X = SL^T + E} \tag{2.6}$$

Where **S** is the scores matrix of dimensions (m x p) containing the coordinates for 'm' samples in the new space (coordinate system) created with 'p' uncorrelated variables. While **L** is the loading matrix of dimensions (p x n) containing the contributions of the 'n' original variables to the 'p' new variables and T represents transpose operation. **E** is the residual error matrix. Mathematically, an eigen vector decomposition is performed of the matrix $\boldsymbol{X^T X}$ to get an approximate $\boldsymbol{\hat{L}}$ of loading matrix $\boldsymbol{L}$**.** The approximate for score matrix $\boldsymbol{\hat{S}}$ is calculated by regressing **X** onto $\boldsymbol{\hat{L}}$**.**

PCA is performed in the following steps [185]:

1) Normalization:

This step normalizes the range of input variables so that each variable contributes equally to the analysis. If the ranges of input variables have significant differences, the variables with more range will have more influence than those with small ranges, this leads to unjust results. This is mathematically done by:

$$q' = \frac{q-\mu}{\sigma} \tag{2.7}$$

Where μ is the mean and σ is the standard deviation.

2) Computation of the covariance matrix:

This matrix computes the relationships (correlations) between variables. It is a *p x p* matrix (where p is the number of input variables). The entries of this matrix are the covariances associated with the input variables. For instance for a 3-dimensional input data having three variables a, b and c, the covariance matric will be given as:

$$\begin{bmatrix} Cov(a,a) & Cov(a,b) & Cov(a,c) \\ Cov(b,a) & Cov(b,b) & Cov(b,c) \\ Cov(c,a) & Cov(c,b) & Cov(c,c) \end{bmatrix} \tag{2.8}$$

3) Computation of the Eigen vectors and Eigen values:

Next step is to compute the eigen vectors and eigen values of the covariance matrix. The eigen values and eigen vectors will determine the principal components (PCs). PCs are the new set of variables constructed using linear combinations of input variables. The PCs are uncorrelated amongst each other, and the first few components contain maximum information about input variables. Which means that a x dimensional data will have x number of PCs and the first PC will have maximum variance (information about input data) and then the second PC will have second most variance and so on. Usually, the first few PCs are used as feature set and other PCs with low variances are discarded. The PCs represent new axes where data direction explains maximum amount of variance. Larger the variance of a line, larger is the scatter of data samples along it and thus more information it carries. The eigen vectors are actually the direction of the axes of most variance (called as PCs), and the eigen values are the coefficients of the eigen vectors, representing magnitude of variance of each PC.

4) Feature vector:

The feature vector has eigen vectors at its columns (the first few ones which are kept). The number of columns in feature vector decides the new dimension of the data.

5) Recasting of data along PC axes:

The input data is still in its original form. To re-orient the input data from original axes to the PC axes, feature vector is used. The transpose of feature vector is multiplied with the transpose of the feature vector to get the transformed data with reduced dimensions having maximum information about input data.

2.2.2. Machine Learning based Modelling:

2.2.2.1. Multivariate regression algorithms:

To interpret spectra in such a way that a desired result is deduced from it, a mathematical function that transforms the predictors (independent variables *x*) and the predicted (dependent variable *y*) is generally needed:

$$y = f(x) \tag{2.9}$$

The coefficients of f(x) are computed from training samples (spectra and reference values both).

2.2.2.1.1. Multiple linear regression (MLR):

The linear regression finds a linear relationship between dependent and independent variable. In case of one independent variable, it is called simple linear regression. While in case of multiple independent variables, it is called as multiple linear regression. Linear regression finds a linear function between the independent and dependent variable, with coefficients computed from the data. Given a dataset of m observations and n number of independent variables [183]:

$$\{y_i, x_{i1}, x_{i2}, \dots, x_{in}\} \; for \; i = 1 \; to \; m \tag{2.10}$$

The MLR estimates a linear function of the form:

$$y_i = \propto_0 + \propto_1 x_{i1} + \propto_2 x_{i2} + \cdots + \propto_n x_{in} + \epsilon_i \tag{2.11}$$

Where y is the estimated value to attribute, β_0 is the y-intercept, β_a (a = 1 to n) are the slope coefficients for each independent variable x_{ia} and x_{ia} are the variables representing absorbance values at multiple wavelengths. Moreover, ϵ is the error variable which defines the random deviations from the linear relationship between the dependent and independent variables. Which can be written in matrix notation as:

$$\boldsymbol{y} = \boldsymbol{X} \propto + \boldsymbol{\epsilon} \tag{2.12}$$

Where

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \tag{2.13}$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1n} \\ 1 & x_{21} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{m1} & \cdots & x_{mn} \end{bmatrix} \tag{2.14}$$

$$\propto = \begin{bmatrix} \propto_0 \\ \propto_1 \\ \vdots \\ \propto_n \end{bmatrix} \tag{2.15}$$

$$\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_m \end{bmatrix} \tag{2.16}$$

Where **y** vector, contains the reference values y_i for i = 1 to m, called as the predicted / dependent variable. **X** matrix contains m observations (along the rows) and each observation having n independent variables in the columns. $\propto$ is an n dimensional regression coefficient vector, $\propto_0$ is the intercept. In simple linear regression, m = 1 and the coefficient is called as slope. These coefficients are computed as partial derivatives of the predicted variable w.r.t independent variables. $\boldsymbol{\epsilon}$ is the vector of error terms or noise. This variable accounts for all other factors that

influence the predicted variable y other than the independent variables. The regression coefficients $\boldsymbol{\alpha}$ can be computed using the ordinary least squares approach assuming **X** is full rank, which gives the unbiased estimator $\widehat{\propto}$ of $\boldsymbol{\alpha}$, of the form:

$$\widehat{\propto} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} \tag{2.17}$$

where the subscript -1 represents the inverse of $\boldsymbol{X}^T\boldsymbol{X}$.

MLR based model is the easiest to interpret considering the wavelengths involved. In MLR, wavelength selection can be manually done, which can lead to more robust model as compared to ones which decide the wavelengths based on training data statistics. However, it requires good background understanding of the underlying phenomenon. In case of spectroscopic data, the number of independent variables 'n' may be larger than the observations 'm' in the training set. Which means that number of columns is greater than number of rows in **X** and thus $\boldsymbol{X}^T\boldsymbol{X}$ (covariance matrix) becomes singular and non-invertible, thus a unique solution is not possible. This can be avoided by increasing the number of observations or reducing the spectral features size such that n < m. Moreover, MLR assumes linear relationship between dependent and independent variables and also that the independent variables don't have high correlation among each other. If the correlation is high, the prediction performance may be poor as high correlation results in over fitting model.

Mainly, MLR outperforms others when the instrument has light source with limited number of wavelengths. For instance, the fruit grader, Sumitomo (Japan), uses an MLR model SSC estimation of fruit as it employs multiple NIR diode lasers of peak wavelengths. The internal defect sorter, MAF (Montabaun, France), employs LEDs of distinct peak wavelengths to estimate ratio of absorbance at two wavelengths, also uses MLR model.

2.2.2.1.2. Principal component regression (PCR):

As the MLR assumes that the independent variables are not correlated with each other (which is not always the case), the PCR improves the modelling of MLR by using PCA and instead of

directly using the independent variables for regression, it uses the PCs (subset of PCs) of the independent variables to estimate the model. Usually, the PCs with higher variance are used as the feature set, however, there may be applications where the PCs with low variance are also important for predicting the output. However, manual selection of PCs which are most influential for the output can lead to excellent prediction results. PCR is applied in three steps [185]:

1) First of all, PCA is applied on the independent variables and PCs are calculated whose number equals the number of independent variables. Then a subset of PCs (the feature set) is chosen based on some appropriate criteria for further use.
2) Next, the reference outcomes are regressed against the chosen feature set using linear regression method to get the regression coefficients vector (dimensions equal to the dimension of the feature set).
3) To get the PCR estimator (having dimension equal to the number of independent variables), the regression coefficients vector is transformed using PCA eigen vectors.

Basically, the PCR approximates the regression coefficients **α** (of MLR) by replacing the **X** matrix with the **S** matrix of PCA

$$\widehat{\propto} = (\boldsymbol{S}^T\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{Y} \tag{2.18}$$

PCR usually has lesser standard errors and give better performance on independent test data as compared to MLR, due to use of PCs. The number of PCs used can be decided by evaluating the PCR model's performance over different number of PCs on a validation dataset that is independent of training set.

2.2.2.1.3. Partial Least Squares Regression (PLSR):

Although PCR provides good improvement in prediction performance as compared to MLR, it uses those PCs which have maximum variance of input data and there is no surety that the PCs with maximum variance will also have influence on the prediction of output variable *y*. If the attribute of interest has significant influence on the spectra, the first few PCs will most likely capture it. However, if the variations in the spectra are influenced by other factors as well, then the

first few PCs may not be most useful for predicting the attribute of interest. To overcome this, the PLSR defines new variables by exploiting the relation between X and Y.

The partial least squares regression is least square regression with partial number of variables. It basically projects the dependent and independent variables into a new space and then finds a linear regression model. The new variables are created as orthogonal linear combinations of input variables by capturing maximum covariance between $\boldsymbol{X}$ and $\boldsymbol{Y}$ [179]. The new variables are different form PCs and are called as latent variables (LVs). In PCA space this is achieved by singular value decomposition of $\boldsymbol{X^T Y}$ instead of $\boldsymbol{X^T X}$**.** As a result, bilinear decomposition of both the matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ is performed in scores matrix and loadings matrix similar to PCA. The decomposition in case of NIPALS algorithm is obtained as [183]:

$$\boldsymbol{X = SL^T + E} \tag{2.19}$$

$$\boldsymbol{Y = RO^T + G} \tag{2.20}$$

Where **X** is an m x n matrix of independent variables (m observations and n features), **Y** is an m x l matrix of output variables, **S** and **R** are m x p matrices (projection of X and Y respectively), **L** and **O** are n x p and l x p orthogonal loading matrices and **E** and **G** are error matrices. PLSR is most suited when the number of features (input variables) is greater than the observations and when there exists multicollinearity between X variables, as opposed to MLR. The number of LVs is decided by testing the PLSR model for different number of LVs on independent test data as in the case of PCR. Both PLSR and PCR perform equally good in most cases. As PLSR captures maximal variance in the attribute of interest i.e. covariance between X and Y rather than maximal variance of spectra i.e. variance in X, it is expected that PLSR gives comparable performance to PCR with fewer number of LVs than that of PCs.

PLSR is the most often used chemometric modelling technique in postharvest research. Commercial postharvest Vis-SWNIRS instrumentation currently available exclusively uses PLSR or MLR methods. Other food applications have seen commercialization of neural networks, indicating that there is room for additional research with other multivariate methods in fruit quality estimation applications, with benchmarking to PLSR using the same data set for calibration and independent test sets. Although there are reports of these modelling techniques being used in postharvest applications [186], no consistent advantage over the usage of PLSR has been proven.

2.2.2.1.4. Artificial neural network (ANN):

The ANNs are inspired by the brain's biological neural networks. An ANN is composed of multiple nodes called as the artificial neurons [185]. Each node is connected to other nodes thus can transmit signal to others just like the brain. At each connection, the signal that is transmitted is a real number and output at each node is calculated by some non-linear function applied at the sum of its inputs. The links connecting different nodes are called edges. Nodes and edges have a weight associated to them that is adjusted as training is done. The weight is multiplied by the signal at each connection. The ANN typically has layers, each layer has one/multiple nodes. The signal travels from input layer to output layer after passing from inner/hidden layers. **Figure *2.3*** below shows the typical structure of ANN.

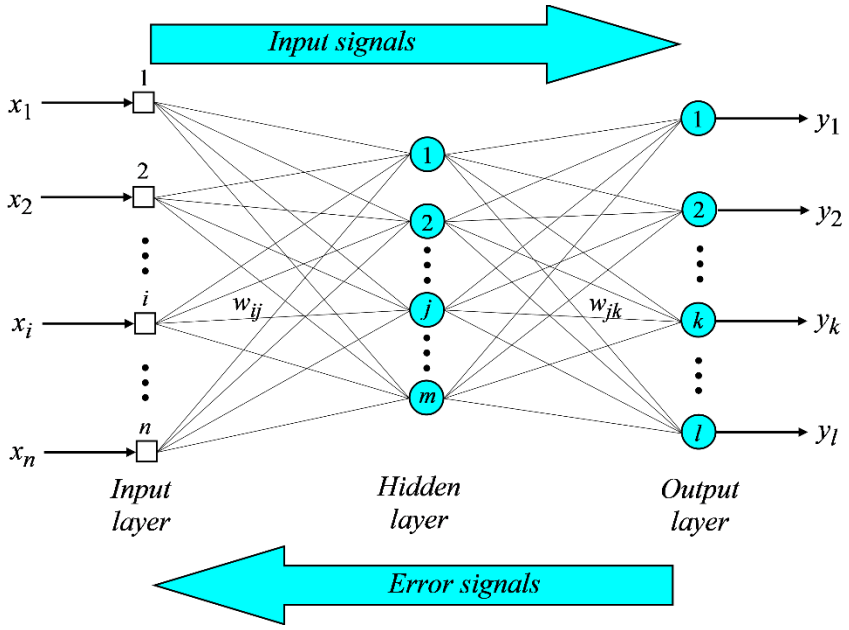


Figure 2.3: Typical structure of ANN

Neural networks are trained just like human brains i.e. by giving them examples, a known input and a known result (without any rules defined). The network is trained by finding the difference between the actual output and predicted output of network and based on the error generated the weights of the network are adjusted. The process is repeated iteratively till the error

is minimal. After sufficient number of iterations a criteria is established and the training is terminated. The ANNs learning resembles the gradient descent approach.

For instance, we have a node with n number of inputs i.e. x_1, x_2, …. , x_n , each input has a respective weight associated with it i.e. w_1, w_2, …. , w_n. the node will sum up all these inputs as:

$$\propto = x_1 w_1 + x_2 w_2 + \cdots + x_n w_n \tag{2.21}$$

Where $\propto$ is the activation or output of the node. After the summation, output is passed through a activation function which includes step, sigmoid, signum or linear function. In the case of step activation function, if the output is greater than a certain threshold the neuron outputs a signal e.g 1 and zero otherwise:

$$x_1 w_1 + x_2 w_2 + \cdots + x_n w_n > T \tag{2.22}$$

Hence, the summation and activation functions together form a node. In 1958, Frank Rosenblatt provided a training algorithm for a simple ANN called as perceptron. A single perceptron can classify inputs into one of two classes. It uses same principle as of a neuron i.e. summation followed by activation function:

$$y = sgn\left(\sum_{i=1}^{2} x_i w_i + \theta\right) \tag{2.23}$$

$$sgn\,(s) = \begin{cases} 1 & if\ s > 0 \\ -1 & otherwise \end{cases} \tag{2.24}$$

Y is the output of perceptron and θ is the bias. For a given training set, both θ and w_i are are updated iteratively. The weights are initialized randomly. μ is the learning rate as in the case of steepest descent algorithm. The algorithm works as follows:

1) Some random sample from training set is taken as input and output is calculated. If output is correct, then nothing is done.
2) If result is incorrect, the weight vector is modified as :

$$w_i = w_i + \mu\, y(n)\, x_i(n) \tag{2.25}$$

$$E = \frac{1}{2}\sum_{m=1}^{N}(y_m - Y_m)^2 \tag{2.26}$$

E is the error function which needs to be minimal. y_m is desired output and Y_m is the predicted output. N is the number of samples in training set. These steps are repeated until the complete training set is classified correctly called as the back-propagation algorithm. The same procedure can be repeated for a multilayer perceptron which one input layer, at least one hidden layer and one output layer. It is to be noted that, ANNs can be used for regression as well as classification. In case of regression the output is computed without the activation function. If large datasets are used, ANNs tend to perform better than PLSR.

2.2.2.2. Classification algorithms:

In machine learning, classification is a two-phase procedure **Figure 2.4**. One is learning phase and other is prediction phase. In learning phase, training data is used to develop a model. In prediction phase, the developed model is used to predict the output for test data[185].

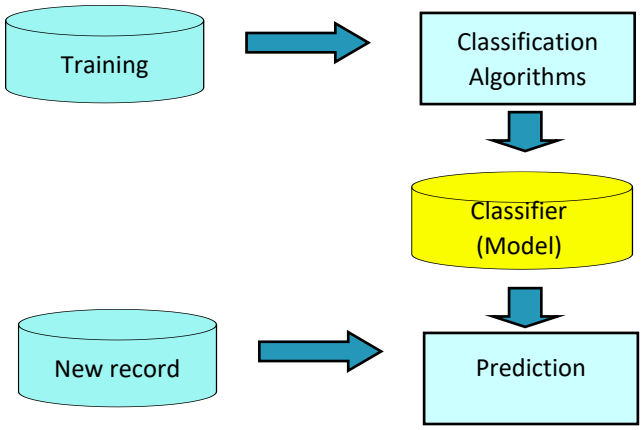


Figure 2.4: Overall framework of classification

2.2.2.2.1. Tree:

Decision trees belong to supervised learning algorithms and are amongst the easiest to understand and implement algorithms [187]. They can be used to solve classification and regression problems both. Decision trees learn decision rules using the training data. Root of the tree is the starting

point to predict a class label (**Figure *2.5***). Root traits are compared with the record's traits. Based on the comparison value, next node to traverse is decided.

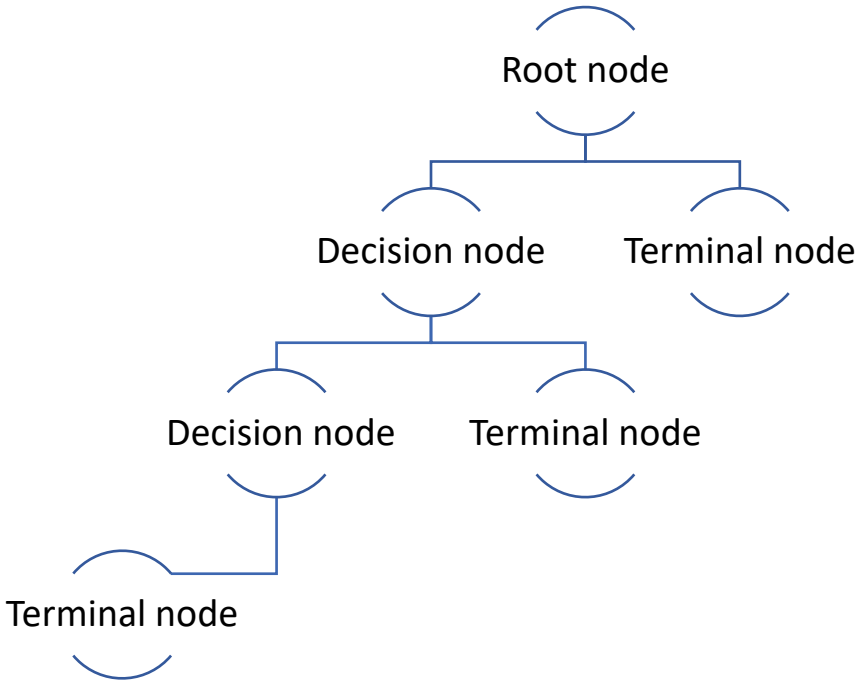


Figure 2.5: Basic layout of decision tree

The basic algorithm includes [188]:

1) Tree is constructed in a divide and conquer manner and top down approach
2) For training at start, all examples are at the root
3) Features are categorical
4) Based on selected features, the examples are divided recursively
5) Test features are selected based on some statistical measure lie the information gain
6) When all samples at a given node belong to same class and there are no remaining features for further division, its time to stop.

2.2.2.2.2. Linear discriminant Analysis (LDA):

LDA is a supervised classification technique used for dimensionality reduction. It overcomes the issues with logistic regression i.e. logistic regression is used for two class classification,

unstable with well-separated classes and if number of observations is lesser than the attributes of interest. The LDA works as follow [189]:

1) Firstly, “between class variance” is calculated which is the distance between the mean of different classes:

$$S_b = \sum_{i=1}^{g} N_i(\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T \quad (2.27)$$

2) Secondly, the “within class variance” is calculated which is the distance between each sample and the mean of that class:

$$S_w = \sum_{i=1}^{g}(N_i - 1)S_i \quad (2.28)$$

3) Finally, the lower dimensional space is created which minimizes the within class variance and maximizes the between class variance. Q is considered as lower dimensional projection also called Fisher’s criterion:

$$Q_{lda} = arg_Q max \frac{|Q^T S_b Q|}{|Q^T S_w Q|} \quad (2.29)$$

2.2.2.2.3. Support vector machine (SVM):

SVM is a supervised machine learning algorithm for classification/regression. It is more suitable for classification but is also used for regression with the name support vector regression (SVR). SVM basically estimates a hyperplane that separates classes of data. To address binary classification issues, SVM employs the "kernel trick," which is based on the addition of one dimension to an existing predictor variable function. When drawing a hyperplane of separation, the points closest to each class' hyperplane are used as support vectors to influence the hyperplane. By default SVM can classify only two classes. However, there exist modifications to tune SVM to classify multiple classes. A binary classifier can be created for each class i.e. if the sample belongs to that class or not of that class. In case of fruits, for instance for apple class, there will be a classifier to predict if it’s an apple or not an apple, for mango class, a classifier will decide if it’s a mango or not and so on.

SVMs perform well for linearly separable data. However, for non-linearly separable data kernelized SVMs are used. The non-linear data in any dimension is mapped to a higher dimension where it is linear. For example, a non linear data in one dimension can be mapped to two dimensions where it will be linearly separable (each 1-D point is mapped to a 2-D ordered pair). The kernel basically tells the similarity between data points w.r.t the two feature spaces i.e. how much two data points of original feature space are similar in the new feature space. Different kernel function exist however, two of are very commonly used [190]:

1) Radial Basis Kernel (RBF):

It is the default kernel which uses an exponentially decaying function to find similarity between two points in the new feature space. Exponential function is applied on the distance between the vectors and the actual input feature space as shown below:

$$K(x, x') = e^{(-\gamma||x-x'||)} \tag{2.30}$$

$||x - x'||$ is the squared Euclidean distance between two points. γ is set manually and should be greater than 0.

2) Polynomial Kernel:

It uses an additional parameter called degree to control the computational cost and complexity.

There are three important parameters in kernel SVM i.e. the Kernel, gamma and the C parameter. Selection of kernel depends upon input data type and the transformation type. Gamma value decides how closely the decision boundary surrounds the data points in original space, low values can lead to low accuracy and high values to over fitting. The C parameter is used to regularize the data. Lower values of C mean more tendency to tolerate errors and hence low accuracy and high value means overfitting.

The SVR is based on the same concepts as SVM, except instead of separating classes, the hyperplane is employed as the regression equation for prediction. In the SVR framework, the epsilon (support vector) is a parameter that specifies the margin of tolerance. The hyperplane is affected by variables in the +/- epsilon region of the hyperplane. A cost parameter is used to weigh the influence of variables outside of this range on the hyperplane's performance, but the hyperplane is still affected by them. LS-SVM (LSSVM) is a variation of the SVM algorithm in which the

hyperplane is solved using a linear equation rather than a quadratic one. There are numerous comparison studies using SVM as a fruit chemometric modelling approach, making it the second most common method.

Pros of SVM include good performance over a range of datasets for both high and low dimensional data and also are versatile in terms of kernel specifications. Cons include more computational complexity, input data needs careful pre-processing, parameter values need to be chosen carefully and are difficult to interpret as they don't provide a direct probability estimator.

2.2.2.2.4. K-Nearest Neighbor (KNN):

KNN is also a supervised learning algorithm and is amongst the simplest ones. It can be used for regression and classification both, however it is mainly used for classification. It is a non-parametric algorithm in a sense that it does not make any assumption about the data. It just stores all the training data examples and when a new sample is given for prediction, it calculates Euclidean distance from all the examples in training data and assign the class that is most similar to the already available classes. This is the reason it is also called as lazy learner algorithm. The KNN classification works as follows [185]:

1) The number K is selected which represents number of neighbors to be checked similarity with.
2) The Euclidean distance of new data point with the K number of neighbors is calculated.
3) Among these neighbors, number of data points in each class is counted.
4) The class which has maximum number of neighbors is assigned to the data point.

In KNN regression, the output value is the average value of K nearest neighbors.

2.2.2.2.5. Ensemble:

Ensemble learning is a general approach that combines prediction from multiple machine learning models to get better results. There are three main categories of ensemble classifiers [185]:

1) Bagging:

   This approach works by applying many decision trees on the same dataset and then averaging the results.

2) Stacking:

   It includes fitting different models on the same dataset and then training another model for best possible combination of the predictions

3) Boosting:

   It involves application of models in a sequential manner which make corrections to the previous model's prediction and then the final prediction is computed as a weighted average.

2.2.3. Benchmarking:

In this field, the aim to develop a chemometric model that relates spectra to attributes of interest [191]. The PLSR is the most widely used chemometric technique used as a regressor in this research area along with standard feature extraction techniques like spectral derivatives, SNV and MSC [191]. If some novel feature extraction technique is reported, benchmarking to a standard and most widely used feature extraction technique is recommended [25] e.g. Sun et al [192] illustrated such benchmarking for temperature correction methods for avocado DM content prediction. The commercially available post-harvest NIRS instruments entirely employ PLSR (or MLR in LED based devices) [25]. Neural networks have also been used commercially for determination of fat products in meat [193]. This suggests that there is capacity for more work using other modelling techniques in postharvest applications however, other techniques should be presented with benchmarking to PLSR on same calibration and independent test datasets [25].

To date, there are numerous publications [25] in this field, however, the comparison of presented models/instruments with other literature reported models/instruments is difficult because of the difference in datasets collected (if available online). In other fields, there are open datasets available online for instance, ImageNet Large Scale Visual Recognition Challenge, ILSVRC, and few open datasets of Vis-NIRS (of a smaller scale) are available online for

applications like soil application [194]. Similar online open access datasets are not available for post-harvest application [25]. Hence comparative studies in this field are not possible.

2.2.4. Calibration set size:

A limited number of observations in the calibration set are used in some NIRS application development reports. Consider the possibility that two random numbers will be arranged in ascending order. Consider what would happen if only three fruit spectra were obtained with dozens of wavelength data points. There's a chance that values at certain wavelengths will trend with attribute level for particular fruits at random. This is an example of regression models that have been overfitted.

It is advised that application studies for assessing intact fruit comprise hundreds of spectra linked to reference values. In order to build a 'robust' model, data from numerous harvests or seasons must be included, resulting in data sets including hundreds of samples (e.g., [48], [152], [195]).

This means that when PLS, MLR, and LS-SVM modelling approaches are compared on the basis of 120 fruits gathered at a single location and time, and the sets of calibration and validation data are divided according to TSS values, the predicted results are overly optimistic for use in independent set prediction, and the proposed solution is not clear whether it will hold up in practice. For an avocado dry matter content application, Blakey [195] gives a valuable case study with over 10,000 samples and the utilisation of independent test sets. As a general rule, cross-validation algorithms assist avoid this problem; however, the cross-validation groups need to be broad and comprise distinct sets of data rather than single sample cross validation.

2.2.5. Independent test sets:

A Vis-NIRS model must be able to withstand a wide range of production, storage, and plant variety circumstances, as well as a wide range of measurement conditions, including ambient

light and temperature. Postharvest industry adoption of the technology requires a demonstration of model robustness, which should be the primary focus of present and future studies. Many reviews have emphasised the importance of using an independent test set, one that has not been used to tune model parameters, to create prediction statistics (e.g., [196]).

A model can be utilised in an NIRS pack-line implementation to predict fruit from various harvest dates, growing sites, and conditions, as well as from different cultivars. As such, it is advised that the test sets chosen reflect the model's intended application. Refractometric assessments of TSS may be affected by changes in the soluble sugar types or levels of organic acids, for example, or the skin characteristics may change with water stress, altering the scattering coefficients of the fruit's skin.

For typical parameters like as TSS and DM, several studies have found that three years of data (growing seasons) is required to develop an accurate model. However, this conclusion may also be linked to the length of a regular funding grant. Samples from a variety of water stress treatments [49] and maturation stages/ripeness stages have been included in the calibration set in an attempt to create robustness within a single season.

Future application development research should cover a wide range of populations. Prediction statistics for these sets, or cross validation findings utilising the independent sets as cross validation groups, should be used to report the outcomes of the study. NIRS model accuracy and precision problems can be teased out by comparing the outcomes of the various test sets (e.g., multiple cultivars, growing conditions, different seasons etc).

2.2.6. Conclusion:

Most of the published research on the measurement of intact fruit internal parameters have used wider wavelength regions including extended NIR region (>1000 nm) [25], e.g. for ‘Valencia’ orange 450-2500nm [197], for citrus 1100-2500nm [198], for ‘Satsuma’ mandarin 400-2350nm [199] and reflection geometry (more details in section 2.1.3). This results in acquisition of spectra having dominant features from fruit skin and hence the models are not robust to skin property variations between populations.

The SWNIR region is used commercially for the assessment of internal quality attributes of intact fruit, in preference to the extended NIR region [25]. Longer wavelength ranges offer narrower and stronger absorption features as compared to SWNIR and thus better evaluation of internal parameters however, the SWNIR wavelengths have greater effective penetration depth into the fruit, hence, offer robustness across independent populations and given the variation in outer layer attributes. The short-wave Vis-NIR option is preferred for commercial purposes due to (currently) lower hardware costs [25], [26]. There is a need to evaluate SWNIRS for fruit quality estimation of Pakistani fruit cultivars.

Moreover, most of the reported literature (section 2.1.3) on non-destructive fruit quality estimation performs quantitative assessment of quality attributes. The potential and benefits of discriminant sorting (based on direct classification) for fruit sweetness classification needs to be analysed as compared to quantitative assessment of quality attribute.

Chapter 3 : SWNIR SPECTROSCOPY FOR FRUIT QUALITY INDEX ESTIMATION FOR PAKISTANI FRUIT CULTIVARS

3.1. Introduction:

The current fruit quality inspection methods in Pakistan are either manual and destructive (based upon physical appearance or internal flesh characteristics) or include weight-based segregation systems. To introduce quality traceable system for each fruit within a pack, efficient non-destructive testing system needs to be adopted by the Pakistani horticulture specially for export quality fruits.

Non-destructive testing using portable near infrared spectroscopy (NIRS) has been used in quantitative measure of fruit quality indices such as DM, SSC, TA , and colour, which are correlated with fruit maturity/quality [2]. Initially, NIRS instruments were laboratory bound, however, portable instrumentation is now available, e.g, Consumer Physics' SciO [200]. A few instruments are designed specifically for use with fruit, e.g., in terms of a partial transmission optical geometry, e.g., Sunforest H-100C [201] and Atago Hikari [202], and ability to operate outdoors with varying ambient temperature and light levels, e.g., Felix Instrument F-750 produce quality meter [159]. The F-750 provides referencing on every sample to address the issues of varying ambient light and temperature. Some spectrometers targeted to fruit applications come with pre-supplied models that the user can not alter (e.g., the Hikari). Other units come with a factory supplied model but allow the user to extend the existing model or develop new models. The F750 unit is supplied with factory installed 'starter' models which are intended for use by the user-horticulturalist. Starter models are available for apple, mango, cherry, grape, avocado, kiwi, mandarin, pear, cherry and persimmon fruit [159]. For ease of practical implementation, a Vis-NIRS model should be 'robust' in use across production environments, seasons, storage conditions, plant varieties, and instruments. Managing multiple models is time-consuming and cumbersome for end-users [177].

To investigate the applicability of SWNIR spectroscopy for non-destructive quality inspection in Pakistan, the industry standard F-750 fruit quality meter is used as a Vis-NIR spectrometer handheld device. This device is featured in many publications [55], [80], [162], [163], [170], [203], [204], all of which involve de-novo development of chemometric models, but

the robustness of the available chemometric models has not been demonstrated in prediction of fruit populations from a range of varieties and growing conditions. To directly adopt available portable NIRS instruments in Pakistani horticulture, the effectiveness of factory supplied models in prediction of Pakistani fruit cultivars needs to be investigated. Therefore, as a first thing, I investigated the robustness of available chemometric models for quality inspection of Pakistani fruit cultivars. And for fruits where available chemometric models failed to predict the quality indices of Pakistani cultivars, I have developed new chemometric models for prediction of fruit quality parameters.

In this chapter, factory supplied starter models for apple, mango and mandarin are assessed relative to locally developed new models. The F-750 Vis-NIR factory-supplied apple, 'Kensington Pride' (KP) mango and mandarin models were evaluated to estimate SSC and DM for local cultivars of apple varieties (i.e., Golden Delicious and Red Delicious), export quality cultivars of mango (i.e., 'Sindhri', 'Samar Bahisht (SB) Chausna' and 'Sufaid Chaunsa') and SSC for an export quality cultivar of mandarin (i.e., Kinnow). Moreover, new models were developed for Pakistani cultivars of mango and mandarin and compared with factory supplied models. Starter model for loquat was not available hence, the applicability of field portable NIRS to the loquat is also investigated. These fruits hold high commercial significance to Pakistan, but also provide examples of fruit with relatively thin and thick skin and with relatively thick and thin edible flesh. These differences in morphology can be expected to impact the non-invasive assessment of flesh characters using NIRS.

3.2. Materials and methods:

3.2.1. Fruit:

Mature, ripened apple (*Malus domestica*) samples were purchased from the local market (120 fruit in total, 30 fruit of each of two varieties, 'Golden Delicious' and 'Red Delicious', on two dates during the month of May 2020). The recommended harvest maturity criterion in Pakistan for apple i.e. 'Red Delicious' cultivar is 12.03 ºBrix and 15.1% DM [205], 'Golden Delicious' is 11.68°Brix and 17.03% DM [206].

Mango (*Mangifera indica L.*) fruit at hard green harvest-maturity stage were harvested from mango orchards located in Muzaffargarh and Multan Districts of Punjab Province, Pakistan. A total of 734 fruit samples of variety 'Sindhri', 'SB Chaunsa' and 'Sufaid Chaunsa' were harvested across two seasons (2019, 2020), with fruit harvested at three stages, i.e., one week before, on and one week after the estimated commercial harvesting date (mango harvest period for Sindhri and Samar Bahisht Chaunsa is 2-3 weeks and for Sufaid chaunsa it is 3-4 weeks). These populations provided a wide range of SSC and DM values. The 600 samples (200 of each variety) harvested in August 2019 were used for PLSR model development. The134 samples (Sindhri 20 samples, SB Chaunsa 73 samples and Sufaid Chaunsa 41 samples) harvested in July 2020 were used for model validation. The recommended harvest maturity criterion in Pakistan for mango i.e. 'Sindhri' cultivar is 6-7.5 ^{o}Brix and 17-20% DM, 'Samar Bahisht (SB) Chaunsa' is 9.0-11.0°Brix and 18-21% DM, and 'Sufaid Chaunsa' is 8-9 °Brix and 24-25% DM ([207], [208]).

Mandarin (Citrus nobilis Lour x Citrus deliciosa Tenora , variety 'Kinnow') ripened samples were purchased from a local market (72 fruit in total, on two dates during the month of Feb 2021). Average peel thickness was 2 - 4 mm. 50 samples were used for model calibration, each sample was scanned from two opposite sides for SSC hence total 100 spectra were used to calibrate model. 22 samples, scanned from two opposite sides for SSC (total 44 spectra), were used for model validation. The recommended harvest maturity criteria for mandarin is 10 ^{o}Brix.

Loquat (*Eriobotrya japonica,* variety 'Tanaka') fruit were harvested during the month of May 2021 from an orchard located in district Garhi Duppata of Azad Jammu Kashmir (AJK) province, Pakistan. Average fruit size of investigated variety is 2.5-5cm with a thin skin and large central seeds. Mature loquat fruit (n=225) were harvested randomly from different locations on different trees, on three dates. Fruit samples from the first two harvest dates (n=157) were used in model development, while fruit samples from the third harvest (n=68) were used in validation. The recommended harvest maturity criteria for loquat (Eriobotrya japonica) is 12 ^{o}Brix [209].

All fruits were transported on the day of collection to a local laboratory and stored at room temperature (25^{o}C) for 24 h prior to analyses to minimize the influence of sample temperature on prediction accuracy [210]. Mangoes were tested in Muhammad Nawaz Shareef University of Agriculture (MNSUA), Multan, and other fruits were tested in National Centre of Robotics and Automation (NCRA), Islamabad.

3.2.2. Collection of Vis-NIR spectra:

Apple, mango, mandarin and loquat fruit were marked on opposite sides (see **Figure *3.1***) at equatorial position and three replicate spectra collected at each position. The three spectra for each position were averaged to provide one spectrum for each position. Vis-NIR spectra (range 400-1150 nm) were collected using the F-750 (Felix Instruments, Camas, WA, USA). This device employs interactance optical geometry and a Carl Zeiss MMS-1 spectrometer, with a pixel spacing of approximately 3.3 nm and a spectral resolution (FWHM) of 8-13 nm. It uses a halogen lamp as a light source.

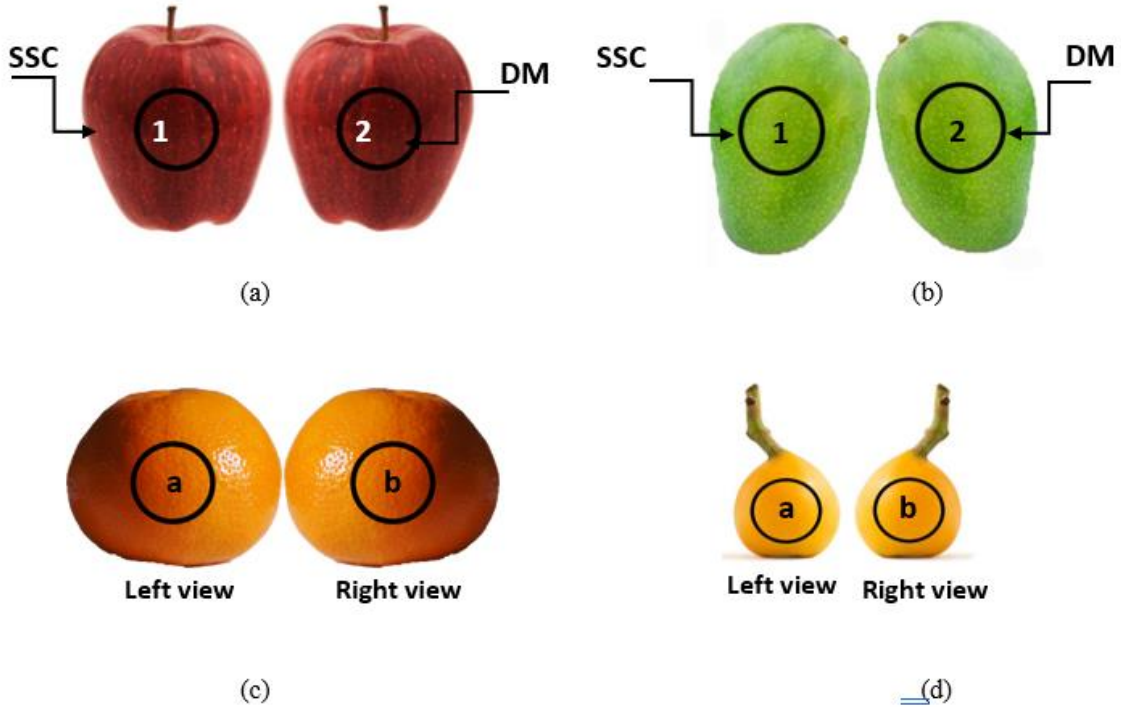


Figure 3.1: Schematic diagram of the marked positions for NIR spectra collection in (a) apples, (b) mangoes, (c) oranges and (d) loquats

3.2.3. Reference measurements:

With apple and mango fruit, one sampling position was used for SSC measurement and other used for dry matter content (DM) estimation using reference methods. For orange and loquat, SSC was assessed of both positions on each fruit. For SSC estimation, the marked region was excised, and skin was removed. The extracted flesh was squeezed using a manual fruit squeezer. SSC was assessed of a sample of the extracted juice using a digital refractometer (Model: PAL-1 [°Brix 0–53%], Atago Co., Ltd, Tokyo, Japan). The refractometer has automatic temperature compensation with range 10-100^{o}C and measurement accuracy of ±0.2%.

For apple and mango DM determination, a core of 26 mm in diameter and 10 mm in depth (Fig. 3.1 a and b) was extracted and the skin (1–2 mm thick) removed using a fruit peeler. Weight was assessed before and after drying in a fan forced oven set at 65°C to constant weight (generally 48 h) [147].

3.3. Chemometric analysis:

The F-750 model builder software was used for building calibration models using calibration datasets **Table *3.2*** for mango, mandarin and loquat fruit. Three models for each fruit type were developed, based on use of raw spectra and pre-processed spectra using Savitzky-Golay first and second derivative smoothing filters. The performance of developed models was evaluated by R^2_{CV} (coefficient of determination of cross validation), R^2_P (coefficient of determination of prediction), $RMSE_{CV}$ (root mean square error of cross validation) and $RMSE_P$ (root mean square error of prediction). This software employs leave one out cross validation. Partial least squares regression (PLSR) models for mango and loquat were developed for SSC and DM using the 729-975 nm region as this is a region containing information on O-H and C-H features of carbohydrates and water [211]. For oranges, the PLSR model was developed using the Vis/NIR region in the range 600-1050 nm (following [212]).

The factory supplied model for apple SSC and DM was indicated to be based on >1500 samples from multiple populations of Granny Smith, Gala, Fuji, Jazz, red Delicious, JonaGold, Golden Delicious and other varieties, with fruit at a range of maturities and at multiple temperatures. The

DM employed 6 latent variables while the SSC model employed 5 latent variables and both models used the 725-975nm range. The supplied mango model of F-750 is a post-harvest model for SSC and DM estimation developed using 196 KP mango samples, used the 801-975nm range and employed 5 latent variables for SSC and DM both. The supplied Mandarin starter model is a SSC demo model developed at room temperature using 78 store bought Mandarin samples of unstated cultivar and used the wavelength region from 702-981nm, employing 6 latent variables.

3.4. Results:

3.4.1. SSC and DM statistics:

The viability of NIRS assessment of a given attribute is dependent on the $RMSE_P$ and the standard deviation (S.D.) of the attribute in the population. This technique is useful only when the S.D. of the attribute of interest is greater than the $RMSE_P$ [25]. **Table *3.1*** lists the SSC and DM statistics of collected datasets from 180 apple, 734 mango, 144 orange and 450 loquat spectra samples. Standard deviation (SD) on SSC measurements was 1.2 ^oBrix in apple, 1.8 ^oBrix in mango, 1.07 ^oBrix in mandarin and 2.88 ^oBrix in loquat, while SD on DM measurements was 1.1 % for apple and 2.5 % for mango. Given expected $RMSE_P$ of under 1 ^oBrix and 1% DM [6], these populations show a sufficiently large distribution range to suit evaluation by NIRS.

Table *3.2* shows statistics for DM of mango with respect to DM model calibration and prediction data sets and SSC statistics of mango, melon, and loquat fruit with respect to the Brix calibration and prediction data sets.

The prediction set SSC and DM distributions were similar to the calibration set for mango and mandarin, while the loquat prediction set had a lower SSC than the calibration set, the distributions were still well overlapped **Table *3.2***.

Table 3.1: Statistics of reference SSC and DM values for apple, mango, mandarin and loquat data sets obtained through destructive testing

Fruits	Variety	No. of Samples	Minimum		Maximum		Mean		S.D.	
			(ºBrix)	(DM%)	(ºBrix)	(DM%)	(ºBrix)	(DM%)	(ºBrix)	(DM%)
Apple	Golden Delicious	60	11.0	13.5	14.7	17.4	12.8	15.1	1.1	1.1
	Red Delicious	60	9.7	12.5	13.5	16.4	11.9	14.6	1.0	1.0
	Combined	120	9.7	12.5	14.7	17.4	11.9	14.6	1.2	1.1
Mango	SB Chaunsa	273	9.4	15.1	19.8	33.7	13.5	23.5	2.99	2.9
	Sufaid Chaunsa	241	5.3	19.1	12.0	30.4	8.5	25.2	1.2	2.1
	Sindhri	220	5.4	11.4	11.2	22.9	8.5	18.2	1.0	2.2
	Combined	734	5.3	11.4	19.8	33.7	10.2	22.3	1.8	2.5
Mandarin	Kinnow	144	8.2	--	13	--	10.35	--	1.07	--
Loquat	Tanaka	450	4.9	--	25.7	--	13.87	--	2.88	--

Table 3.2: Statistics of reference SSC and DM values with respect to calibration and prediction data sets for mango, mandarin, and loquat fruit, obtained through destructive testing.

Fruits	Data set	No. of spectra	Minimum		Maximum		Mean		S.D.	
			ºBrix	DM %	ºBrix	DM %	ºBrix	DM %	ºBrix	DM %
Mango	Calibration	600	5.3	11.4	19.8	32.5	10.0	22.0	1.2	2.2
	Prediction	134	6.7	11.7	19.7	33.7	10.2	22.1	1.0	2.0
Mandarin	Calibration	100	8.4	--	13	--	10.45	--	1.1	--
	Prediction	44	8.2	--	12.2	--	10.13	--	0.98	--
Loquat	Calibration	316	4.9	--	25.7	--	14.57	--	2.89	--
	Prediction	134	8.2	--	24.4	--	13.53	--	2.80	--

3.4.2. Overview on spectra:

The average absorbance spectra (**Figure *3.2***) of apple, mango, orange and loquat fruits are dominated by a peak around 680 nm associated to chlorophyll absorption and a 970 nm peak associated with a second overtone of a O-H stretching vibration [25]. **Figure *3.3*** shows the average absorbance spectra for each investigated variety of (a) apple and (b) mango fruit.

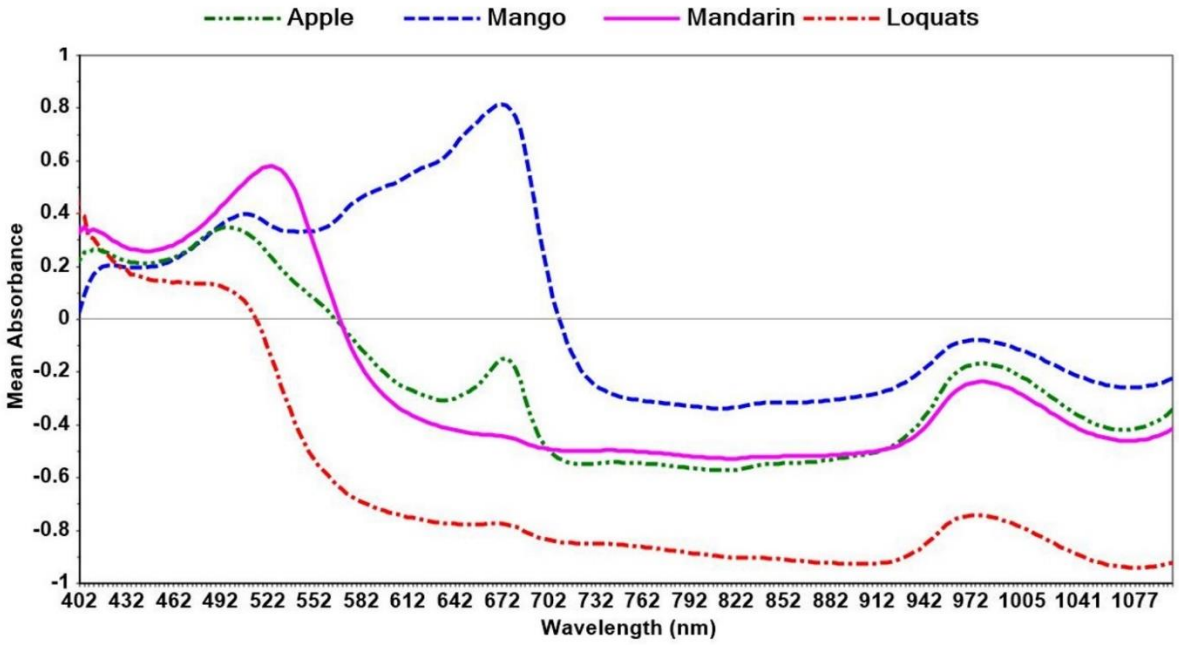
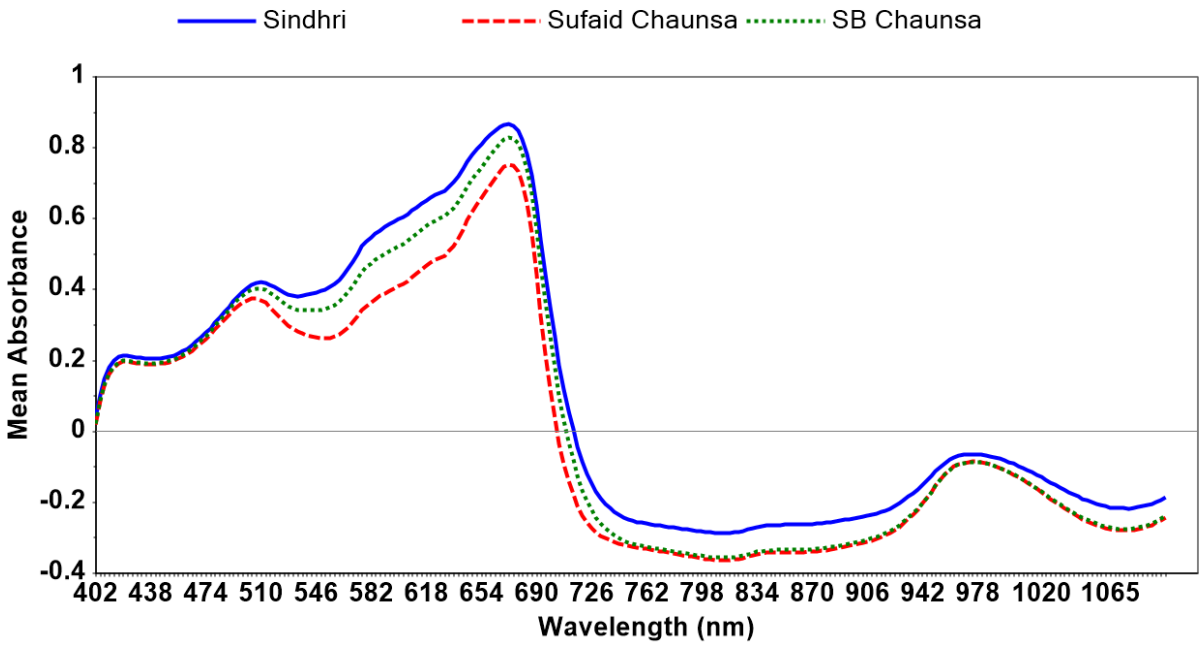
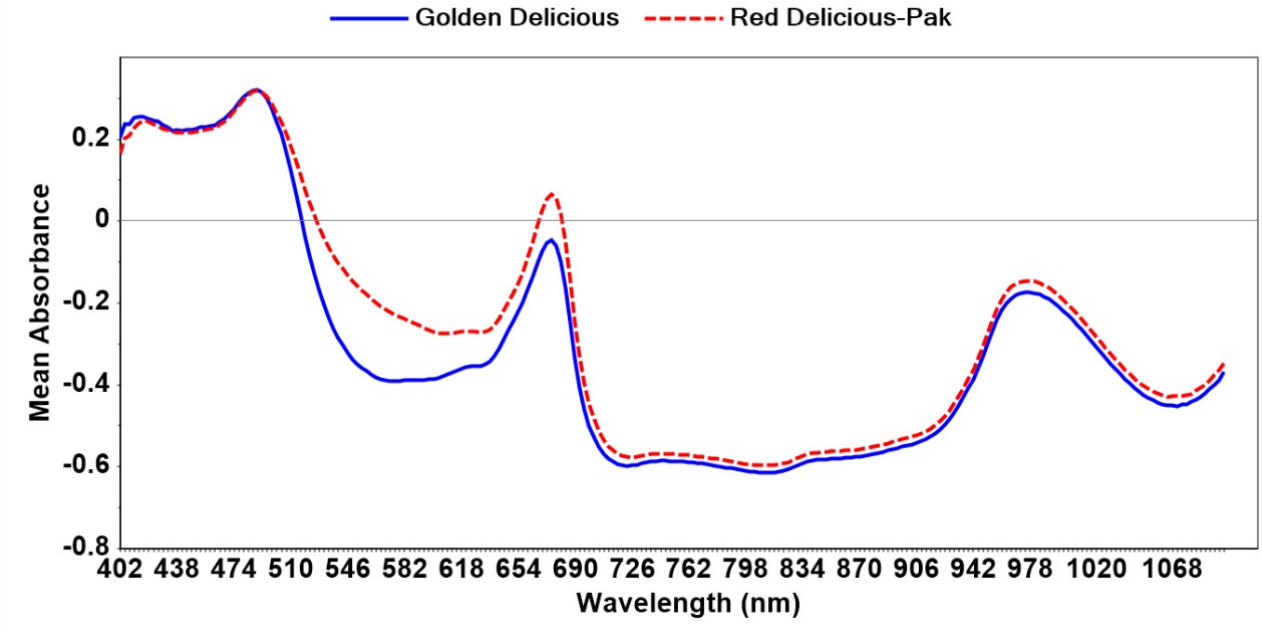


Figure 3.2: Average raw absorbance spectra of all investigated varieties for apple, mango, mandarin and loquat fruit samples.



(a)



(b)

Figure 3.3: Raw absorbance spectra depicting difference in average spectra of investigated varieties of (a) apple and (b) mango samples.

3.4.3. PLSR model results:

3.4.3.1. Apple:

F-750 apple starter PLSR model was validated using 120 samples of 2 varieties .e. 'Golden Delicious' and 'Red Delicious' of apple fruit. **Table *3.3*** shows the combined results of F-750 apple model in terms of R^2 and RMSE for Brix and DM. The values of R^2 (0.90,0.87) and RMSE (0.76 ^{o}Brix, 0.69%) show that the F-750 apple model is a robust model and well suitable for the two investigated varieties i.e., 'Golden Delicious' and 'Red Delicious' cultivated in Pakistan.

Table 3.3: F-750 apple model validation results with two varieties i.e. 'Golden Delicious' and 'Red Delicious' for Brix and DM

Maturity Index	R^2	RMSE
SSC (^{o}Brix)	0.90	0.76
DM (%)	0.87	0.69

3.4.3.2. Mango:

Mango SSC and DM prediction models were built for three local varieties of mango i.e. 'Sindhri', 'SB Chaunsa' and 'Sufaid Chaunsa' using F-750 produce quality meter. Combined variety models were built using PLSR. Three models were built for SSC and DM i.e. with raw spectra, with SG first derivative preprocessed spectra and SG second derivative preprocessed spectra. All models were calibrated using leave one out cross validation. **Table *3.4*** presents cross validation results for combined variety mango model built using 600 mango samples (from 2019 season) in terms R^2 of cross validation (R^2_{CV}) and RMSE of cross validation ($RMSE_{CV}$). From **Table *3.4***, it can be concluded that the PLSR model built with raw spectra for DM prediction and SG second derivative pre-processed spectra for SSC prediction performed comparatively better than the other models with R^2_{CV} of (0.80, 0.88) and $RMSE_{CV}$ of (1.35^{o}, 1.28%) for SSC and DM, respectively. Hence, the model no 1 was selected for DM prediction, model no 3 for SSC prediction and built into the F-750 produce quality meter for further validation.

Table *3.5* depicts the R^2 prediction (R^2_P) and RMSE prediction ($RMSE_P$) values using independent prediction dataset of 2020 season for SSC and DM prediction, respectively. **Table *3.5*** shows the SSC and DM prediction results for each variety separately (Sindhri 20 samples, SB

Chaunsa 73 samples and Sufaid Chaunsa 41 samples) when tested with the developed combined variety PLSR model, as well as prediction results using 134 samples collectively. The prediction results were first collected on-tree using the developed model and then the samples were harvested and immediately transported to laboratory, where the samples were stored on room temperature for 1 hour and then results were collected again using the developed model as well as supplied model. After taking readings from the developed model and supplied mango model, the mango samples were destructively tested to find reference values of SSC and DM. Lab prediction results of developed combined variety model have $RMSE_P$ of 0.95 ^oBrix and 1.17% while on tree results have $RMSE_P$ of 1.25 ^oBrix and 1.24% for Brix and DM respectively. Results of supplied model for the same prediction data set are also reported in **Table *3.5*** which show that DM results of KP model are better than SSC results.

Prediction results over single cultivar and combined cultivars were similar for locally developed combined cultivar model, and superior to the factory supplied model (**Table *3.5***). Results for predictions made outdoors, of fruit on tree were slightly poorer than predictions made of the same fruit under indoor conditions of constant lighting and temperature.

Table 3.4: Cross validation results for SSC and DM prediction models using PLSR for 600 mango samples of 3 cultivars

Model No	Pre-processing	SSC model		DM model	
		Cross validation		Cross Validation	
		R^2_{cv}	$RMSE_{cv}$ (^oBrix)	R^2_{CV}	$RMSE_{CV}$ (%)
1	**None**	0.77	1.38	0.88	1.28
2	**SG 1st Derivative**	0.79	1.35	0.88	1.32
3	**SG 2nd Derivative**	0.80	1.35	0.88	1.31

Table 3.5: Prediction results of SSC and DM with supplied model and developed combined variety model for local varieties

Maturity Index	Prediction data set											
	Developed combined variety model										Factory Model	
	Sindhri samples		SB Chaunsa samples		Sufaid Chaunsa samples		All samples (Prediction in lab)		All samples (Prediction on tree)		All samples (prediction in lab)	
	R^2_P	$RMSE_P$	R^2_P	$RMSE_P$	R^2_P	$RMSE_P$	R^2_P	$RMSE_P$	R^2_P	$RMSE_P$	R^2_P	$RMSE_P$
SSC (°Brix)	0.92	1.15	0.71	1.26	0.93	0.85	0.90	0.95	0.83	1.25	0.38	9.23
DM (%)	0.89	1.62	0.68	1.05	0.74	0.85	0.80	1.17	0.77	1.24	0.70	4.03

3.4.3.3. Mandarin:

Mandarin SSC prediction model was developed using PLSR for Pakistani export quality cultivar i.e. 'Kinnow'. For model calibration, leave-one out cross validation was used on data collected from 100 Kinnow samples. Three models were developed, one with raw spectra and other two with SG first and second derivative pre-processed spectra, respectively. Second derivative pre-processed spectra-based model performed comparatively better with R^2_{CV} of 0.68 and $RMSE_{CV}$ of 0.62^o Brix.

F-750 'Mandarin' SSC starter model and the developed 'Kinnow' model were validated using 44 'Kinnow' samples. Prediction results in **Table *3.6*** show that the supplied model needs re-calibration for the investigated local variety. The locally developed model outperformed the factory supplied model in prediction of a test set with R^2_P of 0.71 and $RMSE_P$ of 0.64 ^{o}Brix.

Table 3.6: Cross validation and prediction results for developed SSC model for 'Kinnow' and prediction results of supplied Mandarin model on same test data

Pre-processing	Developed model				Factory Model	
	Cross validation		Prediction		Prediction	
	R^2_{cv}	$RMSE_{cv}$ (^{o}Brix)	R^2_P	$RMSE_P$ (^{o}Brix)	R^2_P	$RMSE_P$ (^{o}Brix)
None	**0.53**	**0.75**	**-**	**-**	**-**	**-**
SG 1st Derivative	**0.61**	**0.68**	**-**	**-**	**-**	**-**
SG 2nd Derivative	**0.68**	**0.62**	**0.71**	**0.64**	**0.03**	**7.47**

3.4.3.4. Loquat:

Loquats fall under the class of small size (2.5 – 5 cm) and thin skin fruits. The F-750 produce quality meter does not have a factory supplied 'starter' model to predict maturity indices of loquats. To build local calibration model, fruit samples were scanned on their equator region, with around 80% of scanned side illuminated. Three PLSR models were developed, one with raw spectra and other two with SG first and second derivative pre-processed spectra, respectively. Model built using second derivative preprocessing again delivered the best cross validation results

and was further validated with an independent prediction set **Table *3.7*** . Validation results depicted R^2_P 0.90 and $RMSE_P$ 0.95^o Brix.

Table 3.7: Cross validation and prediction results for Brix prediction model using PLSR for loquat fruit

Pre-processing	Cross validation		Prediction	
	R^2_{cv}	$RMSE_{cv}$ (^{o}Brix)	R^2_P	$RMSE_P$ (^{o}Brix)
None	0.85	1.04	-	-
SG 1st Derivative	0.88	0.94	-	-
SG 2nd Derivative	0.89	0.92	0.90	0.95

3.5. Discussion:

3.5.1. Comparison of Vis-NIR spectra:

The fruit types varied in chlorophyll and carotenoid content, as revealed by the size of the absorbance peaks centered on 680 and 550 nm, respectively (**Figure *3.2***). The water absorption peaks in the range 700-1100 nm have lower absorptivity than experienced at higher wavelengths which allows for a longer pathlength of these shorter wavelengths. This is beneficial for estimation of mesocarp DM and/or SSC content of intact fruits. With interactance optical geometry, as employed by the F-750, the effective optical sampling depth is reported to be 10-30 mm for thin skinned fruits [25]. The average NIR absorbance was greater in mango and apple than mandarin and loquat fruits, a result ascribed to a higher light scattering coefficient of skin of mandarin (average rind thickness in the investigated variety was 2-4 mm) relative to apple and mango. It is not clear why the absorbance values of loquats were also lower than that of apple and mango. The lack of impact of the seed on spectra and models suggests that the majority of detected diffusely scattered light originated within the fruit mesocarp (average mesocarp thickness in the investigated variety was <10 mm).

3.5.2. Performance of factory supplied apple, mango and orange models of F-750:

The supplied apple starter model of F-750 was built with an extensive dataset of more than 1500 apples of Granny Smith, Gala, Fuji, Jazz, Red delicious, JonaGold, Golden delicious,

and other cultivars over a wide maturity range and multiple temperatures to estimate SSC and DM. When tested for Red delicious and Golden delicious varieties cultivated in different geological region than those of used in model calibration), the model returned good estimation results with R^2 (0.90,0.87) and RMSE (0.76 ^{o}Brix, 0.69%) (Table 3.3). The reason for such results is, firstly the model was already calibrated for the tested varieties i.e. Red delicious and Golden Delicious; secondly the training dataset was quite extensive with over 1500 apples covering a wide maturity range and thirdly the model was trained over multiple temperatures to reduce the effect of temperature variations while taking readings.

I developed a new model for the three export varieties of Pakistani mango, keeping in view their economic importance. The combined variety model built using 600 samples from 2019 season has R^2_{CV} of (0.80, 0.88) and $RMSE_{CV}$ of (1.35 ^{o}Brix, 1.28%) for SSC and DM, respectively (Table 3.4). Moreover, for SSC estimation, model developed using second derivative of spectra is the best, while for DM estimation, model developed using raw spectra is optimum. The supplied mango model of F-750 is a post-harvest model for SSC and DM estimation developed using 196 KP mango samples. For fair comparison of our developed model and supplied KP model, KP model was validated using the same prediction data set of season 2020 (**Table *3.5***). Results show that DM estimation of KP model is better than SSC estimation for cultivars not included in model calibration. For mangoes, it is observed (see **Table *3.4***) that a single model is not optimum to estimate both SSC and DM.

Supplied Mandarin starter model is a SSC demo model developed at room temperature using 78 store bought Mandarin samples. When validated with a similar variety (i.e., 'Kinnow') in terms of peel thickness (avg thickness 3 mm), the supplied model returned prediction R^2_P of 0.04 and $RMSE_P$ 7.47^{o}Brix (Table 3.6) which shows that Mandarin spectral data is not correlated with Kinnow data. The mandarin model is built upon the spectral range 702-981 nm however, I observed that the developed model of Kinnow performed better for 600-1050 nm range, which is in agreement with the findings of [212].

3.5.3. Comparative performance by fruit type and application to loquat:

A $RMSE_P$ of 0.76, 0.95, 0.64 and 0.95 ^{o}Brix (Table 3.3 – 3.7) was obtained in prediction of SSC for test sets of apple, mango, mandarin and loquat fruit, respectively. And for DM, a

$RMSE_P$ of 0.69% and 1.17% in prediction of test sets of apple and mango fruit, respectively. This result is acceptable for quality classification of these products, given an attribute range (S.D.) of at least twice that of the $RMSE_P$ values [25].

The utility of NIRS assessment (725-975 nm) of SSC in intact loquat was demonstrated, despite the presence of a large seed within a fruit with relatively thin flesh. The $RMSE_P$ value of 0.95° Brix (Table 3.7) for prediction data set agrees with Fu et al. [23] who reported values of 0.96 - 1.21° Brix ($RMSE_P$) for Dahongpao and Jiajiaozhong varieties of loquats with Fourier transform NIR (FT-NIR) spectrometer in diffuse reflectance mode. It is to be noted here that the $RMSE_P$ values reported by Fu et al. [23] used full range reflectance NIR spectra (800-2500 nm) and hence a greater number of spectral features. While, the locally developed model uses 729-975 nm NIR interactance spectra (only 83 spectral features) to build SSC prediction models and achieved $RMSE_P$ of 0.95° Brix.

3.6. Conclusion:

In this chapter, I investigated the effectiveness of handheld near infrared spectroscopy (NIRS) for the evaluation of dry matter (DM) and soluble solids content (SSC) in four types of fruits of significance to Pakistani horticulture, i.e. mango, apple, mandarin and loquat. Lesser NIR absorbance is observed in mandarin and loquat relative to apple and mango fruit, due to higher light scattering coefficient of mandarin's peel (average peel thickness 2-4 mm) and loquat's edible mesocarp (average mesocarp thickness <10mm). The investigated handheld NIRS device (i.e. Felix instruments F750) is supplied with ready to use starter models for many fruits. The starter models (i.e. mango and mandarin) which were based on a small population set of a single cultivar at a single temperature failed to effectively predict the quality indices of local cultivars of mango and mandarin. Conversely, the supplied apple model which was based on a large population set of multiple cultivars at different temperatures performed well in use with locally grown fruit of the same cultivars as used in the calibration dataset. The locally developed mango model was based on samples at a single temperature hence when tested with on-tree fruit, the performance was inferior as compared to that of in-lab testing at single temperature. Based on these results, the portable NIR technology is recommended for use across several commodities, given suitable model development based on samples from multiple cultivars with a range of growing conditions,

and with assessment of samples at different temperatures. Use in quantitative assessment of fruit quality targeted to Pakistan's export markets is also recommended. In the next chapter, a qualitative classification-based decision support system is presented for melon's eating quality classification.

Chapter 4 : DECISION SUPPORT SYSTEM FOR MELON SWEETNESS CLASSIFICATION USING DIRECT CLASSIFICATION APPROACH

4.1. Introduction:

Melons (*Cucumis Melo*) are nutritious, sweet, and amongst the most refreshing summer fruit in Pakistan. Honey melons are cultivated in Sindh, Punjab, and some parts of Kyber Pakhtun Khaw province of Pakistan, harvested from April till June. These are not the same as honey dew melons although they have little resemblance in appearance. Honey melons have creamy net patterns on their skin with a strong fragrant aroma. Melons come under the class of non-climacteric fruits i.e. once harvested they cannot ripe on their own. Hence, for melons maturity (at the time of harvest) and ripeness are the same. The term ripeness is directly related to the eating quality of fruits. Unlike mangoes, and apples it is difficult to know if the melon is sweet just by smelling or looking the fruit from outside. This is usually needed when the melons are graded for packing to the market, or at the consumer end. For now, there exists some melon varieties that are guaranteed to be sweet however for varieties such as "Honey melons", it is difficult to judge their sweetness from smell, skin or firmness.

For convenience of consumers, there is a need to introduce marketing standards that should help differentiate between sweet and poor-quality fruit. Soluble solids content (SSC) has been widely used as melon quality assessment parameter [213]. SSC represents sugars (such as sucrose, fructose etc.), organic acids (such as citric, malic, tartaric etc.), soluble amino acids and other compounds (such as fat, minerals, alcohol, flavonoids etc.). The USA industry has used SSC as quality index for many years for marketing purposes [214]. Moreover, in the absence of sensory analyses, breeders and agronomists depend heavily on SSC as a standard for quality. Chace et. al. [215] first proposed SSC as suitable objective parameter of quality followed by researchers in [216] which affirm that SSC and quality are linearly related. Mutton et. al. [213] suggested that SSC and flesh firmness both are key parameters for melon quality assessment however, flesh firmness changes rapidly during storage thus limiting its use. For Australian market, Mutton et. al. recommended 10^{o} Brix as minimum acceptable quality standard [213], while for USA, it is 9^{o} Brix [214].

Existing conventional methods used to measure internal characteristics of melons are destructive and time consuming [217]. Section 2.1.3.4 of this dissertation presents detailed literature review of the NIRS based non-destructive experiments carried out in literature.

To the best of our knowledge, all efforts done using vis/NIR spectroscopy to define quality standards for melons are based on prediction of SSC by some regression model, and then defining sharp limits on the predicted SSC values to distinguish satisfactory or unsatisfactory class melons as done by researchers in literature [131]. However, there may exist some overlap in SSC values of both classes (satisfactory and unsatisfactory) as SSC does not merely represent sugar content but multiple soluble contents such as sugars, organic acids, soluble amino acids, and other compounds. However, I argue that consumer acceptance is more related to the sweetness level of melons i.e. sugar content than other soluble compounds [37].

In this chapter I define acceptance criteria for melons based on direct classification method to predict the sweetness level using NIR spectroscopy. I have compared performance of our classification-based approach with that of regression-based thresholding methods reported in literature.

4.2. Materials and methods:

4.2.1. Melon samples preparation:

For the experiment, a total of 101 honey melon samples were purchased from local market in five different batches (20 melons each) covering one full season of melons i.e. on 17 April, 1 May, 15 May, 29 May and 12 June 2020. All samples were elliptic and individual fruit weight was around 0.5-1.5 kg. Average rind thickness was 6.68 mm. All samples were transported to a local laboratory (Islamabad, Pakistan) and stored at room temperature (25^{o}C) for 24h to minimize the influence of sample temperature on prediction accuracy [210]. All samples were then marked at 4 equator positions each (almost 90^{o} apart) to cater for within same sample SSC and sweetness variations, for spectral acquisition and destructive testing as shown in **Figure *4.1***. Samples within each fruit were treated as separate spectral sets. Hence, 404 samples were divided into calibration set and prediction set with a of ratio 3:1. To secure better performance of calibration models, the distribution range of SSC values in calibration set was usually much larger than that of prediction

set [46]. For that, the samples were sorted in ascending order of actual SSC values and the second sample of every four samples were added into prediction set. Remaining samples were added to calibration set. This sample partitioning method is reported in [129].

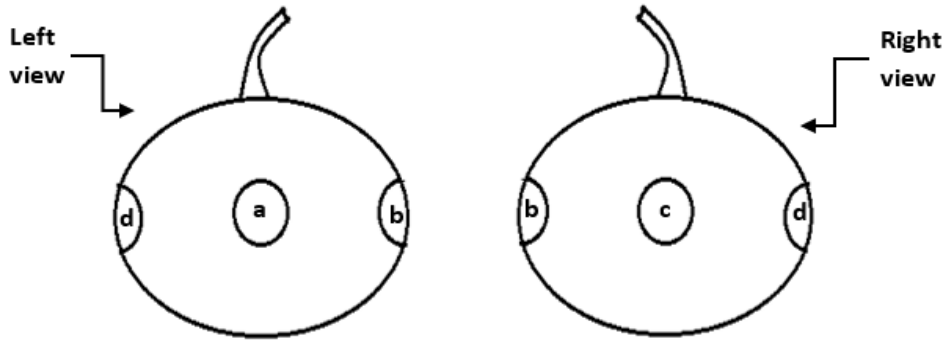


Figure 4.1: Schematic diagram of the marked four positions for NIR spectra collection in honey melons

4.2.2. Collection of Vis-NIR spectra:

Vis-NIR spectra (range 310-1100 nm) were collected using the industry standard F750 (Felix Instruments, Camas, WA, USA) equipped with Carl Zeiss MMS-1 spectrometer employing interactance optical geometry. Each pixel in the receiver has a spatial sampling of 3nm with a spectral resolution of 8-13 nm (approx.). Halogen lamp is used as a light source. The interactance spectra were collected from 101 samples (4 positions each). For each single position, average of three scans was recorded. Spectra collected from each position was treated as separate sample, hence a total of 404 spectral samples were collected.

4.2.3. SSC determination through destructive testing:

After NIR interactance spectra acquisition, the SSC values were immediately measured by conventional destructive testing procedure. For each position, the marked region from where spectral data has been collected, was cut as shown in **Figure *4.2*** and rind was removed. Rind thickness was also measured and recorded using Vernier Calipers. The extracted flesh was squeezed using manual fruit squeezer and the juice was poured onto digital refractometer (Model: PAL-1 [°Brix 0–53%], Atago Co., Ltd, Tokyo, Japan). The refractometer has automatic temperature compensation with range 10-100^{o}C and measurement accuracy of ±0.2%.

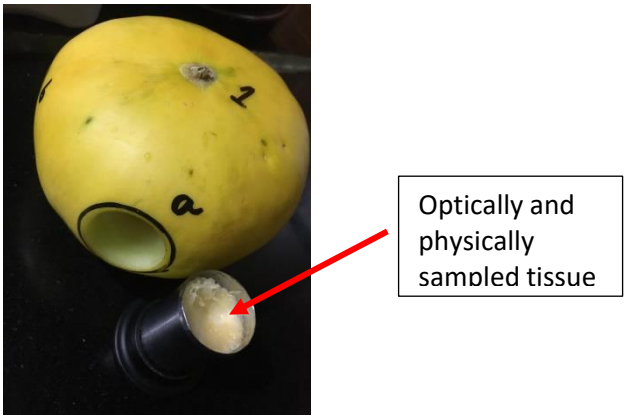


Figure 4.2: Flesh extraction for SSC determination through refractometer

4.2.4. Sensory assessment:

Reference values for sweetness were assessed by a briefly trained 10-member judges panel with age group between 20-60 years. After spectra acquisition and destructive testing, the surrounding region around the marked positions (see **Figure *4.3***) was extracted and cut into 10 equal pieces to be presented to judges. Judges were given distilled water for drinking after every sample assessment to clear previous sample taste. Melons are classified into three classes by sensory evaluation i.e. very sweet, sweet and flat. The sweetness class of each sample was characterized by average score of the panel for that sample. Class wise scoring sheet used for assessment is given in ***Table 4.1***.

Table 4.1: Score distribution for classification of sweetness level of honey melons

Class label	Score
Very sweet	8-10
Sweet	5-7
Flat	0-4

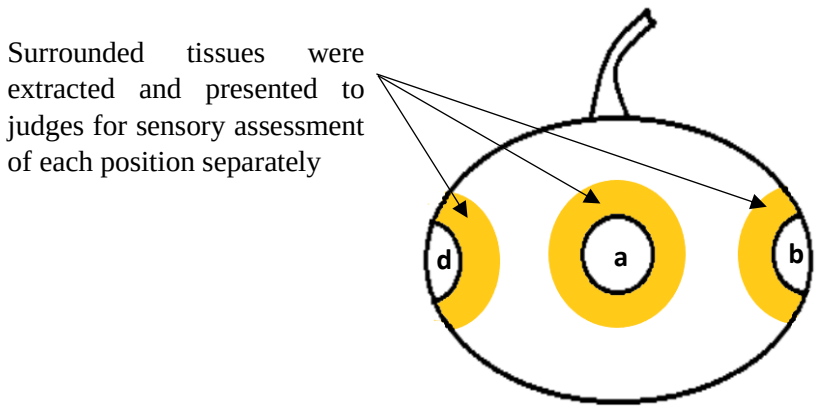


Figure 4.3: Flesh extraction for sensory evaluation of sweetness

4.3. Chemometric analysis:

To establish sweetness quality standards for melons, I propose direct classification-based decision method (method B, **Figure *4.4***) as oppose to literature reported method A of first predicting SSC using regression and then apply sharp thresholds. For Australian market SSC threshold of minimum acceptable quality standard is 10 ^{o}Brix [213], for USA it is 9 ^{o}Brix and for Chinese varieties, satisfactory class samples are above 12 ^{o}Brix and average class samples above 10 ^{o}Brix [131]. As SSC does not represent merely sugar content, I argue that instead of deciding minimum quality standard based on SSC, it may be decided because of direct sensory evaluation. Proposed method B as in **Figure *4.4*** is based on a classification model that classifies input spectral data into satisfactory and unsatisfactory class samples. Before building classification model, spectral pre-processing and PCA were applied to enhance classifier performance.

Investigated pre-processing techniques include moving average smoothing, SG first derivative, SG second derivative, MSC, SNV and normalization were applied one at a time before building calibration model. Performance of models based on each pre-processing technique was compared and analyzed.

Both linear and nonlinear multivariate calibration methods have been used and compared to build regression model for SSC prediction. Linear methods employed include PLSR and MLR while nonlinear methods include support vector machine (SVM) and artificial neural network (ANN). To estimate sweetness level of melons directly from the spectra, several supervised

learning and unsupervised learning classifiers are implemented and compared including tree, ensemble, K nearest neighbor (KNN), linear discriminant analysis (LDA), SVM and ANN.

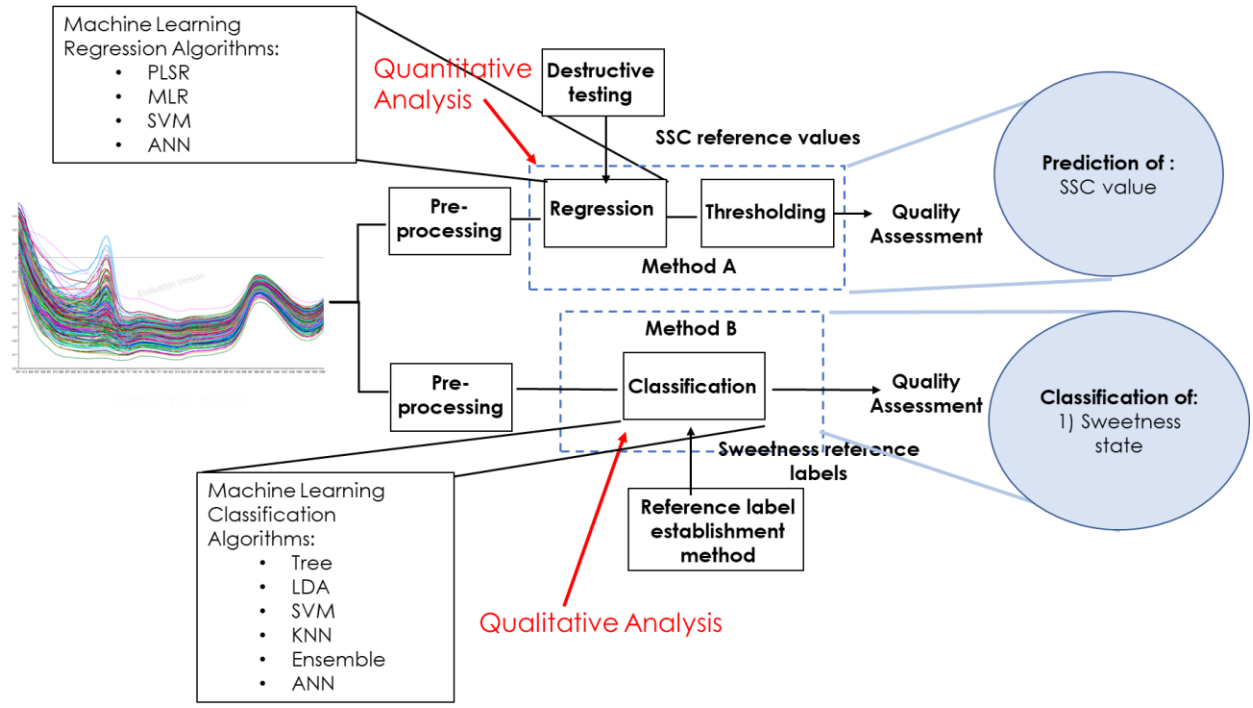


Figure 4.4: Block diagram representing two different definitions of establishing sweetness quality standard for melons

4.4. Results and discussions:

4.4.1. Vis /NIR spectra analysis:

Vis/NIR spectroscopy (range 350-2500 nm) records response of O-H, C-H, C-O and N-H bonds in fruits. Hence these organic molecules absorb energy as they vibrate because of NIR radiation exposure, which is translated into absorbance spectrum by NIR spectrometer. Short wave NIR radiation i.e. 750- 1300nm is considered as the absorbance band of high overtones i.e. 3rd and 4th overtone while common NIR (after 1300 nm) belongs to 1st or 2nd overtone.

The raw absorbance spectra of 101 melons within range 310-1100 nm is shown in **Figure *4.5***. It can be observed that all samples have similar trend in their NIR absorbance spectra. The

spectra before 450 nm and after 1080 nm is noisy. A prominent peak around 675 nm can be observed from **Figure *4.5*** which might be related with the absorbance band of chlorophyll [218]. The absorbance band around 740-760 nm is associated with the 3rd overtone of O-H bond and the band around 840nm is associated with 3rd overtone of C-H bond [219]. The absorbance band around 980 nm is related with water [220]. The penetration depth of NIR in fruits is comparatively greater in the 700-900 nm range [46] hence more information on internal quality attributes can be fetched by using wavelength absorption data of this region. Therefore, regression and classification models were developed using spectral feature region from 729-975 nm as this is the region of carbohydrates such as glucose, fructose and sucrose [211]. To support this, PLSR prediction models were generated using two ranges 1) 729-975 nm and 2) 500-1100 nm and compared in section 4.4.3.

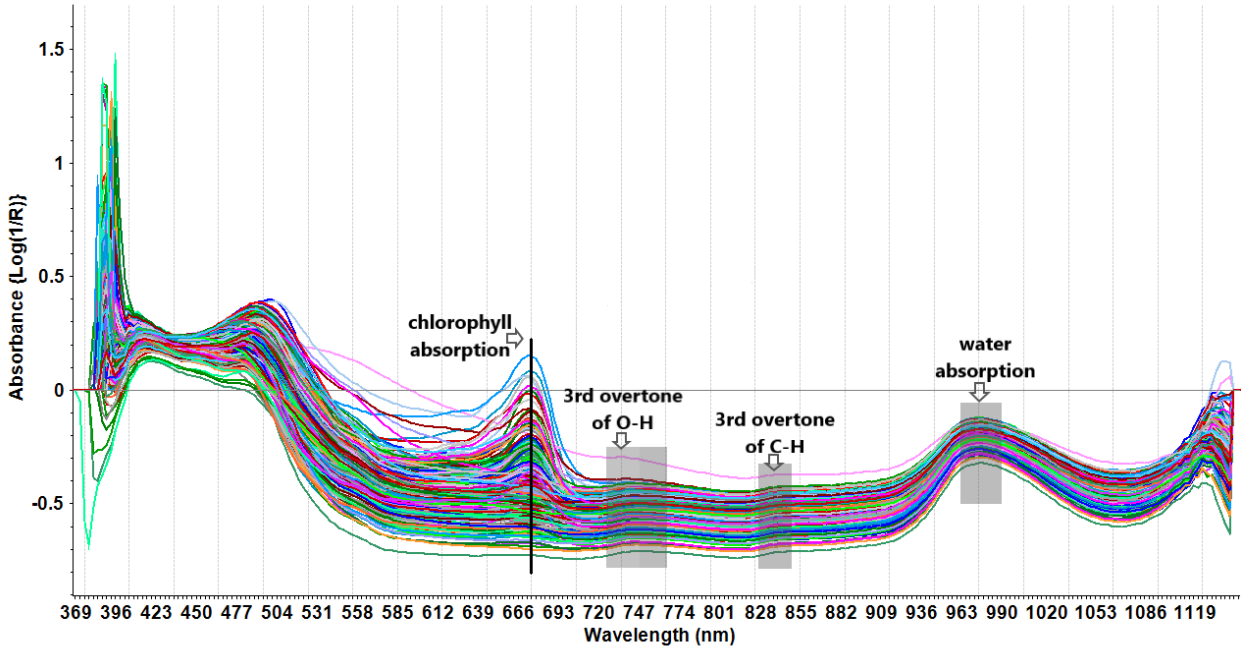


Figure 4.5: Raw absorbance spectra of 101 melons

4.4.2. Statistics of destructive testing and sensory evaluation results:

Table *4.2* lists the SSC statistics of collected datasets from 101 melon samples, which depicts that SSC values from 101 samples were fairly normally distributed around mean 7.99 °Brix with standard deviation (S.D.) 2.52 °Brix. SSC values range from 2.7-14.4, 3.1-13.9 and 2.7-14.4 °Brix

which shows sufficiently large distribution range. Moreover, the SSC range of calibration set was bigger than that of prediction set which helped in development of better prediction models [129], [221].

Table 4.2: Statistics of SSC (^{o}Brix) values of 101 samples in calibration and prediction data sets

Data sets	No. of Samples	Minimum (^{o}Brix)	Maximum (^{o}Brix)	Mean (^{o}Brix)	S.D. (^{o}Brix)
Calibration set	303	2.70	14.4	7.99	2.53
Prediction set	101	3.10	13.9	7.98	2.52
All samples	404	2.70	14.4	7.99	2.52

Figure *4.6* depicts the correlation between the sweetness levels and SSC values. From the fig it is obvious that there is significant overlap in terms of SSC levels between the three melon classes, hence based on only SSC, the sweetness criteria cannot be defined. From 404 samples (101 melons, 4 samples each), 139 samples belonged to flat class, 146 belonged to sweet class and 119 belonged to very sweet class. On average, melons with SSC over 10 ^{o}Brix were very sweet, between 7 -10 ^{o}Brix were sweet and below 7 ^{o}Brix had flat taste.

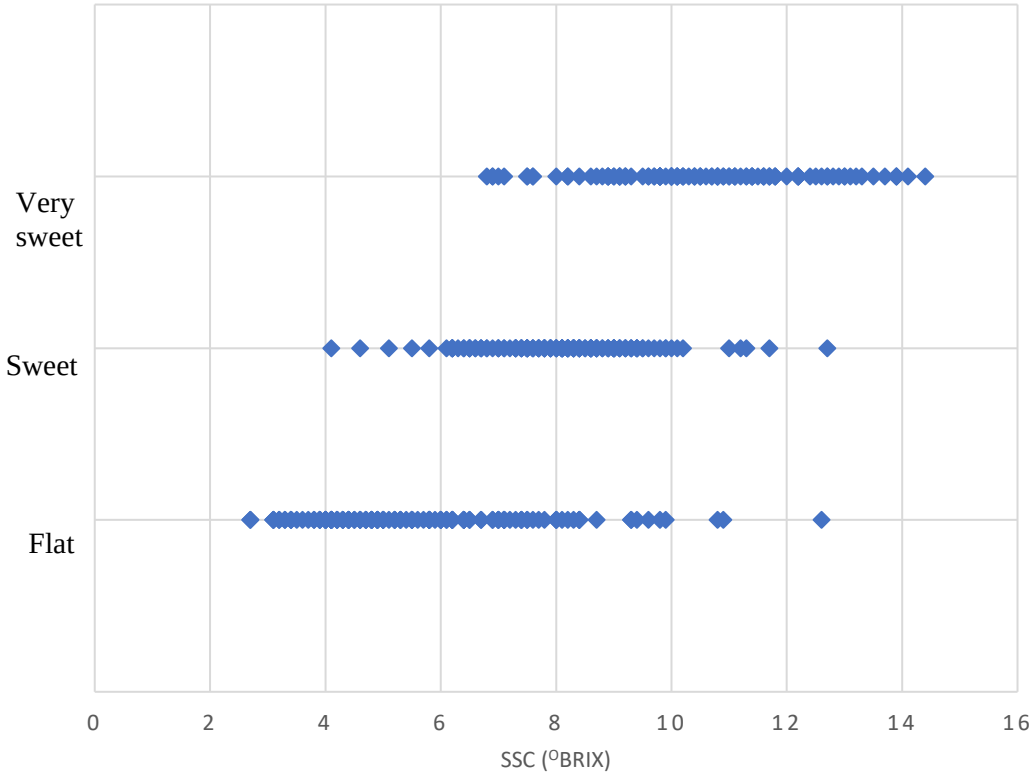


Figure 4.6: Correlation of melon sweetness levels with SSC values

4.4.3. Spectra Pre-processing and principle components extraction using PCA:

Pre-processing techniques were implemented on raw spectra in Unscrambler v11.0 spectral analysis software evaluation version (CAMO PRECESS AS, Oslo, Norway). Smoothing was done using 3 point moving average filter, normalization was performed by unit vector normalization, SG first derivative and second derivative were performed using 9 points and 21 points windows, respectively. After preprocessing, PCA was performed on the resulting transformed spectra for dimensionality reduction. **Table *4.3*** presents contribution rate of first fifteen principle components of raw and pre-processed spectral data. It shows that the first principle component PC-1 offers the main contribution and the total contribution rate of the first 15 PCs is 99.99% in all cases. The first 15 components have non-zero contribution in case of SNV and SG 1st derivative pre-processed data, first 15 components were taken as input to regression and classification algorithms. While for raw spectra and smoothing, normalization, MSC and SG 2nd derivative filter pre-processed data, selected number of PCs are first 9,9,13,13 and 9, components respectively, as other

components have zero contribution. Higher components have very less contribution, however, excluding them from feature set results in lesser accuracy of regression and classification algorithms.

Table 4.3: Contribution rates of first 15 principal components of raw and preprocessed data for different techniques

Principal Component	Preprocessing techniques (Contribution rate %)						
	Raw	**Smoothing**	**Normalization**	**SNV**	**MSC**	**SG 1st derivative**	**SG 2nd derivative**
PC-1	98.4	98.4	84.4	82.3	87.0	86.1	99.7
PC-2	1.32	1.32	13.0	15.7	11.5	7.13	$1.7e^{-1}$
PC-3	$2.5e^{-1}$	$2.5e^{-1}$	2.29	1.59	1.25	5.08	$1.2\ e^{-1}$
PC-4	$2e^{-2}$	$2e^{-2}$	$2.9e^{-1}$	0.40	$2.9e^{-1}$	$9.9e^{-1}$	4.3e-3
PC-5	3.3e-3	3.3e-3	$3e^{-2}$	1.8e-2	1.3e-2	$4.9e^{-1}$	1.6e-3
PC-6	1.1e-3	1.2e-3	6.3e-3	5.1e-3	3.4e-3	6.1e-2	7e-4
PC-7	4e-4	4e-4	2.2e-3	2.0e-3	1.4e-3	4.3e-2	5e-4
PC-8	2e-4	2e-4	1.8e-3	1.5e-3	1.1e-3	3e-2	2e-4
PC-9	2e-4	2e-4	7e-4	7e-4	4e-4	9.3e-3	1e-4
PC-10	-	-	2e-4	2e-4	2e-4	3.6e-3	-
PC-11	-	-	1e-4	2e-4	1e-4	2.4e-3	-
PC-12	-	-	1e-4	1e-4	1e-4	1.3e-3	-
PC-13	-	-	1e-4	1e-4	1e-4	9e-4	-
PC-14	-	-	-	1e-4	-	7e-4	-
PC-15	-	-	-	1e-4	-	6e-4	-

4.4.4. ºBrix based thresholding for sweetness classification:

Four regression techniques i.e. PLSR, MLR, SVM and ANN were implemented and compared in MATLAB (ver R2016a) to build SSC prediction models. Results were compared in terms of correlation coefficient (R) and root mean squared error (RMSE) values. The most celebrated regression technique, PLSR was used to test impact of two different spectral regions on the model. Performance of PLSR model built with 83 spectral features from 729-975 nm range, is compared in table 3 with PLSR model built using spectral data with 200 features (500-1100 nm). Since PCA

is an inherent part of PLSR, each feature here represents a wavelength absorbance value. For model calibration, 10-fold cross validation was used in all models. Calibration and prediction datasets were created using the sampling method discussed in section 4.2.1.

From **Table *4.4***, PLSR model developed using the spectral window 729-975 nm performed comparatively better than the spectral range of 500-1100 nm. Therefore, for rest of the techniques, the spectral range of 729-975 nm was selected as model input features. PLSR performed better when input features were pre-processed with SG second derivative in terms of prediction results i.e. R_P is 0.79, $RMSE_P$ is 1.54 ^oBrix for 725-975 nm and R_P is 0.75, $RMSE_P$ is 1.64 ^oBrix for 500-1100 nm.

Table 4.4: SSC prediction model comparison for the input spectral range 729-975 nm and 500-1050 nm *(* refers to best results)*

Pre-processing	**PLSR**							
	Calibration set				**Prediction set**			
	729-975 nm		**500-1100 nm**		**729-975 nm**		**500-1100 nm**	
	R_{CV}	**$RMSE_{CV}$**	**R_{CV}**	**$RMSE_{CV}$**	**R_P**	**$RMSE_P$**	**R_P**	**$RMSE_P$**
None	0.82	1.44	0.79	1.52	0.75	1.68	0.72	1.72
Smoothing	0.81	1.45	0.79	1.54	0.75	1.69	0.72	1.73
Normalization	0.84	1.38	0.78	1.57	0.69	1.80	0.66	1.78
SNV	0.85	1.34	0.78	1.57	0.70	1.77	0.70	1.78
MSC	0.84	1.35	0.78	1.56	0.69	1.79	0.69	1.81
SG 1st Derivative	0.85	1.32	0.82	1.44	0.72	1.71	0.75	1.64
SG 2nd Derivative	0.83	1.38	0.81	1.48	**0.79***	**1.54***	**0.75**	**1.64**

Performance of SSC prediction model built with MLR, SVM and ANN regressors is presented in **Table *4.5***. PCA is an intrinsic part of PLS regression however, for other investigated regression techniques I have compared their performance with and without PCA in **Table *4.5***. SVM models were built using linear kernel while for PCA enabled input, ANN architecture had input layer with number of inputs equal to number of PCs, one hidden layer with 20 neurons and output layer having single output. Without PCA, ANN architecture had input layer of 83 inputs, one hidden layer of 83 neurons and one output layer having single output. Number of neurons and hidden

layers are chosen heuristically. Overall, all investigated regressors gave better performance with PCA enabled input data while predicting independent test data. With prediction set, MLR had R_P = 0.93 and $RMSE_P$ = 1.63 ^oBrix with 15 principle components as input as compared to R_P = 0.90 and $RMSE_P$ = 1.84 ^oBrix with 83 spectral features as input. Similar trend can be seen for SVM and ANN regressors as well from table 5. With PCA, SVM regression model performed relatively better with MSC pre-processed spectral inputs with R_P = 0.78 and $RMSE_P$ = 1.58 for prediction set, as compared to other pre-processing techniques. ANN performed equally good when input was pre-processed with SG first derivative and SG second derivative filters along with PCA i.e. R_P= 0.76 and $RMSE_P$= 1.63 in both cases. Hence, further comparisons are done for regressors and classifiers with PCA enabled inputs.

Table 4.5: Performance comparison of investigated regressors by first applying different pre-processing techniques with and without PCA on the spectral input data *(* refers to best results)*

Pre--processing		Regressors											
		Calibration set						Prediction set					
		MLR		SVM		ANN		MLR		SVM		ANN	
		R_{CV}	$RMSE_{CV}$	R_{CV}	$RMSE_{CV}$	R_{CV}	$RMSE_{CV}$	R_P	$RMSE_P$	R_P	$RMSE_P$	R_P	$RMSE_P$
None	**With PCA**	0.82	1.47	0.82	1.43	0.89	1.15	0.92	1.67	0.75	1.66	0.74	1.73
	Without PCA	0.89	1.35	0.79	1.58	0.86	1.25	0.91	1.81	0.73	1.78	0.72	1.73
Smoothing	**With PCA**	0.82	1.47	0.82	1.43	0.88	1.21	0.92	1.69	0.74	1.69	0.75	1.69
	Without PCA	0.89	1.36	0.80	1.54	0.84	1.28	0.91	1.82	0.74	1.72	0.70	1.79
Normalization	**With PCA**	0.80	1.56	0.79	1.53	0.86	1.39	0.91	1.81	0.70	1.80	0.74	1.68
	Without PCA	0.88	1.41	0.70	1.80	0.63	1.89	0.90	1.89	0.53	2.16	0.13	2.5
SNV	**With PCA**	0.81	1.52	0.80	1.49	0.83	1.76	0.91	1.78	0.71	1.78	0.71	1.65
	Without PCA	0.88	1.36	0.76	1.65	0.76	1.57	0.90	1.87	0.64	2.02	0.58	2.06
MSC	**With PCA**	0.81	1.53	0.80	1.49	0.84	1.59	0.91	1.80	**0.78***	**1.58***	0.69	1.83
	Without PCA	0.88	1.37	0.73	1.75	0.81	1.65	0.89	1.95	0.60	2.02	0.66	2.08
SG 1st Derivative	**With PCA**	0.83	1.45	0.83	1.42	0.85	1.37	0.91	1.71	0.73	1.73	**0.79***	**1.63***
	Without PCA	0.89	1.33	0.83	1.42	0.82	1.48	0.91	1.78	0.74	1.73	0.78	1.75
SG 2nd Derivative	**With PCA**	0.83	1.43	0.83	1.40	0.84	1.37	**0.93***	**1.63***	0.75	1.66	**0.79***	**1.63***
	Without PCA	0.89	1.22	0.83	1.40	0.81	1.45	0.90	1.84	0.76	1.64	0.72	1.75

To find class labels using the predicted SSC values from regression model, best models for each of the four investigated regressors i.e. PLSR with SG 2nd derivative, MLR with SG 2nd derivative, SVM with MSC and ANN with SG 2nd derivative pre-processing are selected (Table 4.5). Prediction spectral data set was given as input to each model and respective SSC output values were noted. Thresholding was then applied on the predicted SSC values to assign a class label to each sample. Flat class label was assigned to the samples with predicted SSC value below 7 °Brix, sweet class label for samples between 7 – 10 °Brix and above 10 °Brix very sweet class label was assigned. °Brix thresholding-based class labels were then compared with reference class labels (assigned through sensory evaluation) and accuracy was calculated as number of correct labels divided by total number of samples in prediction data set. **Table *4.6*** shows accuracy of °Brix based thresholding with these four regression models. For the three classes, best accuracy was achieved by MLR regressor i.e. 55.45%. With this accuracy rate, almost 44% of the samples will be mis judged which may result in significant waste of melons. The accuracy rate can be improved by reducing the number of classes. Hence, the sweet and very sweet class samples were merged to form satisfactory class samples (all samples above 7 °Brix) and flat class samples represent unsatisfactory class samples (below 7 °Brix). Table 4.6 also shows the accuracy when the samples are classified into two classes. PLSR and MLR gave 80.2% accuracy with satisfactory and unsatisfactory class segregation.

Table 4.6: Accuracy comparison of °Brix based thresholding for melons sweetness classification *(* refers to best results)*

Regressor	**Accuracy (%)**	
	3 class	**2 class**
PLSR	53.47	**80.20***
MLR	**55.45***	**80.20***
SVM	54.46	78.22
ANN	52.48	79.21

4.4.5. Direct sweetness classification:

To predict melon's eating quality in terms of sweetness, multi class classification techniques were implemented. Classifier predictor inputs were spectral absorbance values (raw spectra/pre-

processed spectra) within range 729-975 nm whereas the reference labels were the ones generated by sensory test conducted (section 4.2.4). PCA was performed after pre-processing to optimize performance and first 15 principle components were selected (details in section 4.3.3). Various classification techniques including tree, ensemble, LDA, SVM, KNN and ANN were implemented. 10-fold cross validation was performed for calibration of all classification models. ANN architecture had input layer with number of inputs equal to number of PCs, one hidden layer with 20 neurons and output layer having single output. **Table *4.7*** presents performance comparison of the investigated classifiers in terms of accuracy. SVM and KNN models depicted best prediction accuracy among all others i.e. 70.30% with SG 1st derivative and SG 2nd derivative pre-processed and PCA enabled spectral input data, respectively. Ensemble and ANN classifiers had 64.4% accuracy with PCA enabled raw spectral input. With tree and LDA classifier the best prediction accuracy is 61.39% with 1st derivative preprocessing and PCA.

To improve classification accuracy, this multi class data was converted into binary class data by combining sweet and very sweet samples as satisfactory class and flat class labelled as unsatisfactory class. Performance comparison of same set of classifiers for binary classification is given in **Table *4.8***. Accuracy of all investigated classifiers increased for binary classification. SVM and KNN models again depicted best prediction accuracy among all others i.e. 87.13% and 88.12% with 3-point moving average filtered spectral input data and PCA, respectively.

Table 4.7: Performance comparison in terms of accuracy of different classifiers for 3 class classification *(* refers to best results)*

Pre-processing	Classifiers											
	Calibration set						Prediction set					
	Tree	LDA	SVM	KNN	Ensemble	ANN	Tree	LDA	SVM	KNN	Ensemble	ANN
None	54.1	60.7	69	65	66.7	70.3	44.5	60.4	65.35	67.33	**64.36**	**64.4**
Smoothing	53.1	62	68	62.4	62	75.2	48.51	55.45	67.33	67.33	57.43	63.4
Normalization	48.8	58.7	64.7	62	58.4	73.6	51.49	55.45	64.36	61.39	55.45	60.4
SNV	54.5	61.4	64.7	64	63	62.2	49.50	56.44	58.42	58.42	63.37	56.4
MSC	53.8	63.4	65	66	64.7	68	48.51	57.43	62.38	53.47	62.38	60.4
SG 1st Derivative	51.8	62	67	65.3	64.7	70	**61.39**	**61.39**	**70.30***	66.34	62.38	63.4
SG 2nd Derivative	52.8	61.4	65	63.4	63.4	72.6	49.50	56.44	65.35	**70.30***	55.45	63.4

Table 4.8: Performance comparison in terms of accuracy of different classifiers for binary classification *(* refers to best results)*

Pre-processing	Classifiers											
	Calibration set						Prediction set					
	Tree	LDA	SVM	KNN	Ensemble	ANN	Tree	LDA	SVM	KNN	Ensemble	ANN
None	72.3	82.2	82.8	80.5	83.2	83.2	72.28	80.20	81.19	81.19	80.20	83.2
Smoothing	70.6	82.2	83.2	80.9	80.5	81.2	73.27	**81.19**	**87.13**	**88.12***	79.21	81.2
Normalization	69.6	82.2	82.2	78.9	79.5	79.2	**78.22**	79.21	80.20	84.16	76.24	79.2
SNV	74.6	82.8	83.8	80.2	79.9	83.2	77.23	80.20	77.23	80.20	78.22	83.2
MSC	75.2	82.2	84.2	83.2	81.5	83.2	70.30	**81.19**	78.22	75.25	79.21	83.2
SG 1st Derivative	71.9	82.5	83.5	84.5	82.8	85.1	72.28	**81.19**	86.14	82.18	**83.17**	**85.1**
SG 2nd Derivative	74.3	82.8	83.8	83.5	83.2	82.2	74.26	79.21	83.17	86.14	79.21	82.2

Table *4.9* summarizes best prediction accuracy results from **Table *4.7*** and **Table *4.8*** for each classifier with 3 class and 2 class segregation. KNN and SVM classifiers have 70.30% accuracy for 3 class classification and 88.12% and 87.13% for binary classification, respectively. Comparing **Table *4.6*** and **Table *4.9***, it can be concluded that direct sweetness classification is a better non-destructive quantitative measure as compared to ^{o}Brix based thresholding to classify melons based on sweetness quality. The best results with ^{o}Brix based sweetness estimation are 55.45 % and 80.2% accuracy for 3 class and 2 class melons classification, respectively. However, with direct sweetness classification the results improved to 70.30% and 88.12% for 3 class and 2 class melons classification, respectively.

Since, direct sweetness classification proves out to be better than the ^{o}Brix based sweetness estimation, the classification models were tested again with an independent test data of season 2021. The test data of season 2021 consisted of 9 melons each scanned from 4 sides at equator position (same protocol as of season 2020), hence total 36 spectra samples were acquired, and also sensory evaluation was conducted by the same panel of judges. Among the 36 samples, 6 belonged to very sweet class, 15 to sweet class and 15 to flat class. The direct sweetness prediction results of the selected classifiers for this independent season data set is also presented in **Table *4.9***.

Table 4.9: Comparison of prediction set accuracy for multi class and binary class direct sweetness classification *(* refers to best results)*

Classifier	**Best prediction accuracy of season 2020 data set**		**Prediction accuracy of season 2021 dataset**	
	3 class	**2 class**	**3 class**	**2 class**
Tree	61.39	78.22	58.33	72.22
LDA	61.39	81.19	58.33	77.78
SVM	**70.30***	87.13	63.88	83.33
KNN	**70.30***	**88.12***	**66.67**	**86.11***
Ensemble	64.36	83.17	61.12	80.55
ANN	64.4	85.1	63.88	83.33

4.5. Conclusion:

I proposed direct classification based quantitative measure to predict melons sweetness intensity using NIR spectroscopy. An extensive evaluation was conducted on "honey melon" variety, which is grown in Pakistan. A total of 101 melons with average rind thickness 6.68 mm, were scanned from four sides at equator position. The industry standard F-750 spectrometer employing interactance optical geometry was used to collect spectral data. After spectra collection, destructive testing was performed to find reference SSC values. Sweetness standards were established by sensory evaluation by a panel of judges. Melons with SSC over 10 ^{o}Brix were considered as very sweet class samples; between 7-10 ^{o}Brix as sweet class samples and below 7 ^{o}Brix as flat class samples. Extensive chemometric analysis was carried out to obtain SSC prediction model and direct sweetness classification model for NIR spectroscopy. Raw spectral data (in wavelength range 729-975 nm) was transformed using various pre-processing techniques including smoothing, normalization, SNV, MSC, Savitzky-Golay first derivative and second derivative filtration. PCA was then performed for dimensionality reduction. Suitable SSC calibration model was then obtained by comparing different regression techniques including PLSR, MLR, SVM and ANN. SSC model obtained using MLR gave best results in terms of R for independent prediction samples (R_P=0.93 and $RMSE_P$=1.63 ^{o}Brix) with second derivative spectral pre-processing. To estimate sweetness intensity, ^{o}Brix thresholds were applied on predicted SSC values which resulted in 55.45% classification accuracy for three classes. Moreover, for direct sweetness classification, various classifiers including tree, LDA, SVM, KNN, ensemble and ANN were compared using pre-processed and PCA enabled spectral inputs and reference sweetness labels. For the three sweetness classes, KNN classifier had 70.30% accuracy. Furthermore, I observed that classification accuracy improved by combining sweet and very sweet class into one 'satisfactory' class. For ^{o}Brix thresholding-based classification the accuracy improved to 80.2% and for KNN based direct sweetness classification the accuracy improved to 88.12%. Outcomes of both these methods in terms of prediction accuracy validate that direct classification is a better quantitative measure as compared to ^{o}Brix based thresholding to estimate melons sweetness.

To further validate the performance of proposed direct classification method on other fruits and mixed cultivar datasets, next chapter presents direct sweetness classification of orange cultivars using SWNIRS.

Chapter 5 : DECISION SUPPORT SYSTEM FOR ORANGE SWEETNESS CLASSIFICATION USING DIRECT CLASSIFICATION APPROACH

5.1. Introduction:

Oranges are juicy, refreshing and most loved winter fruit in Pakistan. Pakistan is the 6th largest producer of citrus in the world [1], and around 0.46 million tons of fruit was exported in the year 2020 [222]. Ripeness is very critical as it directly influences the eating quality of harvested fruits [223]. Oranges are non-climacteric fruits i.e., they don't ripe further once they are harvested. In Pakistan, quality inspection for fruits to be exported is still carried out subjectively by the packaging industry by visualizing physical features, such as fruit color, size, sample-based tasting. The method is error prone and tedious. These factors serve as a motivation for automation of testing procedures. To automate the visual quality inspection, one can utilize camera sensors for estimating size, surface characteristics, and texture [224]. For gauging taste, sweetness, or other quality measure, one can utilize infrared spectroscopy-based methods [225]. The non-destructive assessment using NIRS can help to correlate dry matter (DM), SSC, titratable acidity (TA), and color [2] with fruit quality. Such assessment can also help in full batch testing and quality-based segregation as opposed to sample-based manual judgement.

The pulp of oranges is covered inside a thick peel, which makes penetration of NIRS challenging. Since, ripening and harvest maturity is same for non-climacteric fruits, there can be two ways to estimate ripeness/maturity. The first method is to estimate the fruit quality parameters like Brix, TA etc. using machine learning regression algorithm and based on their values judge the sample quality. The second method is to directly classify the eating quality using machine learning classification algorithm, as reported by researchers in [226] for direct sweetness classification of grapes. Like oranges, melons also have thick rind. In the previous chapter, I presented a direct sweetness classifier for melons as opposed to ^{o}Brix based thresholding, using the correlation between short-wave NIR spectroscopy and sensory assessment. The proposed direct sweetness classifier was tested on a single cultivar of melons i.e. 'honey' melons. There is a need to evaluate the correlation of short-wave NIRS and sensory assessment in other fruits as well. Moreover, the potential of short-wave NIRS and direct sweetness classification for mixed cultivar datasets needs to be analyzed.

In this chapter, the effectiveness of short-wave NIR spectroscopy and direct sweetness classification is evaluated for Pakistani cultivars of orange i.e., Blood red, Mosambi and Succari (average peel thickness 6mm). A correlation is developed between quality indices i.e., Brix, TA, Brix:TA and BrimA (Brix minus acids), sweetness of the fruit and NIR spectra which is then classified as sweet, mixed, and acidic using a machine learning classifier based on NIR spectra. I argue that direct classification is more suitable to evaluate orange sweetness as opposed to estimating quality indices.

5.2. Materials and methods:

5.2.1. Fruit samples:

Orange (*Citrus sinenses* (L.) Osbeck), cultivars (cvs.) 'Blood red', 'Mosambi' and 'Succari') ripened samples were harvested from orchard located in Chakwal district of Punjab province on two dates (33 of Blood red, 32 of Mosambi and 27 of Succari; 92 fruits in total). Average peel thickness was 6 mm. Sixty-four samples were used for model calibration, with each fruit scanned on two sides for Brix and TA to give 128 spectra. Twenty-eight samples (total 56 spectra) were used for model validation (see **Table *5.1*** for details). Samples within each fruit were treated as independent spectral set.

Table 5.1: Number of samples of investigated orange cultivars in calibration and prediction datasets.

Cultivar	Number of samples in calibration set	Number of samples in prediction set
Blood red	23	10
Mosambi	22	10
Succari	19	8
Total	64	28

5.2.2. Collection of Vis/NIR spectra:

Orange samples were marked on-tree on opposite sides i.e. sun facing side and non-sun facing side (180^{o} apart approximately) as shown in **Figure 5.1**, to account for within fruit variations. After marking samples on-tree, the oranges were harvested on two dates (both harvest dates were one week apart) and brought to a local laboratory at National Centre of Robotics and Automation (Islamabad, Pakistan) and stored at room temperature for 24 hours to minimize the influence of sample temperature on prediction accuracy [210]. Three spectra were collected from each position and average was computed. Vis-NIR spectra (range 400-1150 nm) were collected using the F-750 (Felix Instruments, Camas, WA, USA). This device employs interactance optical geometry and a Carl Zeiss MMS-1 spectrometer, with a pixel spacing of approximately 3.3 nm and a spectral resolution (FWHM) of 8-13 nm. It uses a halogen lamp as a light source.

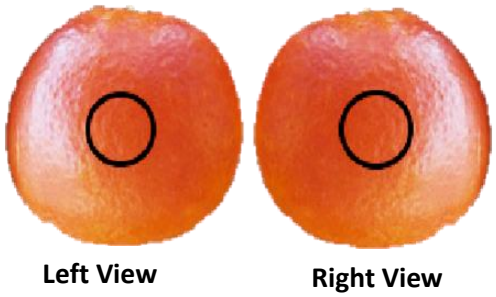


Figure 5.1: Schematic diagram of the marked positions for NIR spectra collection in oranges

5.2.3. Reference measurements:

For reference measurements, the marked region (along with surrounding tissues to get a suitable representation of the core as well) was excised and skin was removed. The extracted flesh was squeezed using a garlic press. Brix was assessed of a sample of the extracted juice using a digital refractometer (Model: PAL-1 [°Brix 0–53%], Atago Co., Ltd, Tokyo, Japan). The refractometer has automatic temperature compensation with range 10-100^{o}C and measurement accuracy of ±0.2%.

TA was measured by manual titration of 10mL of extracted juice with 0.1 M sodium hydroxide (NaOH) using phenolphthalein as an indicator. The acid formula for citrus fruit samples (Eq. 5.1) was applied to calculate TA, expressed as % citric acid [137]

$$TA\ (\%\ citric\ acid) = \frac{0.0064 * titre\ (NaOH) mL}{10mL\ (juice)}\ x\ 100 \quad (5.1)$$

$$Brix\ to\ TA\ ratio\ (maturity\ index) = \frac{Brix}{TA} \quad (5.2)$$

$$BrimA = Brix - k(TA) \quad (5.3)$$

Maturity index and BrimA were then calculated by eq (5.2) and eq (5.3) [104] respectively. The value of k in eq (5.3) is taken as 1.

5.2.4. Sensory Assessment:

Reference values for sweetness were assessed by a briefly trained five judges panel with age between 20 to 50. After spectra acquisition, two slices were cut from the neighbor region from where destructive testing has been performed and presented to two of the judges at random for taste evaluation. Distilled water was provided to judges for drinking after every sample evaluation to clear previous sample taste. Oranges were classified into three classes by sensory evaluation i.e. Sweet, mix (sweet and acidic both) and acidic. The class label of each sample was described by average score of the two judges for that sample. Class wise scoring sheet used for assessment is given in **Table *5.2***.

Table 5.2: Score distribution for classification of sweetness level of oranges

Class label	Score
Sweet	8-10
Mix	5-7
Acidic	0-4

5.2.5. Chemometric Analysis:

A direct sweetness classification method has been proposed in previous chapter for melons sweetness classification as opposed to indirect measure of ^{o}Brix estimation. In this chapter, I have implemented both the methods for quality assessments of oranges as shown in **Figure 5.1**. The first

method shows an indirect approach for classification that estimates fruit quality index parameters values using machine learning regression algorithm and based on those predicted values, the quality of the sample is classified. The second method shows direct classification approach that directly classifies test sample as sweet, acidic or mix class sample, using machine learning classification algorithm.

11-point SG second derivative preprocessing was performed on spectral data. For indirect quality assessment, partial least squares regression was used to build Brix, TA, Brix:TA and BrimA estimation models. PCA was applied on spectral data and then several supervised and unsupervised learning classifiers are implemented and compared including tree, ensemble, KNN, linear discriminant analysis (LDA) and SVM.

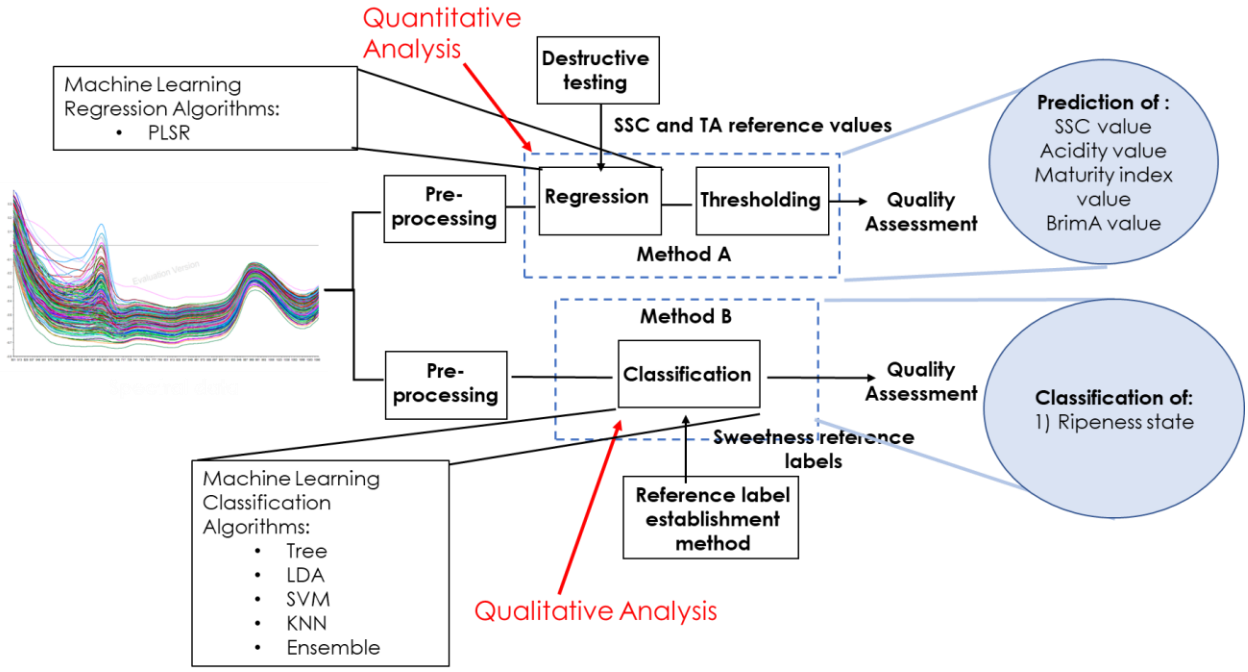


Figure 5.2: Block diagram representing two different methods of orange quality assessment

For indirect method to classify fruits, the Unscrambler v11.0 spectral analysis software evaluation version (CAMO PRECESS AS, Oslo, Norway) was used for building combined variety calibration model using calibration dataset (**Table *5.4***). 11 points Savitzky-Golay second derivative smoothing filter was applied before building model. The performance of developed models was evaluated by R_{CV} (correlation coefficient of cross validation), R_P (correlation coefficient of

prediction), $RMSE_{CV}$ (root mean square error of cross validation) and $RMSE_P$ (root mean square error of prediction). 10-fold cross validation was performed. Partial least squares regression (PLSR) models were developed using the Vis/NIR region in the range 600-1050 nm (following [35]).

For direct classification, MATLAB R 2018a software was used. Input data for both the methods i.e. direct and indirect method of classification was same (pre-processed with 11-point SG second derivative using Unscrambler software). Classification was performed using MATLAB classification learner module with PCA enabled (first 15 principal components were used).

5.3. Results:

5.3.1. Dataset statistics:

Destructive testing statistics of orange quality index parameters i.e. Brix, TA, maturity index and BrimA with respect to the individual variety are shown in **Table *5.3***. The range and mean of Blood red cultivar is relatively low for Brix, Brix:TA and BrimA, and high for TA as compared to other two varieties. **Table *5.3*** shows that the statistics of Succari cultivar are dissimilar from the other two investigated cultivars with respect to TA and Brix:TA i.e. TA range (0.14-0.33%) and mean (0.21%) is lowest and maturity index range (33.64-75.63) and mean (55.38) is highest than that of Blood red and Mosambi cultivars.

Since, Succari cultivar is statistically different from the other two cultivars, the models were built using two different combinations of investigated cultivars i.e. dataset-1 contains all three investigated cultivars and dataset-2 contains only Blood red and Mosambi cultivars. **Table *5.4*** shows data set wise statistics of quality index parameters.

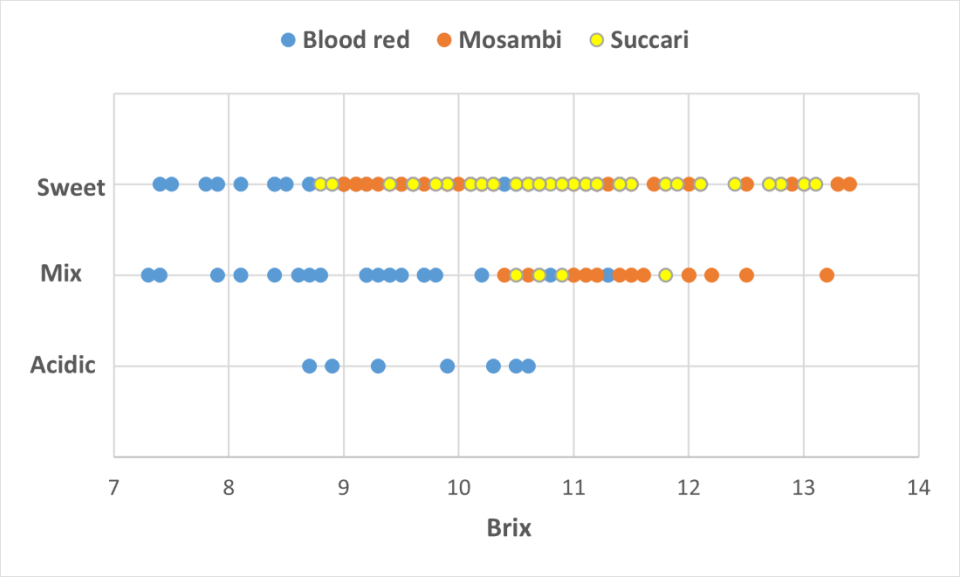
Table 5.3: Statistics of Brix, TA, maturity index and BrimA with respect to the individual investigated varieties of orange

Dataset	Number of Samples	Range				Mean				S.D.			
		Brix (ºBrix)	TA (%)	Brix:TA	BrimA (%)	Brix (ºBrix)	TA (%)	Brix:TA	BrimA (%)	Brix (ºBrix)	TA (%)	Brix:TA	BrimA (%)
Blood red	33	7.3-11.3	0.59-1.98	5.3-12.71	6.53-9.97	9.22	1.03	9.37	8.2	1.04	0.29	1.67	0.88
Mosambi	32	9-13.4	0.4-1.12	9.82-24.69	8.51-12.73	10.98	0.68	16.9	10.31	1.22	0.19	3.57	1.12
Succari	27	8.8-13.1	0.14-0.33	33.64-75.63	8.54-12.9	11.03	0.21	55.38	10.77	1.02	0.04	11.09	0.99

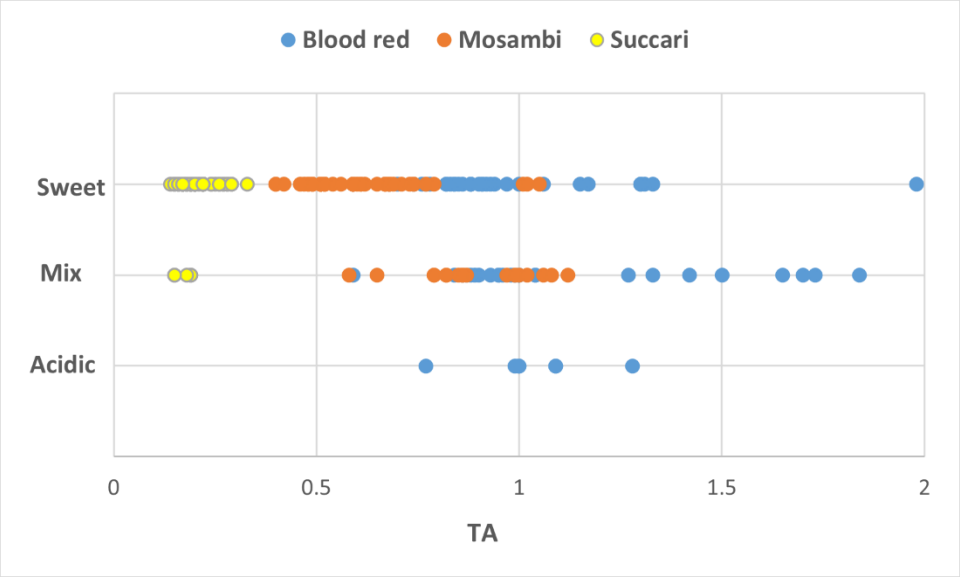
Table 5.4: Statistics of reference values with respect to calibration and prediction data sets

Dataset		Total Samples	Min				Mean				S.D.			
			Brix (ºBrix)	TA (%)	Brix: TA	BrimA (%)	Brix (ºBrix)	TA (%)	Brix: TA	BrimA (%)	Brix (ºBrix)	TA (%)	Brix: TA	BrimA (%)
Dataset1: (Blood red, Mosambi and Succari)	**Calibration**	128	7.4-13.4	0.14-1.98	5.3-75.63	6.53-12.73	10.37	0.69	24.91	9.68	1.39	0.39	20.89	1.49
	Prediction	56	7.3-13.1	0.17-1.5	6.2-65.5	6.56-12.9	10.3	0.62	25.87	9.64	1.36	0.37	19.70	1.55
	Total	184	7.3-13.4	0.14-1.98	5.3-75.63	6.53-12.9	10.36	0.67	25.20	9.70	1.39	0.39	20.48	1.52
Dataset2: (Blood red and Mosambi)	**Calibration**	90	7.4-13.4	0.4-1.98	5.3-22.94	6.53-12.73	10.20	0.89	12.54	9.31	1.52	0.29	4.22	1.53
	Prediction	40	7.3-12.1	0.4-1.5	6.2-24.69	6.56-11.61	9.83	0.79	14.17	9.03	1.20	0.30	5.44	1.25
	Total	130	7.3-13.4	0.4-1.98	5.3-24.69	6.53-12.73	10.08	0.86	13.05	9.23	1.43	0.3	4.67	1.45

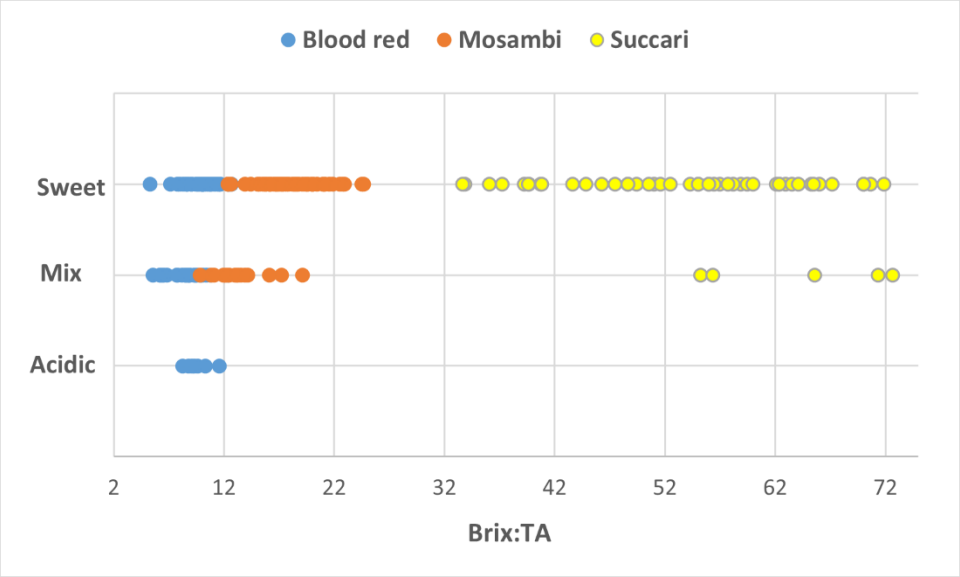
Figure *5.3* shows the correlation of orange sweetness levels and quality index parameters values. From 184 samples (92 oranges, 2 samples each), 129 samples belonged to sweet class, 48 belonged to mix class and 7 belonged to acidic class. From **Table *5.3***, the sweetness levels cannot be concluded based on individual values of Brix, TA, Brix:TA or BrimA, since there is significant overlap between the three sweetness levels and the respective quality indexes. Moreover, it can be concluded that with respect to quality index parameters, Succari cultivar is dissimilar to the other two investigated varieties.



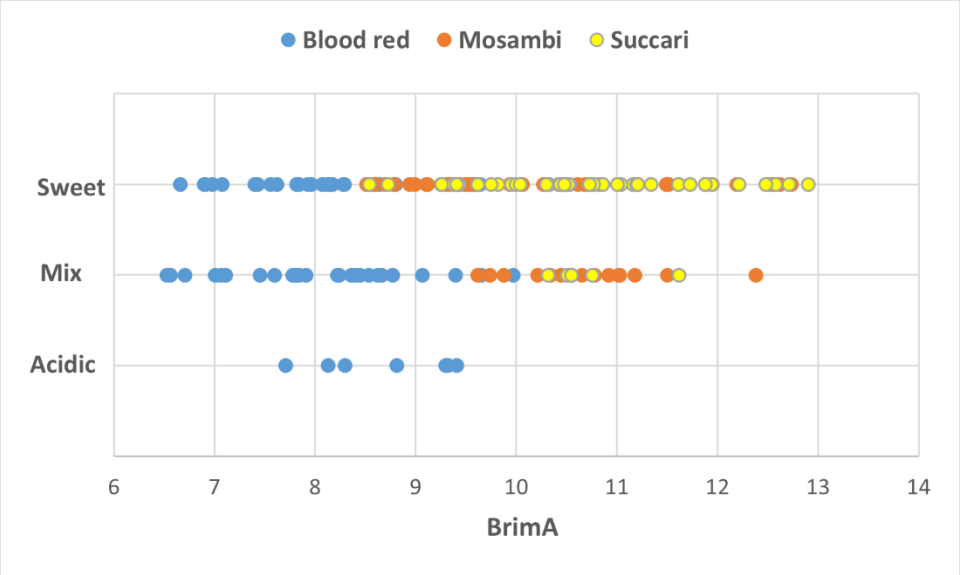
(a)



(b)



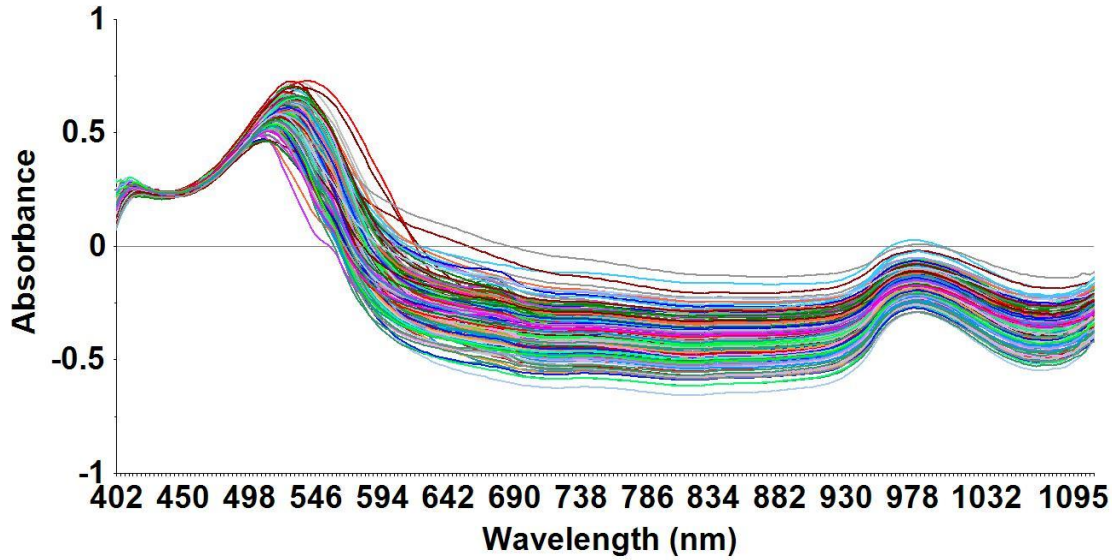
(c)



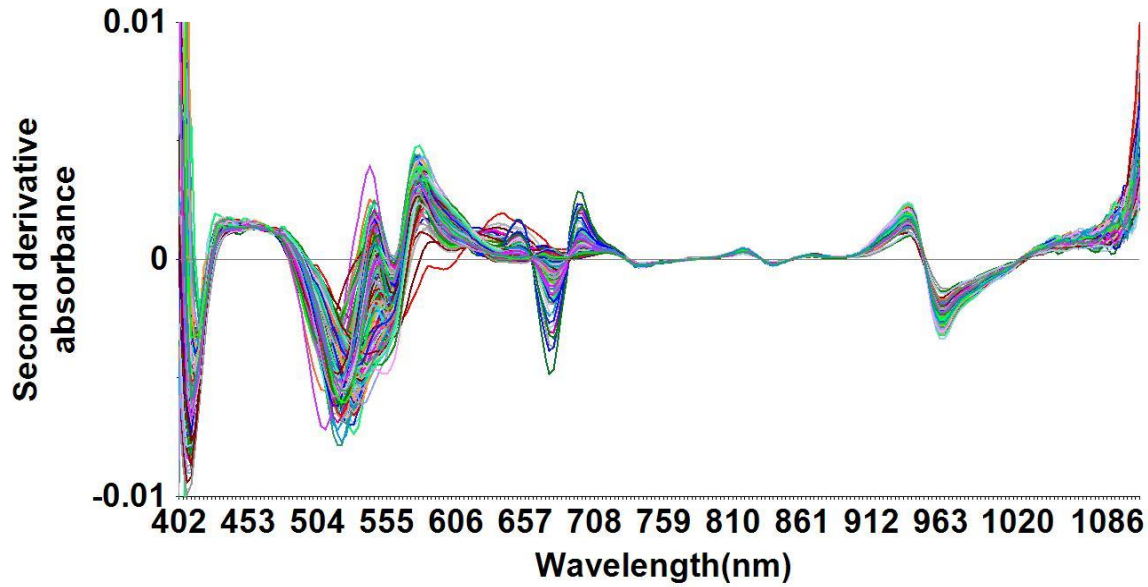
(d)

Figure 5.3: Correlation of orange taste quality levels with (a) Brix, (b) TA, (c) maturity index and (d) BrimA

5.3.2. Overview of Spectra:



(a)



(b)

Figure 5.4: (a) Raw absorbance and (b) Savitzky-Golay second derivative spectra of collected dataset

The absorbance spectra of orange fruit (**Figure *5.4***) is dominated by a peak around 680 nm associated to chlorophyll absorption [218] and a 970 nm peak associated with water absorption band [227].

5.3.3. Indirect method of sample classification results:

Table *5.5* presents the combined variety PLSR model results on Brix, TA, Brix:TA and BrimA with dataset having 184 samples including all three investigated varieties. The cross validation R_{CV} is 0.69, 0.48, 0.5 and 0.66 respectively and $RMSE_{CV}$ is 1.00 ^{o}Brix, 0.34%, 18.06 and 1.13% respectively.

These models include samples of Succari variety as well, which is a statistically incompatible cultivar (with respect to TA and Brix:TA) with the Blood red and Mosambi cultivars. Hence, Table 6 shows PLSR models trained on Blood red and Mosambi cultivars since they are similar to each other w.r.t TA and Brix:TA statistics. **Table *5.6*** shows that excluding Succari samples from dataset and rebuilding PLSR models provided improved results for TA and Brix:TA models. However, Brix and BrimA prediction results were worsened because with respect to Brix, all three investigated varieties have similar statistics. Removing Succari samples reduced the size of data set and hence worse results.

Table 5.5: Cross validation and prediction results for PLSR models developed for dataset1 (Blood red, Mosambi and Succari)

Index	**PLSR model**			
	Cross validation		**Prediction**	
	R_{cv}	**$RMSE_{cv}$ (^{o}Brix/%)**	**R_{P}**	**$RMSE_{P}$ (^{o}Brix/%)**
Brix	0.69	1.00	0.57	1.05
TA	0.48	0.34	0.25	0.48
Brix:TA	0.5	18.06	0.39	20.99
BrimA	0.66	1.13	0.55	1.35

Table 5.6: Cross validation and prediction results for PLSR models developed for dataset2 (Blood red and Mosambi)

Index	PLSR model			
	Cross validation		Prediction	
	R_{cv}	$RMSE_{cv}$ (ºBrix/%)	R_P	$RMSE_P$ (ºBrix/%)
Brix	0.83	1.10	0.43	1.18
TA	0.59	0.23	0.73	0.19
Brix:TA	0.43	3.79	0.66	3.14
BrimA	0.58	1.23	0.29	1.33

5.3.4. Direct classification results:

To predict orange's eating quality in terms of sweetness, multi class classification algorithms were implemented on both datasets. The cross validation and prediction results for both data sets are listed in **Table *5.7*** and **Table *5.8***. For dataset1, ensemble classifier achieved 81.03% accuracy for 3 class classification of independent test data. For dataset2, SVM and KNN both achieved 79.49% accuracy for 3 class classification of independent test data.

Table 5.7: Cross validation and prediction results for 3 class classification for dataset1 (Blood red, Mosambi and Succari cultivars)

Classifiers	Cross validation accuracy (%)	Prediction set accuracy (%)
Tree	57.5	72.41
LDA	56.7	60.34
SVM	64.2	60.34
KNN	63.4	72.41
Ensemble	58.2	**81.03**

Table 5.8: Cross validation and prediction results for 3 class classification for dataset2 (Blood red and Mosambi cultivars)

Classifiers	Cross validation accuracy (%)	Prediction set accuracy (%)
Tree	57.8	64.10
LDA	53.3	76.92
SVM	60	**79.49**
KNN	66.7	**79.49**
Ensemble	57.5	71.79

5.4. Observations and discussion:

5.4.1. Statistics comparison of investigated cultivars:

The "Blood red" variety is the most tasteful (mix to sweet taste) cultivars of orange in Pakistan. **Table *5.3*** shows that its range and mean of TA is high and of Brix, Brix:TA and BrimA is low. From 66 samples of Blood red, 33 belonged to sweet class, 26 belonged to mix class and 7 belonged to acidic class.

The Mosambi cultivar is also segregated as sweet by the judges. It can be seen from **Table *5.3*** that its range and mean of TA is lesser and for Brix its higher than Blood red cultivar hence its flavor is generally more sweater than Blood red variety. Amongst 64 samples of Mosambi, 46 belonged to sweet class and 17 belonged to mix class.

The Succari cultivar is a different cultivar in terms of sweetness from the other two cultivars. Succari samples always have a flat sweet taste due to lack of acids contents. The statistics of quality index parameters also support this claim as its TA range and mean is the lowest and hence Brix:TA values are the highest amongst other investigated. Amongst 54 samples of Succari, 47 belonged to sweet class and only 5 belonged to mix class.

5.4.2. Development of mix cultivar PLSR models:

An attempt was made to predict Brix, TA, Brix:TA and BrimA using PLSR regression models developed for mixed cultivar datasets. Since, Succari cultivar is statistically (w.r.t TA and Brix:TA) and taste wise different from the other two investigated cultivars, PLSR models were built for two datasets, one having mixture of statistically different cultivars i.e. Blood red, Mosambi and Succari and other one having only statistically compatible cultivars i.e. Blood red and Mosambi.

It is observed that since all three investigated cultivars have almost similar Brix and BrimA statistics (**Table *5.3***), hence the model built with data set having all three cultivars achieved better prediction results for Brix and BrimA as compared to the model built with dataset having only two cultivars i.e. Blood red and Mosambi (**Table *5.5*** and **Table *5.6***). This is because dataset 2 has lesser number of samples than dataset1. The TA and Brix:TA results of PLSR models built with only two cultivars data (Blood red and Mosambi) achieved relatively better prediction results than the three cultivar dataset.

5.4.3. Direct vs indirect methods of classification:

Dataset standard deviation (S.D.) is important to determine the value of the NIRS technique for fruit quality assessment [25]. The technique holds significance only when S.D. of the attribute of interest is greater than the measurement $RMSE_P$. Indeed, the prediction set R is directly related to measurement bias corrected $RMSE_P$ and S.D. i.e., for a particular bias corrected $RMSE_P$, higher S.D. will result in higher R_P value [25].

For indirect method of sample classification, it is observed that the R_{CV} and R_P values of the developed PLSR models are low however, the $RMSE_{CV}$ and $RMSE_P$ are below the S.D. of the datasets (for Brix and BrimA considering S.D. of dataset1 and for TA and Brix:TA considering S.D. of dataset2) (see **Table *5.3*** - **Table *5.6***). The low R_P values are because of low S.D. of the collected dataset, which is a limitation for the presented work as well.

I observed (see **Table *5.7*** - **Table *5.8***) good correlation between NIR spectra and sensory assessment as opposed to quality indices. Hence, direct classification is more suitable for orange sweetness classification using NIR spectroscopy as opposed to estimation of quality indices.

5.5. Conclusion:

The research was carried out to develop a correlation between quality indices i.e. Brix, titratable acidity (TA), Brix:TA and BrimA (Brix minus acids), sensory assessment of the fruit and NIR spectra that was then classified as sweet, mixed, and acidic based on NIR spectra for Pakistani cultivars of orange i.e., Blood red, Mosambi and Succari. NIR spectral data was obtained using the industry standard F-750 fruit quality meter (310-1100 nm). Reference Brix and TA measurements were taken using standard destructive testing methods. Reference taste labels i.e. sweet, mix and acidic, were acquired by sensory evaluation of samples by a panel of judges. I observed that Succari cultivar is statistically dissimilar from the other two cultivars w.r.t TA and Brix:TA. Hence, chemometric analysis was carried out on two datasets i.e. dataset1 (Blood red, Mosambi and Succari) and dataset 2 (Blood red and Mosambi samples only), to obtain prediction models for quality indices and sweetness classification model. Better results of partial least squares regression (PLSR) models for Brix and BrimA were achieved for dataset1 with correlation coefficient (R) values of 0.57 and 0.55 on independent test data, respectively. For TA and Brix:TA, better PLSR models were achieved with dataset2 with R values of 0.73 and 0.66 on independent test data, respectively. For direct fruit classification, ensemble classifier achieved 81.03% accuracy with dataset1 and KNN and SVM classifier achieved 79.29% accuracy with dataset2, for 3 class (sweet, mix and acidic) classification on independent test data. I observed good correlation between NIR spectra and sensory assessment as opposed to quality indices. Hence, direct classification is more suitable for orange sweetness classification using NIR spectroscopy as opposed to estimation of quality indices. In the next chapter, SWNIRS based fruit classification system is presented for automated fruit type classification in contrast to computer vision based classification methods.

Chapter 6 : DECISION SUPPORT SYSTEM FOR FRUIT TYPE CLASSIFICATION USING SWNIR SPECTROSCOPY

6.1. Introduction:

The Vis and NIR region of the light spectrum has the range 400-750 nm and 750-2500 nm, respectively. The short wave NIR (SWNIR) or Herschel region lies between 750-1100 nm and the extended NIR region lies between 1100-2500 nm. NIR spectroscopy involves the measurement of the absorbance of light linked with the vibration of molecular bonds. For intact fruits, this usually entails absorption linked with the stretching of O-H and C-H bonds [40], [41], related principally with water and storage reserves (the major macro constituents). The infrared region (>2500 nm) provides the fundamental absorption bands associated with these features with more fine and higher absorption peaks than linked with the overtones observed in the NIR region. NIR region gives lower absorptivity of overtones features than those of infrared (IR) region. Due to this characteristic, the effective pathlengths of NIR radiation through fruit are in the order of millimeters to centimeters rather than micrometers as in the case of IR radiation. Similarly, with the short-wave NIR (SWNIR) region longer effective pathlengths are achieved with higher overtone features as compared to the overtones in the NIR region. This is the reason that SWNIR radiation can be used to estimate fruit parameters of intact fruits [25].

Absorption at 770 nm and 960 nm is associated with a third and second overtone of O-H stretching, respectively and 840nm is associated with the O-H combination feature. While absorption at 910 nm and 1100 nm is associated with a third and second overtone of C-H stretching, respectively [41]. These overtone features have broad peaks with a full width half maximum (FWHM) of 20nm (the reported peaks are centers of the respective bands). However, the overall peak positions shift with temperature and solute concentration because the amount of H-bonding can change which influences the vibration of O-H bonds. In practice, it is hard to interpret short-wave NIR (being second and third overtones, so weak and broad) compared to extended NIR and IR regions. However, the features related to water can be interpreted, which is the main NIR active molecule in fruit. Around 80-90% of fleshy fruit is composed of water. This is the reason that any other parameter is measured with reference to the large absorption features of water. An increase in any other macroconstituent, e.g., SSC and DM, causes a decrease in water content resulting in a negative correlation with water. The penetration depth of NIR in fruits is comparatively greater in

the 700–900 nm range [46] hence more information on internal quality attributes can be fetched by using wavelength absorption data of this region. Understanding the raw absorption spectra is challenging because all absorption characteristics are wide and overlapping. The use of NIRS over IR spectroscopy is successful due to chemometrics, which enabled useful information to be fetched out of the spectra.

Point spectroscopy delivers a sum of light absorption and scattering. The scattering characteristics of tissue affect the effective penetration depth of light into that tissue [228]. However, the quantity of light scattering differs between fruit types [191], [228]. This fact can be used to classify different fruits. In this chapter, the potential of SWNIR spectroscopy, primarily with respect to O-H and C-H overtone features, is analyzed for fruit type classification [38]. Different shallow machine learning architectures were trained to classify fruits using spectral feature vectors obtained by the industry-standard F-750 fruit quality meter. Two types of feature sets are used to compare the classification potential of both sets i.e. (1) 83 features within 725-975nm range and (2) only 4 features at wavelengths 770nm, 840nm, 910nm, and 960nm corresponding to O-H and C-H overtone features.

6.2. Materials and Methods:

6.2.1. Collection of Vis/NIR Spectra:

Vis-NIR spectral data (wavelength range 400-1150nm) is collected using the industry standard F-750 fruit quality meter (Felix Instruments, Camas, WA, USA). The dataset for apple, grapes, mango, melon, orange, loquat, plum and apricot are collected for local cultivars of Pakistan while online available datasets of F-750 [7] for cherry, hass and kiwi have been used. Samples of apple, grapes, mango, melon, orange, loquat, plum and apricot were scanned from equator position with F-750. **Table *6.1*** shows the details of fruit datasets used.

Table 6.1: Details of datasets collected for investigated fruits

Fruit	Variety	Season	Number of Samples	Training and testing dataset samples
Apple	Golden Delicious, Red Delicious-Pak, Red Delicious-Turk	2019	35	Training
		2020	15	Testing

Grapes	Sundar Khani	2020	100	70 Training, 30 Testing
Mango	SB Chaunsa	2019	150	Training
		2020	50	Testing
Melon	Honey melons	2020	150	Training
		2021	50	Testing
Orange	Red-Blood	2020	74	54 Training, 20 Testing
	Mosambi		64	44 Training, 20 Testing
	Succari		54	44 Training, 10 Testing
Loquat	Tanaka	2020	150	Training
		2021	50	Testing
Plum	Fazle Manani	2020	200	150 Training, 50 Testing
Apricot	Badami	2020	200	150 Training, 50 Testing
Cherry	Lapins	2016	200	150 Training, 50 Testing
Hass	unknown	2017	200	150 Training, 50 Testing
Kiwi	General Kiwis	2016	200	150 Training, 50 Testing

6.2.2. Chemometric Analysis:

The basic idea behind finding fruit class from short wave NIR spectral features is to construct a linear predictor function that generates a score from a set of weights that are linearly combined with the set of spectral features of given observations as:

$$Scores\ (X_i, k) = \alpha_k\ .\ X_i \tag{6.1}$$

where "." is the dot operator, X_i is the spectral features vector corresponding to the ith observation, α_k is the coefficients vector corresponding to class k and $Scores\ (X_i, k)$ is the score associated with allocating observation i to class k. The predicted class is the one that has highest score. Machine learning classification algorithms determine the optimal scores and interpret them to assign class to observation i.

For fruit type classification, MATLAB R 2021a software was used. Classification was performed using MATLAB classification learner module with PCA enabled (first 7 principal

components were used). Principle component analysis (PCA) has been widely used with spectroscopic data [4,9] to emphasize variation and bring out strong patterns in the data set. PCA was applied on spectral data and then several supervised and unsupervised machine learning classifiers are implemented and compared including tree, ensemble, K nearest neighbor (KNN), linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), support vector machine (SVM), naïve bayes and artificial neural network (ANN).

6.3. Results:

6.3.1. Overview of spectra:

The average absorbance spectra of all fruits within SWNIR range (**Figure *6.1***) is dominated by small peaks around 770 nm, 840 nm and a strong peak around 960 nm associated with water absorption band [227].

6.3.2. Principal Component Analysis:

PCA scores plot for the collected data set is shown in **Figure 6.2** *(a and b).* Observations that are similar together form a cluster in scores plot of PCA. **Figure 6.2** depicts that all the eleven investigated fruits form well defined clusters in both the cases i.e. (a) with 83 features from 725-975nm wavelength range (3nm resolution) and (b) with only 4 features at wavelengths 770nm, 840nm, 910nm and 960nm. Hence, the plots depict good potential for classification algorithms to distinguish fruits using SWNIR region or simply only the OH and CH overtones features.

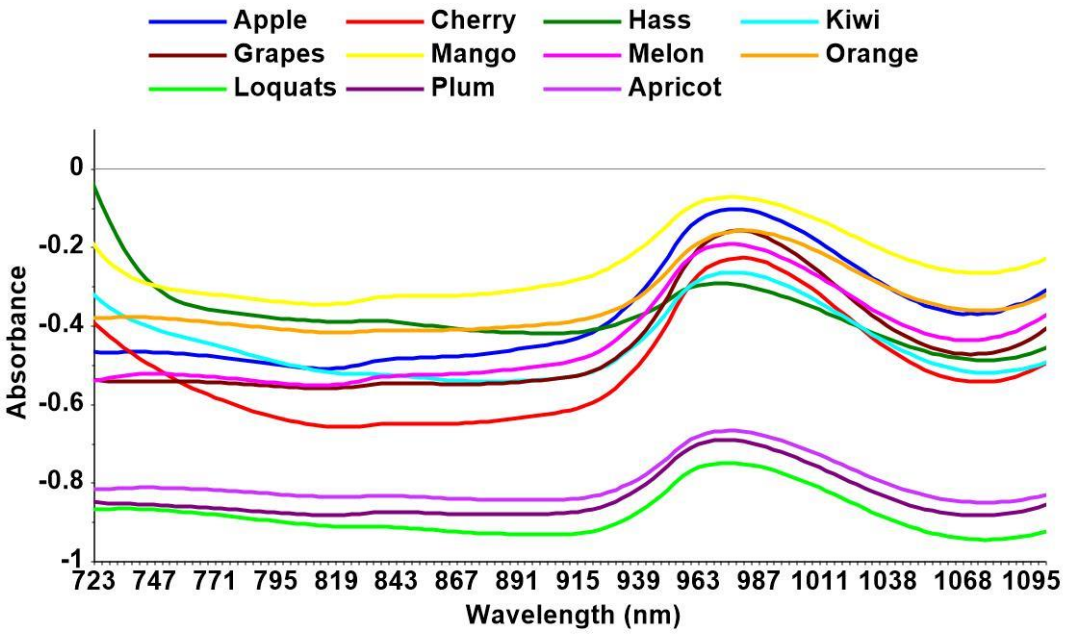
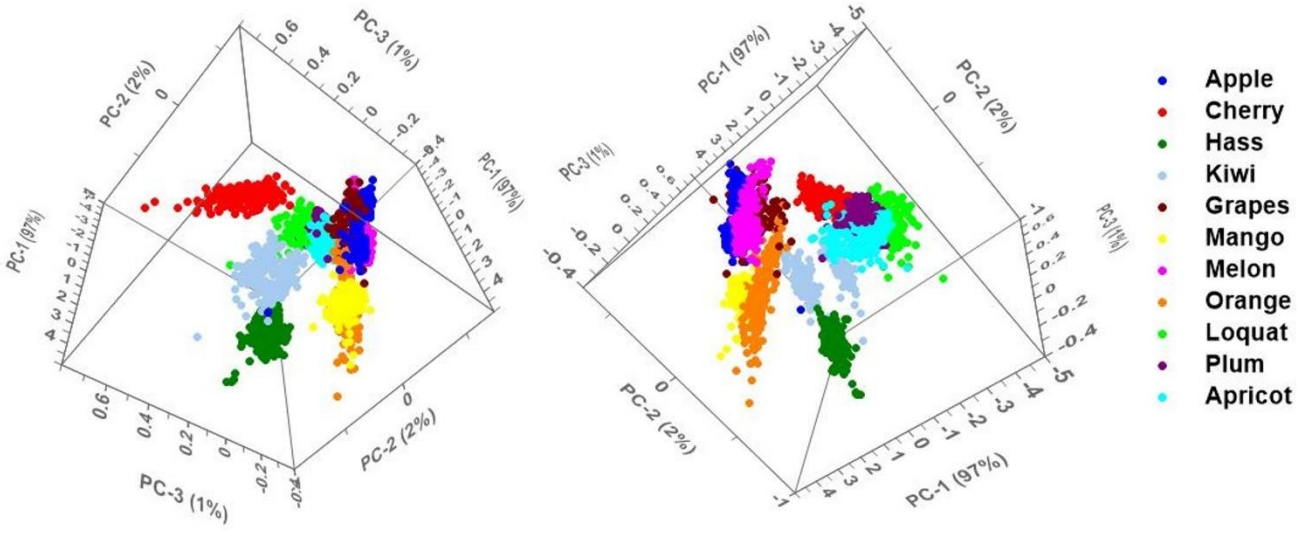
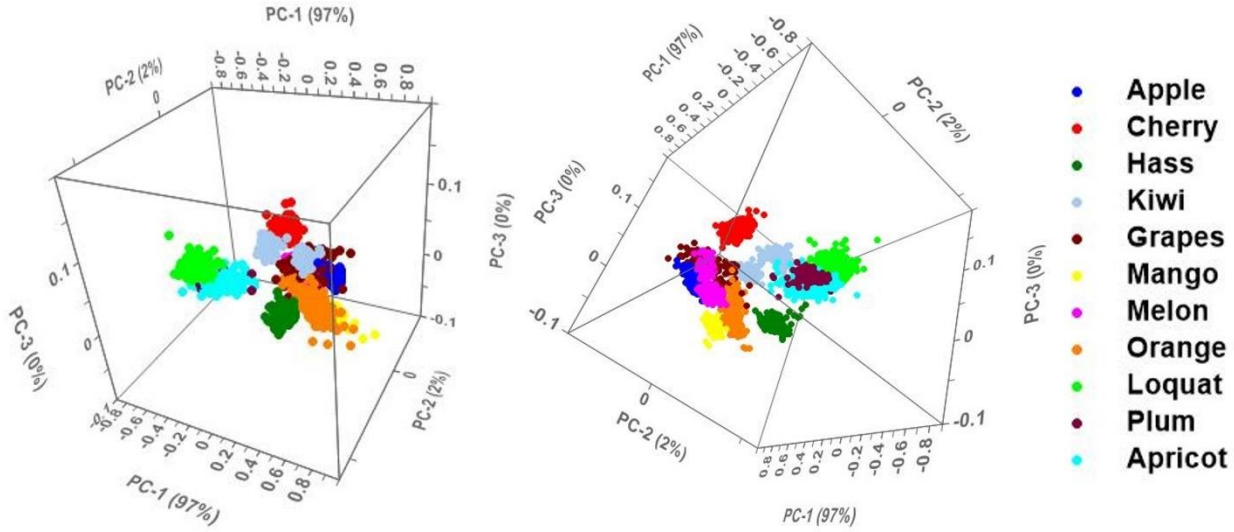


Figure 6.1: Average raw absorbance spectra of collected datasets for the investigated fruits



(a)



(b)

Figure 6.2: 3D scores plot for first three principal components of spectra of the investigated fruits (a) 725-975 nm wavelength range, and (b) only 4 features at wavelengths 770nm, 840nm, 910nm and 960nm

6.3.3. Classification Results:

Table *6.2* shows the classification results in terms of cross validation (CV) accuracy and test data accuracy for 83 features (within range 725-975nm with 3nm resolution) as input to the classifier. Training dataset has 1507 observations for the eleven fruits while the test data has 531 observations (some of them are of independent season). QDA classifier outperformed other investigated classifiers with 100% CV accuracy and 93.02% test accuracy.

The classification results (**Table *6.3***) in terms of CV accuracy and test data accuracy for only 4 features at wavelengths 770nm, 840nm, 910nm and 960nm (OH and CH overtone features) as input to the classifier. QDA classifier outperformed other investigated classifiers with 97.1% CV accuracy and 90.38% test accuracy.

Table 6.2: Classification accuracies using 83 features within wavelength range 725-975nm

Classifier	CV accuracy	Test accuracy
Tree	97.9	88.68
LDA	99.3	90.94

QDA	100	93.02
Naïve Bayes	98.9	91.89
SVM	99.8	92.26
KNN	99.7	93.21
Ensemble	99.3	90.57
ANN	99.4	91.89

Table 6.3: Classification accuracies using 4 features at wavelengths 770nm, 840nm, 910nm and 960nm

Classifier	CV accuracy	Test accuracy
Tree	94.8	87.55
LDA	95.7	87.92
QDA	97.1	90.38
Naïve Bayes	94	81.32
SVM	98.3	90
KNN	97.7	89.81
Ensemble	96.6	89.43
ANN	97.5	89.81

6.4. Observations and Conclusion:

The results obtained from PCA, and classification suggest that fruit classification is mainly a function of OH and CH overtone features. PCA results depict that both the feature sets, i.e., 83 features from the 725-975nm wavelength range and only 4 features representing the OH and CH overtones peaks within the 725-975nm range, form well-defined clusters in the PCA space. The PCA scores plot depicts a good intra-cluster correlation for all the eleven clusters in both cases (**Figure 6.2**). The three-stone fruits i.e. loquat, plum, and apricot have less inter-cluster distance between them (see **Figure 6.2**) and also show similar absorbance behaviour in **Figure *6.1***. This is because all the three investigated stone fruits have similar physical structures i.e. thin peel, a thin pulp (and hence lesser NIR radiation absorbance by pulp) and a big stone/stones in the center of the fruit. Apple, mango, grapes, melon, and orange also have less inter-cluster distances amongst each other (**Figure 6.2**) and also it shows a similar trend in the absorbance spectra (see **Figure *6.1***) i.e. top five peaks at 960nm wavelength related to water absorption band. All these five fruits have thick

pulp and small seeds (the investigated grapes variety is ‘Sundar khani’ which has no seeds). Hass, cherry, and kiwi show good inter-cluster distances from all other clusters. Kiwi’s spectra form two clusters with a less inter-cluster distance between them, the reason is unknown as kiwi’s dataset was taken from an online source with no information available about the cultivars that were scanned.

Results obtained from PCA and classification in case 2 (90.38% test accuracy for 11 classes) i.e. only 4 features at wavelengths 770nm, 840nm, 910nm and 960nm (OH and CH overtone features) provide good motivation for application of low cost LED based devices in fruit classification problems. Spectrometer based portable devices are also available e.g. Felix instruments F-750 [8], Consumer Physics’ SciO [11], Sunforest H-100C [12] and Atago Hikari [13], however these devices are expensive compared to LED based portable devices for assessment of specific pigment in fruit (at two to five wavelengths) i.e. DA meter[14], kiwi meter [14], cherry meter [14], Multiplex 330[15] and FIORAMA [16].

Chapter 7 : CONCLUSION AND FUTURE RECOMMENDATIONS

7.1. Overview of Dissertation:

In this dissertation, I have investigated the SWNIR spectroscopy for non-destructive fruit quality inspection and present a decision support system for the Pakistani horticulture. The case studies involve fruits that are not only of commercial significance to Pakistani horticulture, but also provide examples of fruit with relatively thin and thick skin and with relatively thick and thin edible flesh. These differences in morphology can be expected to impact the non-invasive assessment of flesh characters using SWNIRS. This chapter summarizes the research contributions from chapter 3,4,5 and 6.

The main contributions of this dissertation are as follows:

1. I present a decision support system for quality-based fruit sorting and grading for Pakistani horticulture.
2. I present a novel decision support system to predict sweetness of melons and oranges which is based on direct classification approach.
3. I present SWNIR spectroscopy-based solution to fruit classification problem.

In chapter 2, I investigated the effectiveness of handheld near infrared spectroscopy (NIRS) for the evaluation of DM and SSC in four types of fruits of significance to Pakistani horticulture, i.e. mango, apple, mandarin and loquat. The fruits differed in the applicability of the NIRS technique: apples and mangoes as fruit with a thin skin and thick (> 1 cm) edible tissue, Kinnow mandarin because of their thick peel (>3 mm), and loquats because of their thin (<1 cm) edible tissue, with a large central seed. The investigated handheld NIRS device (i.e. Felix instruments F750) is supplied with ready to use starter models for many fruits. I investigated the performance robustness of supplied apple, mandarin and KP mango model in prediction of SSC and DM of local cultivars of these fruits. The starter models (i.e. mango and mandarin) which were based on a small population set of a single cultivar at a single temperature failed to effectively predict the

quality indices of local cultivars of mango and mandarin. Conversely, the supplied apple model which was based on a large population set of multiple cultivars at different temperatures performed well in use with locally grown fruit of the same cultivars as used in the calibration dataset. Hence, for mango and mandarin, I developed new chemometric models for local cultivars of mango (export culitvars : 'Sindhri', 'Sufaid Chaunsa' and 'SB Chaunsa') and mandarin (export cultivar : 'Kinnow'). I also investigated the feasibility of using SWNIRS for SSC prediction of loquat fruit. The locally developed mango, mandarin and loquat models performed well with the test data. The developed mango, mandarin and loquat models were trained with dataset collected at a single temperature (25^{o}C). To obtain a robust model against sample temperature variations (for on-tree predictions), the model needs to be trained with samples scanned at multiple temperatures. Based on these results, the portable NIR technology is recommended for future use across several commodities, given suitable model development based on samples from multiple cultivars with a range of growing conditions, and with assessment of samples at different temperatures. Use in quantitative assessment of fruit quality targeted to Pakistan's export markets is also recommended.

In chapter 4 and 5, I present a direct sweetness classifier for melons and oranges as opposed to the literature reported regression based indirect measure of the quality attributes like SSC. First extensive evaluation was conducted on "honey melon" variety, which is grown in Pakistan. The industry standard F-750 spectrometer employing interactance optical geometry was used to collect spectral data. After spectra collection, destructive testing was performed to find reference SSC values. Sweetness standards were established by sensory evaluation by a panel of judges. Extensive chemometric analysis was carried out to obtain SSC prediction model and direct sweetness classification model for SWNIRS. Outcomes of both these methods in terms of prediction accuracy validate that direct classification is a better quantitative measure as compared to ^{o}Brix based thresholding to estimate melons sweetness.

Second experiment was conducted on oranges to develop a correlation between quality indices i.e. Brix, TA, Brix:TA and BrimA, sensory assessment of the fruit and NIR spectra that was then classified as sweet, mixed, and acidic based on NIR spectra for Pakistani cultivars of orange i.e., Blood red, Mosambi and Succari. NIR spectral data was obtained using the F-750 fruit quality meter. Reference Brix and TA measurements were taken using standard destructive testing methods. Reference taste labels i.e. sweet, mix and acidic, were acquired by sensory evaluation of

samples by a panel of judges. I observed that Succari cultivar is statistically dissimilar from the other two cultivars w.r.t TA and Brix:TA. Hence, chemometric analysis was carried out on two datasets i.e. dataset1 (Blood red, Mosambi and Succari) and dataset 2 (Blood red and Mosambi samples only), to obtain prediction models for quality indices and sweetness classification model. Best fit partial least squares regression (PLSR) models for Brix and BrimA were achieved for dataset1. For TA and Brix:TA, best fit PLSR models were achieved with dataset2. I observed good correlation between NIR spectra and sensory assessment as opposed to quality indices. Hence, direct classification is more suitable for orange sweetness classification using NIR spectroscopy as opposed to estimation of quality indices. Considering orange fruit, along with sweetness, the consumer acceptance is usually also linked with richness and tartness. As in the case of Succari cultivar which is always sweet but not as tasty as Mosambi and Blood red cultivars. Our proposed sweetness classifier for orange is based on acceptance criteria linked only with sweetness of sample and not the tartness and richness parameters.

In chapter 6, I investigated the feasibility of using SWNIRS spectral features for automated classification of fruit type. Eleven fruits, which include apple, cherry, hass, kiwi, grapes, mango, melon, orange, loquat, plum, and apricot, were used in this study to cover physical characteristics such as peel thinness, pulp, seed thickness, and size. Different shallow machine learning architectures were trained to classify fruits using spectral feature vectors. The results obtained from PCA, and classification suggest that fruit classification is mainly a function of OH and CH overtone features. PCA results depict that both the feature sets, i.e., 83 features from the 725-975nm wavelength range and only 4 features representing the OH and CH overtones peaks within the 725-975nm range, form well-defined clusters in the PCA space. The PCA scores plot depicts a good intra-cluster correlation for all the eleven clusters in both cases. Results obtained from PCA and classification in case 2 (90.38% test accuracy for 11 classes) i.e. only 4 features at wavelengths 770nm, 840nm, 910nm and 960nm (OH and CH overtone features) provide good motivation for application of low cost LED based devices in fruit classification problems. Spectrometer based portable devices can also be used, however these devices are expensive compared to LED based portable devices for assessment of specific pigment in fruit (at two to five wavelengths).

7.2. Future Recommendations:

According to the reported literature and available chemometric models, PLSR has dominated applications of NIRS in horticulture. PLSR models perform well with small training datasets of a particular population and similar conditions. However, the performance of PLSR models is deteriorated for independent test sets of a different population having different growing conditions. In recent applications, ANN and convolutional neural networks (CNN) have shown promise in overcoming these issues. However, ANN and CNN models are more prone to overfitting with smaller datasets as compared to PLSR and MLR models, hence they require larger datasets for training. The ANNs implemented in literature have shallow architectures trained with carefully handcrafted features as input. As a practical example, the Australian mango industry has used ANN model based on 80000 mango spectra samples from 39 instruments, 16 cultivars and 591 populations, for mango DM estimation on a handheld device [183]. The first publication of CNN is reported in 2021 which employs a 1-D CNN and transfer learning approach on a publicly available dataset of 11,691 spectra for mango DM (harvested in four seasons) [229]. This publicly available dataset has been used in many comparative studies [230]. The advantage of using CNN over ANN seems to be the use of raw spectra at the input instead of carefully extracted features using pre-processing algorithms as in case of ANN. Also, standard machine learning algorithms tend to saturate once the dataset size increases beyond a certain limit, whereas deep learning models keeps on learning and updating based on new data and thus improving their performance. Hence, in future, efforts need to be done to collect larger datasets for training which include spectra of multiple cultivars over a range of growing and temperature conditions acquired via multiple instruments. ANN and CNN are the future in this area where extensive datasets may be available and comparative studies need to be done. It is to be noted here that in horticultural applications, collection of such extensive dataset requires years of effort, time and huge amount of funding, and also loss of large number of fruits.

The work presented in this dissertation is the first attempt to introduce non-destructive and efficient decision support system for fruit quality estimation for the Pakistani supply chain, both the upstream in the orchards and downstream in the packaging lines, distribution centres and at the consumer end. The models developed in this dissertation are based on shallow machine learning models like PLSR, KNN etc with smaller but reasonable dataset sizes (comparable to the PLSR

models reported in literature). In future, the research needs to be further continued for extension of available datasets and subsequent use of ANN and deep learning architectures like CNN to achieve robust chemometric models incorporating multiple cultivars data over multiple growing and temperature conditions which can better assist Pakistani supply chain.

REFERENCES

[1] "Agriculture Statistics | Pakistan Bureau of Statistics." https://www.pbs.gov.pk/content/agriculture-statistics (accessed Apr. 05, 2022).

[2] S. Sohaib Ali Shah *et al.*, "Towards fruit maturity estimation using NIR spectroscopy," *Infrared Phys. Technol.*, vol. 111, Dec. 2020, doi: 10.1016/J.INFRARED.2020.103479.

[3] S. Kumar, A. McGlone, C. Whitworth, and R. Volz, "Postharvest performance of apple phenotypes predicted by near-infrared (NIR) spectral analysis," *Postharvest Biol. Technol.*, vol. 100, pp. 16–22, Feb. 2015, doi: 10.1016/J.POSTHARVBIO.2014.09.021.

[4] R. Pourdarbani, S. Sabzi, M. H. Rohban, G. García-Mateos, J. Paliwal, and J. M. Molina-Martínez, "Using metaheuristic algorithms to improve the estimation of acidity in Fuji apples using NIR spectroscopy," *Ain Shams Eng. J.*, vol. 13, no. 6, p. 101776, Nov. 2022, doi: 10.1016/J.ASEJ.2022.101776.

[5] J. Li *et al.*, "Comparison and optimization of models for determination of sugar content in pear by portable Vis-NIR spectroscopy coupled with wavelength selection algorithm," *Springer*, vol. 12, no. 1, pp. 12–22, Jan. 2019, doi: 10.1007/s12161-018-1326-7.

[6] Z. Lu *et al.*, "Nondestructive Testing of Pear Based on Fourier Near-Infrared Spectroscopy," *Foods 2022, Vol. 11, Page 1076*, vol. 11, no. 8, p. 1076, Apr. 2022, doi: 10.3390/FOODS11081076.

[7] C. Watanawan, T. Wasusri, V. Srilaong, W.- Aree, and S. Kanlayanarat, "Near infrared spectroscopic evaluation of fruit maturity and quality of export Thai mango (Mangifera indica L. var. Namdokmai)," *Int. Food Res. J.*, vol. 21, no. 3, pp. 1109–1114, 2014, Accessed: Apr. 05, 2022. [Online]. Available: http://ifrj.upm.edu.my/21 (03) 2014/38 IFRJ 21 (03) 2014 Sirichai 671.pdf

[8] C. Liew and C. Lau, "Determination of quality parameters in Cavendish banana during ripening by NIR spectroscopy," *Int. Food Res. J.*, vol. 19, no. 2, pp. 751–758, 2012, Accessed: Apr. 05, 2022. [Online]. Available: http://www.agris.upm.edu.my:8080/dspace/handle/0/12127

[9] T. Sripaurya, K. Sengchuai, A. Booranawong, and K. Chetpattananondh, "Gros Michel banana soluble solids content evaluation and maturity classification using a developed portable 6 channel NIR device measurement," *Meas. J. Int. Meas. Confed.*, vol. 173, Mar. 2021, doi: 10.1016/J.MEASUREMENT.2020.108615.

[10] M. Li, D. Han, and W. Liu, "Non-destructive measurement of soluble solids content of three melon cultivars using portable visible/near infrared spectroscopy," *Biosyst. Eng.*, vol. 188, pp. 31–39, Dec. 2019, doi: 10.1016/J.BIOSYSTEMSENG.2019.10.003.

[11] S. Y. Kim, S. J. Hong, E. C. Kim, C. H. Lee, and G. Kim, "Neural Network Based Prediction of Soluble Solids Concentrationin Oriental Melon Using VIS/NIR Spectroscopy," *Appl. Eng. Agric.*, vol. 37, no. 4, pp. 653–663, 2021, doi: 10.13031/AEA.14332.

[12] F. Antonucci, F. Pallottino, G. Paglia, A. Palma, S. D'aquino, and P. Menesatti, "Non-destructive Estimation of Mandarin Maturity Status Through Portable VIS-NIR Spectrophotometer", doi: 10.1007/s11947-010-0414-5.

[13] C. Huang, J. Cai, Y. Zhou, H. R. El-Seedi, and Z. Guo, "Fusion models for detection of soluble solids content in mandarin by Vis/NIR transmission spectroscopy combined external factors," *Infrared Phys. Technol.*, vol. 124, p. 104233, Aug. 2022, doi: 10.1016/J.INFRARED.2022.104233.

[14] M. L. Amodio, F. Ceglie, M. M. A. Chaudhry, F. Piazzolla, and G. Colelli, "Potential of NIR spectroscopy for predicting internal quality and discriminating among strawberry fruits from different production systems," *Postharvest Biol. Technol.*, vol. 125, pp. 112–121, Mar. 2017, doi: 10.1016/J.POSTHARVBIO.2016.11.013.

[15] A. G. Saad, M. M. Azam, and B. M. A. Amer, "Quality Analysis Prediction and Discriminating Strawberry Maturity with a Hand-held Vis–NIR Spectrometer," *Food Anal. Methods*, vol. 15, no. 3, pp. 689–699, Mar.

2022, doi: 10.1007/S12161-021-02166-2/FIGURES/8.

[16] C. Camps and D. Christen, “Non-destructive assessment of apricot fruit quality by portable visible-near infrared spectroscopy,” *LWT - Food Sci. Technol.*, vol. 42, no. 6, pp. 1125–1131, Jul. 2009, doi: 10.1016/J.LWT.2009.01.015.

[17] M. B. Buyukcan and I. Kavdir, “Prediction of some internal quality parameters of apricot using FT-NIR spectroscopy,” *J. Food Meas. Charact.*, vol. 11, no. 2, pp. 651–659, Jun. 2017, doi: 10.1007/S11694-016-9434-9/FIGURES/7.

[18] G. Costa, C. Andreotti, O. Miserocchi, … M. N.-… S. on K., and U. 1999, “Near-infrared (NIR) methods to determine kiwifruit field harvest date and maturity parameters in cool store.” Accessed: Apr. 05, 2022. [Online]. Available: https://www.actahort.org/books/498/498_26.htm

[19] A. M. Afonso, M. D. Antunes, S. Cruz, A. M. Cavaco, and R. Guerra, “Non-destructive follow-up of ‘Jintao’ kiwifruit ripening through VIS-NIR spectroscopy – individual vs. average calibration model’s predictions,” *Postharvest Biol. Technol.*, vol. 188, p. 111895, Jun. 2022, doi: 10.1016/J.POSTHARVBIO.2022.111895.

[20] F. Visconti and J. M. de Paz, “Non-destructive assessment of chloride in persimmon leaves using a miniature visible near-infrared spectrometer,” *Comput. Electron. Agric.*, vol. 164, Sep. 2019, doi: 10.1016/J.COMPAG.2019.104894.

[21] R. Guidetti, R. Beghi, L. B.-T. of the ASABE, and U. 2010, “Evaluation of grape quality parameters by a simple Vis/NIR system,” *elibrary.asabe.org*, vol. 53, no. 2, pp. 477–484, 2010, Accessed: Apr. 05, 2022. [Online]. Available: https://elibrary.asabe.org/abstract.asp?aid=29556

[22] T. Basile, A. D. Marsico, and R. Perniola, “Use of Artificial Neural Networks and NIR Spectroscopy for Non-Destructive Grape Texture Prediction,” *Foods 2022, Vol. 11, Page 281*, vol. 11, no. 3, p. 281, Jan. 2022, doi: 10.3390/FOODS11030281.

[23] X. P. Fu *et al.*, “Determination of soluble solid content and acidity of loquats based on FT-NIR spectroscopy,” *J. Zhejiang Univ. Sci. B*, vol. 10, no. 2, pp. 120–125, Feb. 2009, doi: 10.1631/JZUS.B0820097.

[24] C. L. Y. Amuah *et al.*, “Feasibility Study of the Use of Handheld NIR Spectrometer for Simultaneous Authentication and Quantification of Quality Parameters in Intact Pineapple Fruits,” 2019, doi: 10.1155/2019/5975461.

[25] K. B. Walsh, J. Blasco, M. Zude-Sasse, and X. Sun, “Visible-NIR ‘point’ spectroscopy in postharvest fruit and vegetable assessment: The science behind three decades of commercial use,” *Postharvest Biol. Technol.*, vol. 168, p. 111246, Oct. 2020, doi: 10.1016/J.POSTHARVBIO.2020.111246.

[26] K. B. Walsh, V. A. McGlone, and D. H. Han, “The uses of near infra-red spectroscopy in postharvest decision support: A review,” *Postharvest Biol. Technol.*, vol. 163, May 2020, doi: 10.1016/J.POSTHARVBIO.2020.111139.

[27] “Australian Mango Industry Association.” https://www.industry.mangoes.net.au/ (accessed Apr. 06, 2022).

[28] B. A. J. G. Jacobs *et al.*, “Estimation of the prior storage period of lamb’s lettuce based on visible/near infrared reflectance spectroscopy,” *Postharvest Biol. Technol.*, vol. 113, pp. 95–105, Mar. 2016, doi: 10.1016/J.POSTHARVBIO.2015.11.007.

[29] K. Hameed, D. Chai, and A. Rassau, “A comprehensive review of fruit and vegetable classification,” *Image Vis. Comput.*, vol. 80, pp. 24–44, 2018, doi: 10.1016/j.imavis.2018.09.016.

[30] M. TUCERYAN and A. K. JAIN, “TEXTURE ANALYSIS,” *Handb. Pattern Recognit. Comput. Vis.*, pp. 235–276, Aug. 1993, doi: 10.1142/9789814343138_0010.

[31] A. Schmitz, Y. Bansho, K. Noda, … H. I.-2014 I.-R., and U. 2014, “Tactile object recognition using deep learning and dropout,” *2014 IEEE-RAS Int. Conf. Humanoid Robot.*, 2014, Accessed: Apr. 05, 2022. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7041493/

[32] H. Liu, Y. Yu, F. Sun, … J. G.-T. on A. S., and U. 2016, "Visual–tactile fusion for object recognition," *ieeexplore.ieee.org*, vol. 14, no. 2, pp. 996–1008, 2017, Accessed: Apr. 05, 2022. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7462208/

[33] H. Zhang, J. Wu, Z. Zhao, Z. W.-P. B. and Technology, and U. 2018, "Nondestructive firmness measurement of differently shaped pears with a dual-frequency index based on acoustic vibration," *Postharvest Biol. Technol.*, vol. 138, pp. 11–18, 2018, doi: 10.1016/j.postharvbio.2017.12.002.

[34] M. Lashgari and J. A. Journal, A Maleki, "Application of acoustic impulse response in discrimination of apple storage time using neural network.," *Int. Food Res. J.*, vol. 24, no. 3, pp. 1075–1080, 2017, Accessed: Apr. 05, 2022. [Online]. Available: http://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=19854668&asa=Y&AN=124235826&h=FxHVna%2FsbxzZBXiVrtAIf58ONMyiAhMTku6dStlwh9cuSK%2Fzj3ReDhOWbEqn3iXCzc39vVfIcFrauNNg4DSwQw%3D%3D&crl=c

[35] J. ; Foerster, M. ; Geyer, O. ; Schlüter, P. Fey, and M. Kiefer, "Acoustic resonance analysis for quality characterization of fruits and vegetables," *Crop. Mach.*, vol. 65, no. 2, pp. 96–98, 2010.

[36] A. Y. Groot, TH de, E Woudenberg, "Urban objects classification with an experimental acoustic sensor network," *IEEE Sens. J.*, vol. 15, no. 5, pp. 3068–3075, 2015, Accessed: Apr. 05, 2022. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7001237/

[37] A. Zeb *et al.*, "Is this melon sweet? A quantitative classification for near-infrared spectroscopy," *Infrared Phys. Technol.*, vol. 114, 2021, doi: 10.1016/j.infrared.2021.103645.

[38] A. Zeb, W. S. Qureshi, A. Ghafoor, and D. O. Sullivan, "Learning Fruit Class from Short Wave Near Infrared Spectral Features, an AI Approach Towards Determining Fruit Type," in *2022 8th International Conference on Mechatronics and Robotics Engineering (ICMRE)*, Mar. 2022, pp. 193–196. doi: 10.1109/ICMRE54455.2022.9734107.

[39] M. Zude-Sasse, I. Truppel, and B. Herold, "An approach to non-destructive apple fruit chlorophyll determination," *Postharvest Biol. Technol.*, vol. 25, no. 2, pp. 123–133, 2002, doi: 10.1016/S0925-5214(01)00173-9.

[40] S. Kawano, "Non-Destructive NIR Quality Evaluation of Fruits and Vegetables in Japan," *NIR news*, vol. 5, no. 6, pp. 10–12, Dec. 1994, doi: 10.1255/NIRN.278.

[41] M. Golic, K. Walsh, and P. Lawson, "Short-wavelength near-infrared spectra of sucrose, glucose, and fructose with respect to sugar concentration and temperature," *Appl. Spectrosc.*, vol. 57, no. 2, pp. 139–145, Feb. 2003, doi: 10.1366/000370203321535033.

[42] KH NORRIS, "Direct spectrophotometric determination of moisture content of grain and seeds," *Proc. 1963 Int. Symp.*, vol. 4, no. 19, 1965, Accessed: Apr. 05, 2022. [Online]. Available: https://ci.nii.ac.jp/naid/20001547711/

[43] G. Birth, GS, Dull, "Nondestructive spectrophotometric determination of dry matter in onions," *J. Am. Soc. Hortic. Sci.*, vol. 110, no. 2, pp. 297–303, 1985, Accessed: Apr. 06, 2022. [Online]. Available: https://www.researchgate.net/profile/Stanley-Kays/publication/284681634_Nondestructive_spectrophotometric_determination_of_dry_matter_in_onions/links/57460daa08ae9f741b43133d/Nondestructive-spectrophotometric-determination-of-dry-matter-in-onions.pdf

[44] G. Dull, G. Birth, D. Smittle, and L. RG, "Near infrared analysis of soluble solids in intact cantaloupe," *J. Food Sci.*, vol. 54, no. 2, pp. 393–395, 1989, Accessed: Apr. 06, 2022. [Online]. Available: https://agris.fao.org/agris-search/search.do?recordID=US8934755

[45] S. Kawano, H. Watanabe, and M. Iwamoto, "Determination of sugar content in intact peaches by near infrared spectroscopy with fiber optics in interactance mode," *J. Japanese Soc. Hortic. Sci.*, vol. 61, no. 2, pp. 445–451, 1992, Accessed: Apr. 06, 2022. [Online]. Available: https://www.jstage.jst.go.jp/article/jjshs1925/61/2/61_2_445/_article/-char/ja/

[46] J. Lammertyn, A. Peirs, and B. J. De, “Light penetration properties of NIR radiation in fruit with respect to non-destructive quality assessment,” *Postharvest Biol. Technol.*, vol. 18, pp. 121–132, 2000, Accessed: Apr. 06, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S092552149900071X

[47] U. Acharya, K. Walsh, and P. S. Infrared, “Robustness of partial least-squares models to change in sample temperature: I. A comparison of methods for sucrose in aqueous solution,” *J. near infrared Spectrosc.*, vol. 22, no. 4, pp. 279–286, 2014, Accessed: Apr. 06, 2022. [Online]. Available: https://www.osapublishing.org/abstract.cfm?uri=jnirs-22-4-279

[48] P. P. Subedi, K. B. Walsh, and G. Owens, “Prediction of mango eating quality at harvest using short-wave near infrared spectrometry,” *Postharvest Biol. Technol.*, vol. 43, no. 3, pp. 326–334, Mar. 2007, doi: 10.1016/J.POSTHARVBIO.2006.09.012.

[49] N. T. Anderson, P. P. Subedi, and K. B. Walsh, “Manipulation of mango fruit dry matter content to improve eating quality,” *Sci. Hortic. (Amsterdam).*, vol. 226, pp. 316–321, Dec. 2017, doi: 10.1016/J.SCIENTA.2017.09.001.

[50] P. Timkhum and A. Terdwongworakul, “Non-destructive classification of durian maturity of ‘Monthong’ cultivar by means of visible spectroscopy of the spine,” *J. Food Eng.*, vol. 112, no. 4, pp. 263–267, Oct. 2012, doi: 10.1016/J.JFOODENG.2012.05.018.

[51] V. A. McGlone, R. B. Jordan, and P. J. Martinsen, “Vis/NIR estimation at harvest of pre- and post-storage quality indices for ‘Royal Gala’ apple,” *Postharvest Biol. Technol.*, vol. 25, no. 2, pp. 135–144, 2002, doi: 10.1016/S0925-5214(01)00180-6.

[52] P. P. Subedi and K. B. Walsh, “Assessment of sugar and starch in intact banana and mango fruit by SWNIR spectroscopy,” *Postharvest Biol. Technol.*, vol. 62, no. 3, pp. 238–245, Dec. 2011, doi: 10.1016/J.POSTHARVBIO.2011.06.014.

[53] Z. Guo *et al.*, “Quantitative detection of apple watercore and soluble solids content by near infrared transmittance spectroscopy,” *J. Food Eng.*, vol. 279, p. 109955, Aug. 2020, doi: 10.1016/J.JFOODENG.2020.109955.

[54] C. Shenderey *et al.*, “NIRS Detection of Moldy Core in Apples,” *Food Bioprocess Technol.*, vol. 79, no. 3, 2010, doi: 10.1007/s11947-009-0256-1.

[55] B. P. Khatiwada, P. P. Subedi, C. Hayes, L. C. C. Carlos, and K. B. Walsh, “Assessment of internal flesh browning in intact apple using visible-short wave near infrared spectroscopy,” *Postharvest Biol. Technol.*, vol. 120, pp. 103–111, Oct. 2016, doi: 10.1016/J.POSTHARVBIO.2016.06.001.

[56] P. Subedi, K. Walsh, and H. DW, “Assessment of Titratable Acidity in Fruit Using Short Wave near Infrared Spectroscopy. Part B: Intact Fruit Studies,” *Journl near infrared Spectrosc.*, vol. 20, no. 4, pp. 459–463, 2012, Accessed: Apr. 06, 2022. [Online]. Available: https://opg.optica.org/jnirs/abstract.cfm?uri=jnirs-20-4-459

[57] P. Subedi, K. Walsh, and D. Hopkins, “Assessment of Titratable Acidity in Fruit Using Short Wave near Infrared Spectroscopy. Part A: Establishing a Detection Limit Based on Model Solutions,” *J. Near Infrared Spectrosc.*, vol. 20, no. 4, pp. 449–457, 2012, Accessed: Apr. 06, 2022. [Online]. Available: https://opg.optica.org/jnirs/abstract.cfm?uri=jnirs-20-4-449

[58] P. P. Subedi and K. B. Walsh, “Non-invasive techniques for measurement of fresh fruit firmness,” *Postharvest Biol. Technol.*, vol. 51, no. 3, pp. 297–304, Mar. 2009, doi: 10.1016/J.POSTHARVBIO.2008.03.004.

[59] T. Ignat, Z. Schmilovitch, J. Fefoldi, B. Steiner, and S. Alkalai-Tuvia, “Non-destructive measurement of ascorbic acid content in bell peppers by VIS-NIR and SWIR spectrometry,” *Postharvest Biol. Technol.*, vol. 74, pp. 91–99, Dec. 2012, doi: 10.1016/J.POSTHARVBIO.2012.06.010.

[60] K. Walsh, V. McGlone, D. H.-P. B. and Technology, and undefined 2020, “The uses of near infra-red spectroscopy in postharvest decision support: A review,” *Elsevier*, Accessed: Apr. 05, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925521419308129

[61] T. Ignat *et al.*, "Forecast of Apple Internal Quality Indices at Harvest and During Storage by VIS-NIR Spectroscopy," *Food Bioprocess Technol.*, vol. 7, no. 10, pp. 2951–2961, Sep. 2014, doi: 10.1007/S11947-014-1297-7/TABLES/4.

[62] A. M. Rady, S. Sugiharto, and A. A. Adedeji, "Evaluation of Carrot Quality Using Visible-Near Infrared Spectroscopy and Multivariate Analysis," *J. Food Res.*, vol. 7, no. 4, 2018, doi: 10.5539/jfr.v7n4p80.

[63] H. Xiao *et al.*, "Quality assessment and discrimination of intact white and red grapes from Vitis vinifera L. at five ripening stages by visible and near-infrared spectroscopy," *Sci. Hortic. (Amsterdam).*, vol. 233, pp. 99–107, Mar. 2018, doi: 10.1016/J.SCIENTA.2018.01.041.

[64] V. Cortés *et al.*, "Prediction of the level of astringency in persimmon using visible and near-infrared spectroscopy," *J. Food Eng.*, vol. 204, pp. 27–37, Jul. 2017, doi: 10.1016/J.JFOODENG.2017.02.017.

[65] L. Pan, R. Lu, Q. Zhu, J. M. McGrath, and K. Tu, "Measurement of moisture, soluble solids, sucrose content and mechanical properties in sugar beet using portable visible and near-infrared spectroscopy," *Postharvest Biol. Technol.*, vol. 102, pp. 42–50, Apr. 2015, doi: 10.1016/J.POSTHARVBIO.2015.02.005.

[66] S. Fan, B. Zhang, J. Li, C. Liu, W. Huang, and X. Tian, "Prediction of soluble solids content of apple using the combination of spectra and textural features of hyperspectral reflectance imaging data," *Postharvest Biol. Technol.*, vol. 121, pp. 51–61, Nov. 2016, doi: 10.1016/J.POSTHARVBIO.2016.07.007.

[67] G. Brown, L. Segerlind, and S. R, "Near-infrared reflectance of bruised apples," *Trans. ASAE*, vol. 17, no. 1, pp. 0017–0019, 1974, Accessed: Apr. 05, 2022. [Online]. Available: https://elibrary.asabe.org/abstract.asp?aid=36775

[68] W. Bilanski, C. Pen, and F. DR, "Apple bruise detection using optical reflectance parameters," *Can. Agric. Eng.*, vol. 26, no. 2, pp. 111–114, 2013, Accessed: Apr. 05, 2022. [Online]. Available: https://agris.fao.org/agris-search/search.do?recordID=US201302647854

[69] C. Pen, W. Bilanski, and D. F. The, "Classification analysis of good and bruised peeled apple tissue using optical reflectance," *Trans. ASAE*, vol. 28, no. 1, pp. 0326–0330, 1985, Accessed: Apr. 05, 2022. [Online]. Available: https://elibrary.asabe.org/abstract.asp?aid=32251

[70] E. Moons, G. Sinnaeve, and P. D. International, "Non destructive visible and NIR spectroscopy measurement for the determination of apple internal quality," *XXV Int. Hortic. Congr. Part 7 Qual. Hortic. Prod.*, vol. 517, pp. 441–448, 2000, Accessed: Apr. 05, 2022. [Online]. Available: https://www.actahort.org/books/517/517_56.htm

[71] J. Lammertyn, B. Nicolaï, K. Ooms, S. V De, and B. J. De, "Non-destructive measurement of acidity, soluble solids, and firmness of Jonagold apples using NIR-spectroscopy," *Trans. ASAE*, vol. 41, no. 4, pp. 1089–1094, 1998, Accessed: Apr. 05, 2022. [Online]. Available: https://elibrary.asabe.org/abstract.asp?aid=17238

[72] M. Zude-Sasse, "Comparison of indices and multivariate models to non-destructively predict the fruit chlorophyll by means of visible spectrometry in apple fruit," *Anal. Chim. Acta*, vol. 481, no. 1, pp. 119–126, 2003, doi: 10.1016/S0003-2670(03)00070-9.

[73] Y. Liu and Y. Ying, "Use of FT-NIR spectrometry in non-invasive measurements of internal quality of 'Fuji' apples," *Postharvest Biol. Technol.*, vol. 37, no. 1, pp. 65–71, Jul. 2005, doi: 10.1016/J.POSTHARVBIO.2005.02.013.

[74] Y. Liu, Y. Ying, H. Yu, and X. Fu, "Comparison of the HPLC method and FT-NIR analysis for quantification of glucose, fructose, and sucrose in intact apple fruits," *J. Agric. Food Chem.*, vol. 54, no. 8, pp. 2810–2815, Apr. 2006, doi: 10.1021/JF052889E.

[75] M. R. Mogollon, A. F. Jara, C. Contreras, and J. P. Zoffoli, "Quantitative and qualitative VIS-NIR models for early determination of internal browning in 'Cripps Pink' apples during cold storage," *Postharvest Biol. Technol.*, vol. 161, Mar. 2020, doi: 10.1016/J.POSTHARVBIO.2019.111060.

[76] K. Walsh and P. Subedi, "In-field monitoring of mango fruit dry matter for maturity estimation," *ISHS Acta*

Hortic., vol. 1119, pp. 273–278, 2016, Accessed: Apr. 05, 2022. [Online]. Available: https://www.actahort.org/books/1119/1119_38.htm

[77] J. Guthrie and K. Walsh, "Non-invasive assessment of pineapple and mango fruit quality using near infra-red spectroscopy," *Austrailian J. Exp. Agric.*, vol. 37, no. 2, pp. 253–263, 1997, doi: 10.1071/EA96026.

[78] W. Spreer and J. Müller, "Estimating the mass of mango fruit (Mangifera indica, cv. Chok Anan) from its geometric dimensions by optical measurement," *Comput. Electron. Agric.*, vol. 75, no. 1, pp. 125–131, Jan. 2011, doi: 10.1016/J.COMPAG.2010.10.007.

[79] M. Nagle, B. Mahayothee, P. Rungpichayapichet, S. Janjai, and J. Müller, "Effect of irrigation on near-infrared (NIR) based prediction of mango maturity," *Sci. Hortic. (Amsterdam).*, vol. 125, no. 4, pp. 771–774, Jul. 2010, doi: 10.1016/J.SCIENTA.2010.04.044.

[80] S. H. E. J. Gabriëls, P. Mishra, M. G. J. Mensink, P. Spoelstra, and E. J. Woltering, "Non-destructive measurement of internal browning in mangoes using visible and near-infrared spectroscopy supported by artificial neural network analysis," *Postharvest Biol. Technol.*, vol. 166, Aug. 2020, doi: 10.1016/J.POSTHARVBIO.2020.111206.

[81] B. Park, J. Abbott, K. Lee, C. Choi, and KH Choi, "Near-infrared diffuse reflectance for quantitative and qualitative measurement of soluble solids and firmness of Delicious and Gala apples," *Trans. ASAE*, vol. 46, no. 6, pp. 1721–1731, 2003, Accessed: Apr. 05, 2022. [Online]. Available: https://elibrary.asabe.org/abstract.asp?aid=15628

[82] A. Pissard, V. Baeten, P. Dardenne, P. Dupont, M. L.- BASE, and undefined 2018, "Use of NIR spectroscopy on fresh apples to determine the phenolic compounds and dry matter content in peel and flesh," *popups.uliege.be*, Accessed: Apr. 05, 2022. [Online]. Available: https://popups.uliege.be/1780-4507/index.php?id=16241

[83] Y. Wu, L. Li, L. Liu, and Y. Liu, "Nondestructive measurement of internal quality attributes of apple fruit by using NIR spectroscopy," *Multimed. Tools Appl.*, vol. 78, no. 4, pp. 4179–4195, Feb. 2019, doi: 10.1007/S11042-017-5388-0.

[84] Y. Zhang, J. F. Nock, Y. Al Shoffe, and C. B. Watkins, "Non-destructive prediction of soluble solids and dry matter contents in eight apple cultivars using near-infrared spectroscopy," *Postharvest Biol. Technol.*, vol. 151, pp. 111–118, May 2019, doi: 10.1016/J.POSTHARVBIO.2019.01.009.

[85] N. Nguyen Do Trong, C. Erkinbaev, M. Tsuta, J. De Baerdemaeker, B. Nicolaï, and W. Saeys, "Spatially resolved diffuse reflectance in the visible and near-infrared wavelength range for non-destructive quality assessment of 'Braeburn' apples," *Postharvest Biol. Technol.*, vol. 91, pp. 39–48, May 2014, doi: 10.1016/J.POSTHARVBIO.2013.12.004.

[86] W. Lan, B. Jaillais, A. Leca, C. M. G. C. Renard, and S. Bureau, "A new application of NIR spectroscopy to describe and predict purees quality from the non-destructive apple measurements," *Food Chem.*, vol. 310, Apr. 2020, doi: 10.1016/J.FOODCHEM.2019.125944.

[87] X. Xu, J. Mo, L. Xie, and Y. Ying, "Influences of Detection Position and Double Detection Regions on Determining Soluble Solids Content (SSC) for Apples Using On-line Visible/Near-Infrared (Vis/NIR) Spectroscopy", doi: 10.1007/s12161-019-01530-7.

[88] Z. Schmilovitch, A. Mizrach, A. Hoffman, H. Egozi, and Y. Fuchs, "Determination of mango physiological indices by near-infrared spectrometry," *Postharvest Biol. Technol.*, vol. 19, no. 3, pp. 245–252, Jul. 2000, doi: 10.1016/S0925-5214(00)00102-2.

[89] E. J. N. Marques, S. T. De Freitas, M. F. Pimentel, and C. Pasquini, "Rapid and non-destructive determination of quality parameters in the 'Tommy Atkins' mango using a novel handheld near infrared spectrometer," *Food Chem.*, vol. 197, pp. 1207–1214, Apr. 2016, doi: 10.1016/J.FOODCHEM.2015.11.080.

[90] S. Saranwong, J. Sornsrivichai, and S. Kawano, "Prediction of ripe-stage eating quality of mango fruit from its harvest quality measured nondestructively by near infrared spectroscopy," *Postharvest Biol. Technol.*, vol.

31, no. 2, pp. 137–145, 2004, doi: 10.1016/J.POSTHARVBIO.2003.08.007.

[91] Y. Q. Polinar, K. F. Yaptenco, E. K. Peralta, and J. U. Agravante, "Near-infrared spectroscopy for non-destructive prediction of maturity and eating quality of 'Carabao'mango (Mangifera indica L.) fruit," *VI-Postharvest Technol. Process Eng.*, vol. 21, no. 1, p. 209, 2019, Accessed: Apr. 05, 2022. [Online]. Available: http://cigrjournal.org/index.php/Ejounral/article/view/5152

[92] V. Cortés, C. Ortiz, N. Aleixos, J. Blasco, S. Cubero, and P. Talens, "A new internal quality index for mango and its prediction by external visible and near-infrared reflection spectroscopy," *Postharvest Biol. Technol.*, vol. 118, pp. 148–158, Aug. 2016, doi: 10.1016/J.POSTHARVBIO.2016.04.011.

[93] T. Nordey, J. Joas, F. Davrieux, M. Chillet, and M. Léchaudel, "Robust NIRS models for non-destructive prediction of mango internal quality," *Sci. Hortic. (Amsterdam).*, vol. 216, pp. 51–57, Feb. 2017, doi: 10.1016/J.SCIENTA.2016.12.023.

[94] D. G. Abdullah Al-Sanabani, M. I. Solihin, L. P. Pui, W. Astuti, C. K. Ang, and L. W. Hong, "Development of non-destructive mango assessment using Handheld Spectroscopy and Machine Learning Regression," *J. Phys. Conf. Ser.*, vol. 1367, no. 1, p. 012030, Nov. 2019, doi: 10.1088/1742-6596/1367/1/012030.

[95] A. Munawar, D. von Hörsten, J. Wegener, E. Pawelzik, and D. Morlein, "Rapid and non-destructive prediction of mango quality attributes using Fourier transform near infrared spectroscopy and chemometrics," *Eng. Agric. Environ. food*, vol. 9, no. 3, pp. 208–215, 2016, Accessed: Apr. 05, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1881836615300264

[96] J. P. dos Santos Neto, M. W. D. de Assis, I. P. Casagrande, L. C. Cunha Júnior, and G. H. de Almeida Teixeira, "Determination of 'Palmer' mango maturity indices using portable near infrared (VIS-NIR) spectrometer," *Postharvest Biol. Technol.*, vol. 130, pp. 75–80, Aug. 2017, doi: 10.1016/J.POSTHARVBIO.2017.03.009.

[97] S. Neto *et al.*, "Cold storage of 'Palmer'mangoes sorted based on dry matter content using portable near infrared (VIS-NIR) spectrometer," *Wiley Online Libr.*, vol. 42, no. 6, Jun. 2018, doi: 10.1111/jfpp.13644.

[98] R. Bahari, F. Khairul, K. Kasim, X. Sy, P. S. Woo, and M. F. Ibrahim, "pH Prediction of Perlis Sunshine Mango Using NIR Spectrometer," *iopscience.iop.org*, doi: 10.1088/1757-899X/705/1/012021.

[99] V. Jayasena and I. Cameron, "° Brix/acid ratio as a predictor of consumer acceptability of Crimson Seedless table grapes," *J. Food Qual.*, vol. 31, no. 6, pp. 736–750, Dec. 2008, doi: 10.1111/j.1745-4557.2008.00231.x.

[100] J. Fernández-Novales, M. I. López, M. T. Sánchez, J. A. García-Mesa, and V. González-Caballero, "Assessment of quality parameters in grapes during ripening using a miniature fiber-optic near-infrared spectrometer," *Int. J. Food Sci. Nutr.*, vol. 60, no. SUPPL. 7, pp. 265–277, Sep. 2009, doi: 10.1080/09637480903093116.

[101] A. Guesalaga, E. Agosin, J. Herrera, A. Guesalaga, and E. Agosin, "Shortwave–near infrared spectroscopy for non-destructive determination of maturity of wine grapes," *iopscience.iop.org*, vol. 14, pp. 689–697, 2003, doi: 10.1088/0957-0233/14/5/320.

[102] A. Guesalaga, E. Agosin, M. Larraín, A. R. Guesalaga, and E. Agosín, "A multipurpose portable instrument for determining ripeness in wine grapes using NIR spectroscopy," *ieeexplore.ieee.org*, vol. 57, no. 2, 2008, doi: 10.1109/TIM.2007.910098.

[103] F. Cao, D. Wu, and Y. He, "Soluble solids content and pH prediction and varieties discrimination of grapes based on visible-near infrared spectroscopy," *Comput. Electron. Agric.*, vol. 71, no. SUPPL. 1, pp. 15–18, Apr. 2010, doi: 10.1016/J.COMPAG.2009.05.011.

[104] A. F. Omar, "Spectroscopic profiling of soluble solids content and acidity of intact grape, lime, and star fruit," *Sens. Rev.*, vol. 33, no. 3, pp. 238–245, 2013, doi: 10.1108/02602281311324690/FULL/HTML.

[105] R. Ferrer-Gallego, J. Miguel Hernández-Hierro, J. C. Rivas-Gonzalo, and M. Teresa Escribano-Bailón, "Determination of phenolic compounds of grape skins during ripening by NIR spectroscopy," *Elsevier*, vol. 44, pp. 847–853, 2011, doi: 10.1016/j.lwt.2010.12.001.

[106] B. Kemps, L. Leon, S. Best, J. De Baerdemaeker, and B. De Ketelaere, "Assessment of the quality parameters in grapes using VIS/NIR spectroscopy," *Biosyst. Eng.*, vol. 105, no. 4, pp. 507–513, 2010, doi: 10.1016/J.BIOSYSTEMSENG.2010.02.002.

[107] J. Fernández-Novales, T. Garde-Cerdán, J. Tardáguila, G. Gutiérrez-Gamboa, E. P. Pérez-Álvarez, and M. P. Diago, "Assessment of amino acids and total soluble solids in intact grape berries using contactless Vis and NIR spectroscopy during ripening," *Talanta*, vol. 199, pp. 244–253, Jul. 2019, doi: 10.1016/J.TALANTA.2019.02.037.

[108] D. dos Santos Costa, N. F. Oliveros Mesa, M. Santos Freire, R. Pereira Ramos, and B. J. Teruel Mederos, "Development of predictive models for quality and maturation stage attributes of wine grapes using vis-nir reflectance spectroscopy," *Postharvest Biol. Technol.*, vol. 150, pp. 166–178, Apr. 2019, doi: 10.1016/J.POSTHARVBIO.2018.12.010.

[109] B. Baca-Bocanegra, J. M. Hernández-Hierro, J. Nogales-Bueno, and F. J. Heredia, "Feasibility study on the use of a portable micro near infrared spectroscopy device for the 'in vineyard' screening of extractable polyphenols in red grape skins," *Talanta*, vol. 192, pp. 353–359, Jan. 2019, doi: 10.1016/J.TALANTA.2018.09.057.

[110] P. M.-O. for N. Resources, U. Agriculture, U. And, and U. 2007, "In-field spectrophotometric measurement to estimate maturity stage of wine grapes," *Opt. Nat. Resour. Agric. Foods II*, vol. 6761, p. 676118, Sep. 2007, doi: 10.1117/12.735547.

[111] V. Geraudie *et al.*, "A revolutionary device for predicting grape maturity based on NIR spectrometry," *hal.archives-ouvertes.fr*, 2010, Accessed: Apr. 05, 2022. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00468877/

[112] A. J. Daniels, C. Poblete-Echeverría, U. L. Opara, and H. H. Nieuwoudt, "Measuring Internal Maturity Parameters Contactless on Intact Table Grape Bunches Using NIR Spectroscopy," *Front. Plant Sci.*, vol. 10, Nov. 2019, doi: 10.3389/FPLS.2019.01517/FULL.

[113] C. Kanchanomai, D. Naphrom, S. Ohashi, and N. Kazuhiro, "Nondestructive determination of quality management in table grapes using near infrared spectroscopy (NIRS) technique," *Int. J. Food Eng.*, vol. 5, no. 1, 2019, Accessed: Apr. 05, 2022. [Online]. Available: http://www.ijfe.org/uploadfile/2019/0322/20190322031304124.pdf

[114] F. Chauchard, R. Cogdill, S. Roussel, J. M. Roger, B. Maurel, and V. Bellon-Maurel, "Application of LS-SVM to non-linear phenomena in NIR spectroscopy: development of a robust and portable sensor for acidity prediction in grapes," *Elsevier*, vol. 71, no. 2, pp. 141–150, 2004, doi: 10.1016/j.chemolab.2004.01.003ï.

[115] J. Yu, H. Wang, X. Sun, and W. Huang, "Parameter optimization in soluble solid content prediction of entire bunches of grape based on near infrared spectroscopic technique," *J. Food Meas. Charact.*, vol. 11, no. 4, pp. 1676–1680, Dec. 2017, doi: 10.1007/S11694-017-9547-9.

[116] G. Ripoll, M. Vazquez, and M. Vilanova, "Ultraviolet–visible-near infrared spectroscopy for rapid determination of volatile compounds in white grapes during ripening," *Tech. Sci. vitiv*, vol. 32, no. 1, p. 53, 2017, doi: 10.1051/ctv/20173201053.

[117] J. Véstia, J. M. Barroso, H. Ferreira, L. Gaspar, and A. E. Rato, "Predicting calcium in grape must and base wine by FT-NIR spectroscopy," *Food Chem.*, vol. 276, pp. 71–76, Mar. 2019, doi: 10.1016/J.FOODCHEM.2018.09.116.

[118] J. Guthrie, B. Wedding, and K. Walsh, "Robustness of NIR calibrations for soluble solids in intact melon and pineapple," *J. Near Infrared Spectrosc.*, vol. 6, no. 1–4, pp. 259–265, 1998, doi: 10.1255/JNIRS.145.

[119] J. Sugiyama, "Visualization of sugar content in the flesh of a melon by near-infrared imaging," *J. Agric. Food Chem.*, vol. 47, no. 7, pp. 2715–2718, Jul. 1999, doi: 10.1021/JF981079I.

[120] C. Greensill, P. Wolfs, C. Spiegelman, and K. Walsh, "Calibration transfer between PDA-based NIR spectrometers in the NIR assessment of melon soluble solids content," *Appl. Spectrosc.*, vol. 55, no. 5, pp.

647–653, 2001, Accessed: Apr. 05, 2022. [Online]. Available: https://www.osapublishing.org/abstract.cfm?uri=as-55-5-647

[121] J. Guthrie, C. Liebenberg, and K. Walsh, "NIR model development and robustness in prediction of melon fruit total soluble solids," *Austrailian J. Agric. Res.*, vol. 57, pp. 1–8, 2006, doi: 10.1071/AR05123.

[122] R. L. Long, K. B. Walsh, R. L. Long, and K. B. Walsh, "Limitations to the measurement of intact melon total soluble solids using near infrared spectroscopy," *Aust. J. Agric. Res.*, vol. 57, no. 4, pp. 403–410, Apr. 2006, doi: 10.1071/AR05285.

[123] K. Flores, M. Sanchez, D. Perez-Marin, M. Lopez, J. Guerrero, and A. Garrido-Varo, "Prediction of total soluble solid content in intact and cut melons and watermelons using near infrared spectroscopy," *J. near infrared Spectrosc.*, vol. 16, no. 2, pp. 91–98, 2008, Accessed: Apr. 05, 2022. [Online]. Available: https://www.osapublishing.org/abstract.cfm?uri=jnirs-16-2-91

[124] S. R. Suh, K.-H. Lee, H. Yu, H. S. Shin, Y. S. Choi, and S. N. Yoo, "A melon fruit grading machine using a miniature VIS/NIR spectrometer: 1. Calibration models for the prediction of soluble solids content and firmness," *koreascience.or.kr*, vol. 37, no. 3, pp. 166–176, 2012, doi: 10.5307/JBE.2012.37.3.166.

[125] S. R. Suh, K.-H. Lee, H. Yu, H. S. Shin, S. N. Yoo, and Y. S. Choi, "A Melon Fruit Grading Machine Using a Miniature VIS/NIR Spectrometer: 2. Design Factors for Optimal Interactance Measurement Setup," *koreascience.or.kr*, vol. 37, no. 3, pp. 177–183, 2012, doi: 10.5307/JBE.2012.37.3.177.

[126] H. Tian, C. Wang, H. Zhang, Z. Yu, and J. Li, "Measurement of soluble solids content in melon by transmittance spectroscopy," *Sens. Lett.*, vol. 10, pp. 570–573, 2012, Accessed: Apr. 05, 2022. [Online]. Available: https://www.ingentaconnect.com/contentone/asp/senlet/2012/00000010/F0020001/art00083

[127] M. T. Sánchez, I. Torres, M. J. De la Haba, and D. Pérez-Marín, "First steps to predicting pulp colour in whole melons using near-infrared reflectance spectroscopy," *Biosyst. Eng.*, vol. 123, pp. 12–18, 2014, doi: 10.1016/J.BIOSYSTEMSENG.2014.04.010.

[128] L. Khurnpoon and P. Sirisomboon, "Rapid evaluation of the texture properties of melon (Cucumis melo L. Var. reticulata cv. Green net) using near infrared spectroscopy," *J. Texture Stud.*, vol. 49, no. 4, pp. 387–394, Aug. 2018, doi: 10.1111/JTXS.12329.

[129] R. Hu, L. Zhang, Z. Yu, Z. Zhai, and R. Zhang, "Optimization of soluble solids content prediction models in 'Hami' melons by means of Vis-NIR spectroscopy and chemometric tools," *Infrared Phys. Technol.*, vol. 102, Nov. 2019, doi: 10.1016/J.INFRARED.2019.102999.

[130] D. Zhang, L. Xu, Q. Wang, X. Tian, J. L.-F. A. Methods, and undefined 2019, "The optimal local model selection for robust and fast evaluation of soluble solid content in melon with thick peel and large size by vis-NIR spectroscopy," *Springer*, vol. 12, no. 1, pp. 136–147, Jan. 2019, doi: 10.1007/s12161-018-1346-3.

[131] J. Lu *et al.*, "Nondestructive determination of soluble solids and firmness in mix-cultivar melon using near-infrared CCD spectroscopy," *J. Innov. Opt. Health Sci.*, vol. 8, no. 6, Nov. 2015, doi: 10.1142/S1793545815500327.

[132] M. Li, R. R. Pullanagari, T. Pranamornkith, I. J. Yule, and A. R. East, "Quantitative prediction of post storage 'Hayward' kiwifruit attributes using at harvest Vis-NIR spectroscopy," *J. Food Eng.*, vol. 202, pp. 46–55, Jun. 2017, doi: 10.1016/J.JFOODENG.2017.01.002.

[133] Y. Uwadaira, Y. Sekiyama, and A. Ikehata, "An examination of the principle of non-destructive flesh firmness measurement of peach fruit by using VIS-NIR spectroscopy," *Heliyon*, vol. 4, no. 2, Feb. 2018, doi: 10.1016/J.HELIYON.2018.E00531.

[134] W. Guo, F. Zhao, and J. Dong, "Nondestructive Measurement of Soluble Solids Content of Kiwifruits Using Near-Infrared Hyperspectral Imaging," *Food Anal. Methods*, vol. 9, no. 1, pp. 38–47, Jan. 2016, doi: 10.1007/S12161-015-0165-Z.

[135] J. Li, H. Zhang, B. Zhan, Y. Zhang, R. Li, and J. Li, "Nondestructive firmness measurement of the multiple

cultivars of pears by Vis-NIR spectroscopy coupled with multivariate calibration analysis and MC-UVE-SPA method," *Infrared Phys. Technol.*, vol. 104, Jan. 2020, doi: 10.1016/J.INFRARED.2019.103154.

[136] V. Lafuente, L. J. Herrera, M. del M. Pérez, J. Val, and I. Negueruela, "Firmness prediction in Prunus persica 'Calrico' peaches by visible/short-wave near infrared spectroscopy and acoustic measurements using optimised linear and non-linear chemometric models," *J. Sci. Food Agric.*, vol. 95, no. 10, pp. 2033–2040, Aug. 2015, doi: 10.1002/JSFA.6916.

[137] M. T. Sánchez, M. J. De la Haba, and D. Pérez-Marín, "Internal and external quality assessment of mandarins on-tree and at harvest using a portable NIR spectrophotometer," *Comput. Electron. Agric.*, vol. 92, pp. 66–74, 2013, doi: 10.1016/J.COMPAG.2013.01.004.

[138] O. O. Olarewaju, I. Bertling, and L. S. Magwaza, "Non-destructive evaluation of avocado fruit maturity using near infrared spectroscopy and PLS regression models," *Sci. Hortic. (Amsterdam).*, vol. 199, pp. 229–236, Feb. 2016, doi: 10.1016/J.SCIENTA.2015.12.047.

[139] J. Wang, J. Wang, Z. Chen, and D. Han, "Development of multi-cultivar models for predicting the soluble solid content and firmness of European pear (Pyrus communis L.) using portable vis–NIR spectroscopy," *Postharvest Biol. Technol.*, vol. 129, pp. 143–151, Jul. 2017, doi: 10.1016/J.POSTHARVBIO.2017.03.012.

[140] P. R. Santagapita, U. Tylewicz, V. Panarese, P. Rocculi, and M. Dalla Rosa, "Non-destructive assessment of kiwifruit physico-chemical parameters to optimise the osmotic dehydration process: A study on FT-NIR spectroscopy," *Biosyst. Eng.*, vol. 142, pp. 101–109, Feb. 2016, doi: 10.1016/J.BIOSYSTEMSENG.2015.12.011.

[141] P. Andrade *et al.*, "Robust PLS models for soluble solids content and firmness determination in low chilling peach using near-infrared spectroscopy (NIR)," *Elsevier*, vol. 111, pp. 345–351, 2016, doi: 10.1016/j.postharvbio.2015.08.006.

[142] M. Amodio, F. Piazzolla, F. Colantuono, and G. Colelli, "The use of rapid FT-NIR methods to predict soluble solids, pH, titratable acidity and phenols of clingstone peaches ('Baby Gold 9')," *ISHS Acta Hortic. 1194*, pp. 1111–1118, 2018, Accessed: Apr. 05, 2022. [Online]. Available: https://www.actahort.org/books/1194/1194_159.htm

[143] L. M. Yuan, L. Sun, J. R. Cai, and H. Lin, "A Preliminary Study on Whether the Soluble Solid Content and Acidity of Oranges Predicted by Near Infrared Spectroscopy Meet the Sensory Degustation," *J. Food Process Eng.*, vol. 38, no. 4, pp. 309–319, Aug. 2015, doi: 10.1111/JFPE.12104.

[144] L. M. Yuan, F. Mao, X. Chen, L. Li, and G. Huang, "Non-invasive measurements of 'Yunhe' pears by vis-NIRS technology coupled with deviation fusion modeling approach," *Postharvest Biol. Technol.*, vol. 160, Feb. 2020, doi: 10.1016/J.POSTHARVBIO.2019.111067.

[145] J. G. Kim *et al.*, "Application of NIR-Spectroscopy to predict the harvesting maturity, fruit ripening and storage ability of Ca-chitosan treated baby kiwifruit," *academicjournals.org*, vol. 9, no. 4, pp. 44–53, 2018, doi: 10.5897/JSPPR2018.0257.

[146] S. Xudong, Z. Hailiang, and L. Yande, "Nondestructive assessment of quality of Nanfeng mandarin fruit by a portable near infrared spectroscopy," *Int. J. Agric. Biol. Eng.*, vol. 2, no. 1, 2009, doi: 10.3965/j.issn.1934-6344.2009.01.065-071.

[147] P. P. Subedi and K. B. Walsh, "Assessment of avocado fruit dry matter content using portable near infrared spectroscopy: Method and instrumentation optimisation," *Postharvest Biol. Technol.*, vol. 161, Mar. 2020, doi: 10.1016/J.POSTHARVBIO.2019.111078.

[148] Y. Shao, Y. He, Y. Bao, and J. Mao, "Near-infrared spectroscopy for classification of oranges and prediction of the sugar content," *Int. J. food Prop.*, vol. 12, no. 3, pp. 644–658, Jul. 2009, doi: 10.1080/10942910801992991.

[149] K. Miyamoto, M. Kawauchi, and T. Fukuda, "Classification of high acid fruits by partial least squares using the near infrared transmittance spectra of intact satsuma mandarins," *J. near infrared Spectrosc.*, vol. 6, no. 1,

pp. 267–271, 1998, Accessed: Apr. 05, 2022. [Online]. Available: https://www.osapublishing.org/abstract.cfm?uri=jnirs-6-1-267

[150] J. Song, G. Li, X. Yang, X. Liu, and L. Xie, "Rapid analysis of soluble solid content in navel orange based on visible-near infrared spectroscopy combined with a swarm intelligence optimization method," *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, vol. 228, Mar. 2020, doi: 10.1016/J.SAA.2019.117815.

[151] J. L. Aleixandre-Tudó, L. Castelló-Cogollos, J. L. Aleixandre, and R. Aleixandre-Benavent, "Bibliometric insights into the spectroscopy research field: A food science and technology case study," *Appl. Spectrosc. Rev.*, vol. 55, no. 9, pp. 873–906, Oct. 2019, doi: 10.1080/05704928.2019.1694936.

[152] A. Peirs, J. Tirry, B. Verlinden, P. Darius, and B. M. Nicolaï, "Effect of biological variability on the robustness of NIR models for soluble solids content of apples," *Postharvest Biol. Technol.*, vol. 28, no. 2, pp. 269–280, May 2003, doi: 10.1016/S0925-5214(02)00196-5.

[153] P. N. Schaare and D. G. Fraser, "Comparison of reflectance, interactance and transmission modes of visible-near infrared spectroscopy for measuring internal properties of kiwifruit (Actinidia chinensis)," *Postharvest Biol. Technol.*, vol. 20, no. 2, pp. 175–184, 2000, doi: 10.1016/S0925-5214(00)00130-7.

[154] D. G. Fraser, R. B. Jordan, R. Künnemeyer, and V. A. McGlone, "Light distribution inside mandarin fruit during internal quality assessment by NIR spectroscopy," *Postharvest Biol. Technol.*, vol. 27, no. 2, pp. 185–196, Feb. 2003, doi: 10.1016/S0925-5214(02)00058-3.

[155] X. Sun, Y. Liu, Y. Li, M. Wu, and D. Zhu, "Simultaneous measurement of brown core and soluble solids content in pear by on-line visible and near infrared spectroscopy," *Postharvest Biol. Technol.*, vol. 116, pp. 80–87, Jun. 2016, doi: 10.1016/J.POSTHARVBIO.2016.01.009.

[156] D. Lorente, P. Escandell-Montero, S. Cubero, J. Gómez-Sanchis, and J. Blasco, "Visible-NIR reflectance spectroscopy and manifold learning methods applied to the detection of fungal infections on citrus fruit," *J. Food Eng.*, vol. 163, pp. 17–24, Oct. 2015, doi: 10.1016/J.JFOODENG.2015.04.010.

[157] L. Salguero-Chaparro and F. Peña-Rodríguez, "On-line versus off-line NIRS analysis of intact olives," *LWT - Food Sci. Technol.*, vol. 56, no. 2, pp. 363–369, May 2014, doi: 10.1016/J.LWT.2013.11.032.

[158] H. Yan, H. S.-N. news, and undefined 2018, "Hand-held near-infrared spectrometers: State-of-the-art instrumentation and practical applications," *journals.sagepub.com*, vol. 29, no. 7, pp. 8–12, Nov. 2018, doi: 10.1177/0960336018796391.

[159] "F-750 Produce Quality Meter | Tools for Applied Food Science | felixinstruments.com." https://felixinstruments.com/food-science-instruments/portable-nir-analyzers/f-750-produce-quality-meter/ (accessed Apr. 05, 2022).

[160] A. Scalisi and M. G. O'Connell, "Application of visible/NIR spectroscopy for the estimation of soluble solids, dry matter and flesh firmness in stone fruits," *J. Sci. Food Agric.*, vol. 101, no. 5, pp. 2100–2107, Mar. 2021, doi: 10.1002/JSFA.10832.

[161] K. Ncama, L. S. Magwaza, S. Z. Tesfay, A. Mditshwa, and N. Mbili, "In-situ assessment of harvest maturity of 'Hass' avocado (Persea americana) using portable Vis-NIR spectrometer," *Acta Hortic.*, vol. 1299, pp. 339–345, Dec. 2020, doi: 10.17660/ACTAHORTIC.2020.1299.51.

[162] I. R. Donis-González, C. Valero, M. A. Momin, A. Kaur, and D. C. Slaughter, "Performance Evaluation of Two Commercially Available Portable Spectrometers to Non-Invasively Determine Table Grape and Peach Quality Attributes," *Agron. 2020, Vol. 10, Page 148*, vol. 10, no. 1, p. 148, Jan. 2020, doi: 10.3390/AGRONOMY10010148.

[163] K. Ncama, L. S. Magwaza, C. A. Poblete-Echeverría, H. H. Nieuwoudt, S. Z. Tesfay, and A. Mditshwa, "On-tree indexing of 'Hass'' avocado fruit by non-destructive assessment of pulp dry matter and oil content,'" *Biosyst. Eng.*, vol. 174, pp. 41–49, Oct. 2018, doi: 10.1016/J.BIOSYSTEMSENG.2018.06.011.

[164] P. F. Ndlovu, L. S. Magwaza, S. Z. Tesfay, and R. R. Mphahlele, "Rapid visible–near infrared (Vis–NIR)

spectroscopic detection and quantification of unripe banana flour adulteration with wheat flour," *J. Food Sci. Technol.*, vol. 56, no. 12, pp. 5484–5491, Dec. 2019, doi: 10.1007/S13197-019-04020-0/FIGURES/4.

[165] P. Mishra, E. Woltering, and N. El Harchioui, "Improved prediction of 'Kent' mango firmness during ripening by near-infrared spectroscopy supported by interval partial least square regression," *Infrared Phys. Technol.*, vol. 110, Nov. 2020, doi: 10.1016/J.INFRARED.2020.103459.

[166] A. Goke, S. Serra, and S. Musacchi, "Postharvest Dry Matter and Soluble Solids Content Prediction in d'Anjou and Bartlett Pear Using Near-infrared Spectroscopy," *HortScience*, vol. 53, no. 5, pp. 669–680, May 2018, doi: 10.21273/HORTSCI12843-17.

[167] S. Escribano, W. V. Biasi, R. Lerud, D. C. Slaughter, and E. J. Mitcham, "Non-destructive prediction of soluble solids and dry matter content using NIR spectroscopy and its relationship with sensory quality in sweet cherries," *Postharvest Biol. Technol.*, vol. 128, pp. 112–120, Jun. 2017, doi: 10.1016/J.POSTHARVBIO.2017.01.016.

[168] P. M. A. Toivonen, A. Batista, and B. Lannard, "Development of a predictive model for 'lapins' sweet cherry dry matter content using a visible/near-infrared spectrometer and its potential application to other cultivars," *Can. J. Plant Sci.*, vol. 97, no. 6, pp. 1030–1035, May 2017, doi: 10.1139/CJPS-2017-0013/ASSET/IMAGES/LARGE/CJPS-2017-0013F2.JPEG.

[169] S. Fan *et al.*, "Non-destructive evaluation of soluble solids content of apples using a developed portable Vis/NIR device," *Biosyst. Eng.*, vol. 193, pp. 138–148, May 2020, doi: 10.1016/J.BIOSYSTEMSENG.2020.02.017.

[170] C. Bonomelli, R. Mogollón, S. T. de Freitas, J. P. Zoffoli, and C. Contreras, "Nutritional Relationships in Bitter Pit-Affected Fruit and the Feasibility of Vis-NIR Models to Determine Calcium Concentration in 'Fuji' Apples," *Agron. 2020, Vol. 10, Page 1476*, vol. 10, no. 10, p. 1476, Sep. 2020, doi: 10.3390/AGRONOMY10101476.

[171] R. Mogollón, C. Contreras, M. L. da Silva Neta, E. J. N. Marques, J. P. Zoffoli, and S. T. de Freitas, "Non-destructive prediction and detection of internal physiological disorders in 'Keitt' mango using a hand-held Vis-NIR spectrometer," *Postharvest Biol. Technol.*, vol. 167, p. 111251, Sep. 2020, doi: 10.1016/J.POSTHARVBIO.2020.111251.

[172] A. M. Alhamdan and A. Atia, "Non-destructive method to predict Barhi dates quality at different stages of maturity utilising near-infrared (NIR) spectroscopy," *https://doi.org/10.1080/10942912.2017.1387794*, vol. 20, pp. S2950–S2959, Jan. 2018, doi: 10.1080/10942912.2017.1387794.

[173] A. M. Alhamdan, A. Fickak, and A. R. Atia, "Evaluation of sensory and texture profile analysis properties of stored Khalal Barhi dates nondestructively using Vis/NIR spectroscopy," *J. Food Process Eng.*, vol. 42, no. 6, p. e13215, Oct. 2019, doi: 10.1111/JFPE.13215.

[174] "DA Meter." https://www.trturoni.com/en/fruit-veg-ripeness-quality-control/da-meter (accessed Apr. 07, 2022).

[175] "Home | itk - Predict and Decide." https://www.itk.fr/en/ (accessed Apr. 07, 2022).

[176] "CP - Our Products." http://cp-info.de/Our-Products/ (accessed Apr. 07, 2022).

[177] N. Anderson and K. Walsh, "The evolution of chemometrics coupled with near infrared spectroscopy for fruit quality evaluation," *J. near infrared Spectrosc.*, vol. 0, no. 0, pp. 1–15, 2022, Accessed: Apr. 05, 2022. [Online]. Available: https://opg.optica.org/abstract.cfm?uri=jnirs-30-1-3

[178] F. Bexiga *et al.*, "A TSS classification study of 'Rocha' pear (Pyrus communis L.) based on non-invasive visible/near infra-red reflectance spectra," *Postharvest Biol. Technol.*, vol. 132, pp. 23–30, Oct. 2017, doi: 10.1016/J.POSTHARVBIO.2017.05.014.

[179] W. Saeys, N. Nguyen Do Trong, R. Van Beers, and B. M. Nicolaï, "Multivariate calibration of spectroscopic sensors for postharvest quality evaluation: A review," *Postharvest Biol. Technol.*, vol. 158, Dec. 2019, doi:

10.1016/J.POSTHARVBIO.2019.110981.

[180] H. Wang, J. Peng, C. Xie, Y. Bao, and Y. He, "Fruit Quality Evaluation Using Spectroscopy Technology: A Review," *Sensors 2015, Vol. 15, Pages 11889-11927*, vol. 15, no. 5, pp. 11889–11927, May 2015, doi: 10.3390/S150511889.

[181] B. M. Nicolaï *et al.*, "Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review," *Postharvest Biol. Technol.*, vol. 46, no. 2, pp. 99–118, Nov. 2007, doi: 10.1016/J.POSTHARVBIO.2007.06.024.

[182] "Zeaiter, M., & Rutledge, D. (2009). Preprocessing... - Google Scholar." https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Zeaiter%2C+M.%2C+%26+Rutledge%2C+D.+%282009%29.+Preprocessing+Methods.+Comprehensive+Chemometrics.&btnG= (accessed Apr. 05, 2022).

[183] N. Anderson and K. Walsh, "Review: The evolution of chemometrics coupled with near infrared spectroscopy for fruit quality evaluation," *Journal of Near Infrared Spectroscopy*, 2022. https://opg.optica.org/jnirs/abstract.cfm?uri=jnirs-30-1-3 (accessed Apr. 19, 2022).

[184] A. Savitzky and M. J. E. Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures," *Anal. Chem.*, vol. 36, no. 8, pp. 1627–1639, Jul. 1964, doi: 10.1021/AC60214A047.

[185] M. Batta, "No Title," *Int. J. Sci. Res.*, vol. 9, pp. 381–386, 2020, [Online]. Available: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.researchgate.net/profile/Batta-Mahesh/publication/344717762_Machine_Learning_Algorithms_-A_Review/links/5f8b2365299bf1b53e2d243a/Machine-Learning-Algorithms-A-Review.pdf?eid=5082902844932096

[186] M. I. Mukarev and K. B. Walsh, "Prediction of Brix Values of Intact Peaches with Least Squares-Support Vector Machine Regression Models:," *http://dx.doi.org/10.1255/jnirs.1026*, vol. 20, no. 6, pp. 647–655, Jan. 2012, doi: 10.1255/JNIRS.1026.

[187] A. Parmar, R. Katariya, and V. Patel, "A Review on Random Forest: An Ensemble Classifier," *Lect. Notes Data Eng. Commun. Technol.*, vol. 26, pp. 758–763, 2019, doi: 10.1007/978-3-030-03146-6_86/COVER.

[188] R. Panigrahi *et al.*, "Performance Assessment of Supervised Classifiers for Designing Intrusion Detection Systems: A Comprehensive Review and Recommendations for Future Research," *Math. 2021, Vol. 9, Page 690*, vol. 9, no. 6, p. 690, Mar. 2021, doi: 10.3390/MATH9060690.

[189] A. J. Izenman, "Linear Discriminant Analysis," pp. 237–280, 2013, doi: 10.1007/978-0-387-78189-1_8.

[190] S. Suthaharan, "Support Vector Machine," pp. 207–235, 2016, doi: 10.1007/978-1-4899-7641-3_9.

[191] R. Lu, R. Van Beers, W. Saeys, C. Li, and H. Cen, "Measurement of optical properties of fruits and vegetables: A review," *Postharvest Biol. Technol.*, vol. 159, Jan. 2020, doi: 10.1016/J.POSTHARVBIO.2019.111003.

[192] X. Sun, P. Subedi, and K. B. Walsh, "Achieving robustness to temperature change of a NIRS-PLSR model for intact mango fruit dry matter content," *Postharvest Biol. Technol.*, vol. 162, p. 111117, Apr. 2020, doi: 10.1016/J.POSTHARVBIO.2019.111117.

[193] S. Anderson *et al.*, "Determination of Fat,Moisture, and Protein in Meat and Meat Products by Using the FOSS FoodScan Near-Infrared Spectrophotometer with FOSS Artificial Neural Network Calibration Model and Associated Database: Collaborative Study," *J. AOAC Int.*, vol. 90, no. 4, pp. 1073–1083, Jul. 2007, doi: 10.1093/JAOAC/90.4.1073.

[194] J. A. Fernández Pierna and P. Dardenne, "Soil parameter quantification by NIRS as a Chemometric challenge at 'Chimiométrie 2006,'" *Chemom. Intell. Lab. Syst.*, vol. 91, no. 1, pp. 94–98, Mar. 2008, doi: 10.1016/J.CHEMOLAB.2007.06.007.

[195] R. J. Blakey, "Evaluation of avocado fruit maturity with a portable near-infrared spectrometer," *Postharvest Biol. Technol.*, vol. 121, pp. 101–105, Nov. 2016, doi: 10.1016/J.POSTHARVBIO.2016.06.016.

[196] P. Williams, P. Dardenne, and P. Flinn, “Tutorial: Items to be included in a report on a near infrared spectroscopy project:,” *J. near infrared Spectrosc.*, vol. 25, no. 2, pp. 85–90, Apr. 2017, doi: 10.1177/0967033517702395.

[197] K. Ncama, U. L. Opara, S. Z. Tesfay, O. A. Fawole, and L. S. Magwaza, “Application of Vis/NIR spectroscopy for predicting sweetness and flavour parameters of ‘Valencia’ orange (Citrus sinensis) and ‘Star Ruby’ grapefruit (Citrus x paradisi Macfad),” *J. Food Eng.*, vol. 193, pp. 86–94, Jan. 2017, doi: 10.1016/J.JFOODENG.2016.08.015.

[198] J. C. Tewari, V. Dixit, B. K. Cho, and K. A. Malik, “Determination of origin and sugars of citrus fruits using genetic algorithm, correspondence analysis and partial least square combined with fiber optic NIR spectroscopy,” *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.*, vol. 71, no. 3, pp. 1119–1127, Dec. 2008, doi: 10.1016/J.SAA.2008.03.005.

[199] A. H. Gómez, Y. He, and A. G. Pereira, “Non-destructive measurement of acidity, soluble solids and firmness of Satsuma mandarin using Vis/NIR-spectroscopy techniques,” *J. Food Eng.*, vol. 77, no. 2, pp. 313–319, Nov. 2006, doi: 10.1016/J.JFOODENG.2005.06.036.

[200] “SCiO - The World’s Only Pocket-Sized NIR Micro Spectrometer.” https://www.consumerphysics.com/ (accessed Apr. 05, 2022).

[201] “Sunforest.” http://www.sunforest.kr/category_main.php?sm_idx=168 (accessed Apr. 05, 2022).

[202] “Digital Hand-held ‘Pocket’ IR Brix Meter.” http://atago.net/product/?l=ue&f=products_pal_hikari.html (accessed Apr. 05, 2022).

[203] I. S. Minas, F. Blanco-Cipollone, and D. Sterle, “Accurate non-destructive prediction of peach fruit internal quality and physiological maturity with a single scan using near infrared spectroscopy,” *Food Chem.*, vol. 335, Jan. 2021, doi: 10.1016/J.FOODCHEM.2020.127626.

[204] A. Scalisi and M. G. O’Connell, “Application of visible/NIR spectroscopy for the estimation of soluble solids, dry matter and flesh firmness in stone fruits,” *J. Sci. Food Agric.*, vol. 101, no. 5, pp. 2100–2107, Mar. 2021, doi: 10.1002/JSFA.10832.

[205] I. Jan *et al.*, “RESPONSE OF APPLE CULTIVARS TO DIFFERENT STORAGE DURATIONS Prime Minister Youth Skill Development Program Phase 3-Batch 2 (National Vocational & Technical Training Commission (NAVTTC), Pakistan) View project Tissue culture of medicinal/horticultural plants View project RESPONSE OF APPLE CULTIVARS TO DIFFERENT STORAGE DURATIONS,” vol. 28, no. 2, 2012, Accessed: Nov. 01, 2022. [Online]. Available: https://www.researchgate.net/publication/230824022

[206] “Sugar composition of apple cultivars and its rel… — Library of Science.” https://bibliotekanauki.pl/articles/828070 (accessed Nov. 02, 2022).

[207] Amin and Muhammad, “Integrated approaches for improving fruit quality and shelflife of two commercial mango cultivars of Pakistan.,” 2012, Accessed: Nov. 02, 2022. [Online]. Available: http://prr.hec.gov.pk/jspui/handle/123456789/13987

[208] M. Amin, A. U. Malik, M. S. Khalid, and R. Anwar, “Fruit harvest maturity indicators for mango cultivars ‘Sindhri’ and ‘Samar Bahisht Chaunsa,’” *Acta Hortic.*, vol. 992, pp. 561–568, 2013, doi: 10.17660/ACTAHORTIC.2013.992.69.

[209] “Loquat in Pakistan.” https://www.slideshare.net/MuhammadRiaz89/loquat-in-pakistan (accessed Nov. 02, 2022).

[210] D. Cozzolino *et al.*, “Effect of temperature variation on the visible and near infrared spectra of wine and the consequences on the partial least square calibrations developed to measure chemical composition,” *Anal. Chim. Acta*, vol. 588, no. 2, pp. 224–230, Apr. 2007, doi: 10.1016/J.ACA.2007.01.079.

[211] H. Cen and Y. He, “Theory and application of near infrared reflectance spectroscopy in determination of food quality,” *Trends Food Sci. Technol.*, vol. 18, no. 2, pp. 72–83, Feb. 2007, doi: 10.1016/J.TIFS.2006.09.003.

[212] A. Wang, D. Hu, and L. Xie, “Comparison of detection modes in terms of the necessity of visible region (VIS) and influence of the peel on soluble solids content (SSC) determination of navel orange using VIS-SWNIR spectroscopy,” *J. Food Eng.*, vol. 126, pp. 126–132, Apr. 2014, doi: 10.1016/J.JFOODENG.2013.11.011.

[213] L. L. Mutton, B. R. Cullis, and A. B. Blakeney, “The objective definition of eating quality in rockmelons (Cucumis melo),” *J. Sci. Food Agric.*, vol. 32, no. 4, pp. 385–391, Apr. 1981, doi: 10.1002/JSFA.2740320412.

[214] “Postharvest Technology of Horticultural Crops - Adel A. Kader - Google Books.” https://books.google.com.pk/books?hl=en&lr=&id=O1zhx2OWftQC&oi=fnd&pg=PA287&dq=+A.+Kader.+(1985).+Standardization+and+inspection+of+fresh+fruits+and+vegetables,+Postharvest+Technology+of+Horticultural+Crops,+287–299.&ots=4iw590zkKO&sig=KjXL-zYTuR0n94dPYMHZyJjmG_A&redir_esc=y#v=onepage&q&f=false (accessed Apr. 11, 2022).

[215] E. Chace, C. Church, and F. Denny, “Relation between the composition of California cantaloupes and their commercial maturity,” 1924. https://agris.fao.org/agris-search/search.do?recordID=US201300427302 (accessed Apr. 11, 2022).

[216] T. M. Cu, R. R. En, C. E. And, R. U. Ssell, and L. A. Rs, “REFRACTIVE INDEX AS AN ESTIMATE OF QUALITY BETWEEN AND WITHIN MUSKMELON FRUITS,” *Plant Physiol.*, vol. 16, no. 3, p. 611, Jul. 1941, doi: 10.1104/PP.16.3.611.

[217] J. Sun, B. Ma, J. Dong, R. Zhu, R. Zhang, and W. Jiang, “Detection of internal qualities of hami melons using hyperspectral imaging technology based on variable selection algorithms,” *J. Food Process Eng.*, vol. 40, no. 3, p. e12496, Jun. 2017, doi: 10.1111/JFPE.12496.

[218] P. Stchur, D. Cleveland, J. Zhou, and R. G. Michel, “A REVIEW OF RECENT APPLICATIONS OF NEAR INFRARED SPECTROSCOPY, AND OF THE CHARACTERISTICS OF A NOVEL PbS CCD ARRAY-BASED NEAR-INFRARED SPECTROMETER,” *http://dx.doi.org/10.1081/ASR-120016293*, vol. 37, no. 4, pp. 383–428, Dec. 2007, doi: 10.1081/ASR-120016293.

[219] B. Jamshidi, S. Minaei, E. Mohajerani, and H. Ghassemian, “Reflectance Vis/NIR spectroscopy for nondestructive taste characterization of Valencia oranges,” *Comput. Electron. Agric.*, vol. 85, pp. 64–69, Jul. 2012, doi: 10.1016/J.COMPAG.2012.03.008.

[220] A. M. Cavaco, P. Pinto, M. D. Antunes, J. M. da Silva, and R. Guerra, “‘Rocha’ pear firmness predicted by a Vis/NIR segmented model,” *Postharvest Biol. Technol.*, vol. 51, no. 3, pp. 311–319, Mar. 2009, doi: 10.1016/J.POSTHARVBIO.2008.08.013.

[221] Y. Liu, X. Chen, and A. Ouyang, “Nondestructive determination of pear internal quality indices by visible and near-infrared spectrometry,” *LWT - Food Sci. Technol.*, vol. 41, no. 9, pp. 1720–1725, Nov. 2008, doi: 10.1016/J.LWT.2007.10.017.

[222] “| Ministry of Finance | Government of Pakistan |.” https://www.finance.gov.pk/survey_2021.html (accessed Apr. 12, 2022).

[223] D. C. Slaughter, “Nondestructive Maturity Assessment Methods for Mango: A Review of Literature and Identification of Future Research Needs,” 2009.

[224] S. Cubero, W. S. Lee, N. Aleixos, F. Albert, and J. Blasco, “Automated Systems Based on Machine Vision for Inspecting Citrus Fruits from the Field to Postharvest—a Review,” *Food Bioprocess Technol.*, vol. 9, no. 10, pp. 1623–1639, Oct. 2016, doi: 10.1007/S11947-016-1767-1/TABLES/2.

[225] L. S. Magwaza, U. L. Opara, H. Nieuwoudt, P. J. R. Cronje, W. Saeys, and B. Nicolaï, “NIR Spectroscopy Applications for Internal and External Quality Analysis of Citrus Fruit-A Review,” *Food Bioprocess Technol.*, vol. 5, no. 2, pp. 425–444, Feb. 2012, doi: 10.1007/S11947-011-0697-1/TABLES/6.

[226] G. P. Parpinello, G. Nunziatini, A. D. Rombolà, F. Gottardi, and A. Versari, “Relationship between sensory and NIR spectroscopy in consumer preference of table grape (cv Italia),” *Postharvest Biol. Technol.*, vol. 83, pp. 47–53, Sep. 2013, doi: 10.1016/J.POSTHARVBIO.2013.03.013.

[227] G. Giovanelli, N. Sinelli, R. Beghi, R. Guidetti, and E. Casiraghi, “NIR spectroscopy for the optimization of postharvest apple management,” *Postharvest Biol. Technol.*, 2014, doi: 10.1016/j.postharvbio.2013.07.041.

[228] R. Cubeddu *et al.*, “Time-Resolved Reflectance Spectroscopy Applied to the Nondestructive Monitoring of the Internal Optical Properties in Apples:,” *http://dx.doi.org/10.1366/0003702011953496*, vol. 55, no. 10, pp. 1368–1374, Aug. 2001, doi: 10.1366/0003702011953496.

[229] P. Mishra and D. Passos, “Realizing transfer learning for updating deep learning models of spectral data to be used in new scenarios,” *Chemom. Intell. Lab. Syst.*, vol. 212, May 2021, doi: 10.1016/J.CHEMOLAB.2021.104283.

[230] N. Anderson, K. Walsh, and P. Subedi, “Mango DMC and spectra Anderson et al. 2020,” 2020.